

RESEARCH ARTICLE

Functional and genomic characterisation of a xenograft model system for the study of metastasis in triple-negative breast cancer

Cameron N. Johnstone^{1,2,3,4,5,6,7,*‡}, Andrew D. Pattison^{6,8}, Kylie L. Gorringer^{1,2}, Paul F. Harrison⁹, David R. Powell^{8,9}, Peter Lock¹⁰, David Baloyan⁴, Matthias Ernst^{4,5}, Alastair G. Stewart⁷, Traude H. Beilharz^{6,8,*‡} and Robin L. Anderson^{1,2,3,4,5,*‡}

ABSTRACT

Triple-negative breast cancer (TNBC) represents 10-20% of all human ductal adenocarcinomas and has a poor prognosis relative to other subtypes. Hence, new molecular targets for therapeutic intervention are necessary. Analyses of panels of human or mouse cancer lines derived from the same individual that differ in their cellular phenotypes but not in genetic background have been instrumental in defining the molecular players that drive the various hallmarks of cancer. To determine the molecular regulators of metastasis in TNBC, we completed a rigorous *in vitro* and *in vivo* characterisation of four populations of the MDA-MB-231 human breast cancer line ranging in aggressiveness from non-metastatic to spontaneously metastatic to lung, liver, spleen and lymph node. Single nucleotide polymorphism (SNP) array analyses and genome-wide mRNA expression profiles of tumour cells isolated from orthotopic mammary xenografts were compared between the four lines to define both cell autonomous pathways and genes associated with metastatic proclivity. Gene set enrichment analysis (GSEA) demonstrated an unexpected association between both ribosome biogenesis and mRNA metabolism and metastatic capacity. Differentially expressed genes or families of related genes were allocated to one of four categories, associated with either metastatic initiation (e.g. *CTSC*, *ENG*, *BMP2*), metastatic virulence (e.g. *ADAMTS1*, *TIE1*), metastatic suppression (e.g. *CST1*, *CST2*, *CST4*, *CST6*, *SCNNA1*, *BMP4*) or metastatic avirulence (e.g. *CD74*). Collectively, this model system based on MDA-MB-231 cells should be useful for the assessment of gene function in the

metastatic cascade and also for the testing of novel experimental therapeutics for the treatment of TNBC.

This article has an associated First Person interview with the first author of the paper.

KEY WORDS: Triple-negative, Breast cancer, Metastasis, Mouse model, Xenograft

INTRODUCTION

Genetic and genomic analyses have shown that mammary ductal adenocarcinoma, the major form of breast cancer in humans, clusters into several distinct subtypes that are associated with significant differences in patient mortality rates (Curtis et al., 2012; Perou et al., 2000; Sorlie et al., 2001). Unlike hormone-receptor-positive tumours that express the oestrogen receptor (ER) and progesterone receptor (PR), or HER-2-positive neoplasms that express the erb-b2 transmembrane receptor, triple-negative breast cancers (TNBCs) are negative for all three receptors (Nielsen et al., 2004; Prat et al., 2013). Since approved targeted therapies are lacking, the standard treatment modalities for TNBC are adjuvant radio- and chemotherapy following surgery. Despite aggressive intervention, TNBC displays the worst outcome of all breast cancer subtypes due to its propensity for early relapse and the development of resistance to chemotherapeutic drugs (Carey et al., 2007; Eckhardt et al., 2012; Liedtke et al., 2008). The discovery of new molecular targets for TNBC is also complicated by considerable inter- and intra-tumour heterogeneity. Indeed, genome-wide expression and DNA copy number analyses have described at least four distinct variants (Hennessy et al., 2009; Lehmann et al., 2011, 2016; Prat et al., 2010; Turner and Reis-Filho, 2013).

To discover the cellular alterations that lead to acquisition of intrinsic malignant cancer cell phenotypes, such as unlimited growth potential, metabolic rewiring, invasive capacity, metastatic ability and chemoresistance (Hanahan and Weinberg, 2011), analyses of panels of congenic breast cancer lines derived from the same individual but differing in these cellular phenotypes have been conducted (Aslakson and Miller, 1992; Chang et al., 2007; Eckhardt et al., 2005; Halldorsson et al., 2017; Johnstone et al., 2015; Minn et al., 2005b; Morris et al., 1993; Santner et al., 2001; Weaver et al., 1995). This has led not only to the elucidation of gene networks underlying these processes, but also to potential new molecular targets for therapy (Eckhardt et al., 2005; Johnstone et al., 2015; Minn et al., 2005a).

Here, we comprehensively characterised the orthotopic growth and metastatic ability of four different variants of the human TNBC cell line MDA-MB-231, originally isolated from the pleural effusion of a patient with metastatic breast cancer (Cailleau et al., 1974). Genome-wide expression profiling was completed by RNA-sequencing

¹Cancer Research Division, Peter MacCallum Cancer Centre, Victorian Comprehensive Cancer Centre, Parkville, Victoria 3050, Australia. ²Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Victoria 3050, Australia. ³Department of Pathology, University of Melbourne, Parkville, Victoria 3010, Australia. ⁴Olivia Newton-John Cancer Research Institute, Heidelberg, Victoria 3084, Australia. ⁵School of Cancer Medicine, La Trobe University, Bundoora, Victoria 3086, Australia. ⁶Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia. ⁷Department of Pharmacology & Therapeutics, University of Melbourne, Parkville, Victoria 3010, Australia. ⁸Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia. ⁹Monash Bioinformatics Platform, Monash University, Clayton, Victoria 3800, Australia. ¹⁰LIMS Bioimaging Facility, La Trobe Institute for Molecular Science, La Trobe University, Bundoora, Victoria 3086, Australia.

*These authors are co-senior authors

‡Authors for correspondence (cameron.johnstone@onjcri.org.au; traude.beilharz@monash.edu; robin.anderson@onjcri.org.au)

© C.N.J., 0000-0002-6464-8661; K.L.G., 0000-0001-5681-2022; P.L., 0000-0003-4421-3609; T.H.B., 0000-0002-8942-9502; R.L.A., 0000-0002-6841-7422

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

(RNA-Seq) on primary tumour cells isolated from non-metastatic MDA-MB-231_ATCC (231_ATCC) (Cailleau et al., 1974), moderately metastatic MDA-MB-231_LNA (231_LNA) and highly-metastatic MDA-MB-231_LM2 (231_LM2) (Minn et al., 2005a), and MDA-MB-231_HM.LNm5 (231_HM.LNm5) (Chang et al., 2007, 2008; Fietz et al., 2017) mammary xenografts. The DNA content and genomic DNA copy number alterations were also analysed in each of the four lines. Differentially expressed genes or gene families were allocated to one of four categories, associated with either metastatic initiation, metastatic virulence, metastatic suppression or metastatic avirulence (Nguyen et al., 2009).

RESULTS

Characterisation of spontaneous metastatic potential of MDA-MB-231-derived lines

To identify genes involved in progression and metastasis of human TNBC, we firstly comprehensively characterised four human breast cancer xenograft models for differences in spontaneous metastatic capacity. Each tumour line was derived from MDA-MB-231 human TNBC cells (Fig. S1), originally isolated from the pleural effusion of a patient with metastatic breast cancer (Cailleau et al., 1974). Each of the four lines was shown to match MDA-MB-231 by short tandem repeat (STR) profiling (see Materials and Methods). 231_ATCC is an early-passage line (passage number < 20). The 231_I line is a population of MDA-MB-231 cells subcultured more than 50 times *in vitro* (Mongroo et al., 2004), and its daughter line, 231_LNA, was isolated from a lymph node metastasis that developed from a 231_I primary orthotopic tumour (Fig. S1, Table S1). 231_LM2 cells were isolated from an experimental lung metastasis that arose in mice inoculated with a reporter-gene-tagged version of the parental line, and have been extensively characterised elsewhere (Minn et al., 2005a). Lastly, 231_HM.LNm5 cells were isolated in our laboratory from a lymph node metastasis in a mouse bearing a MDA-MB-231HM primary mammary tumour (Fietz et al., 2017). The MDA-MB-231HM line was originally isolated from a spontaneous lung metastasis that arose following six rounds of *in vivo* passaging, whereby spontaneous secondary lesions forming in the lung were isolated and expanded *ex vivo* and subsequently re-implanted into the mammary gland (Chang et al., 2007, 2008). We and others have documented their aggressive metastatic phenotype in mice (Chang et al., 2015; Fietz et al., 2017; Jin et al., 2012; Le et al., 2016).

The relative abilities of the four tumour lines to metastasise spontaneously *in vivo* were compared side-by-side in a surgical resection model whereby single orthotopic mammary tumours were removed at a similar size (Fig. 1A, Figs S1 and S2). NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) immunodeficient mice were used because they were previously shown to support increased metastasis of breast cancer lines (Iorns et al., 2012; Puchalapalli et al., 2016). Whole animal bioluminescence imaging (BLI) revealed that mice bearing 231_ATCC tumours failed to show evidence of either local or distant recurrence by 22 days after resection, whereas each of the three other tumour lines demonstrated both local recurrence and distant relapse in the abdominal and thoracic cavities (Fig. 1B). *Ex vivo* fluorescent imaging revealed metastatic deposits in lung, liver and spleen from animals inoculated with either 231_LNA, 231_LM2 or 231_HM.LNm5 cells (Fig. S3), which was associated with considerable whole-organ hypertrophy of liver and spleen (data not shown). No metastatic lesions were found in secondary organs of mice bearing 231_ATCC tumours (Fig. S3). These findings were confirmed and extended by histological analyses of lung,

liver and spleen from tumour-bearing mice (Fig. 2). In addition to soft-tissue metastasis, 231_HM.LNm5 tumours also disseminated to spine in a minority of individuals (Fig. 2E). We also observed significant differences in the incidence of ipsilateral lymph node metastasis among the different models (Le et al., 2016). Metastasis to the draining lymph node in the axilla was observed in the 231_LM2 and 231_HM.LNm5 models but not in the 231_ATCC or 231_LNA lines (Fig. S4). Altogether, this extensive *in vivo* assessment demonstrates that the 231_ATCC model is incapable of spontaneous metastasis in NSG mice, whereas 231_LNA has an intermediate metastatic capacity, and the 231_LM2 and 231_HM.LNm5 tumours are highly metastatic.

231_ATCC was the slowest proliferating line, both *in vivo* (Fig. 1A) and in three-dimensional (3D) culture *in vitro* (Fig. 3C). 231_LNA and 231_LM2 formed loosely adherent invasive structures when cultured on a basement membrane gel *in vitro* (Fig. 3A), as expected for cancer cells with metastasising capability. However, despite the rapid proliferation and aggressive metastatic phenotype of 231_HM.LNm5 cells *in vivo*, they unexpectedly formed non-invasive near-spherical clusters in 3D culture, a phenotype that closely resembled the morphology of non-metastatic 231_ATCC cells (Fig. 3A). Both parental 231_HM cells and the 231_HM.LNm5 line showed substantially reduced *in vitro* migration towards serum than the other three lines (Fig. 3B). Indeed, we showed recently that 231_HM.LNm5 cells are less motile than the 231_ATCC cells *in vitro* (Fietz et al., 2017).

Analysis of DNA content and DNA copy number in MDA-MB-231 variants

MDA-MB-231 cells are near triploid and display trisomy of most autosomes (Satya-Prakash et al., 1981). To assess potential heterogeneity within each MDA-MB-231-derived population, cell cycle analysis was conducted on asynchronously growing untagged versions of the four lines by flow cytometry. 231_ATCC, 231_I (from which 231_LNA was isolated) and 231_LM2 showed nearly identical DNA content and cell cycle distributions, with each suggestive of homogeneous or near-homogeneous populations (Fig. 4A-C). In contrast, parental 231_HM cells exhibited a cell cycle distribution indicative of two distinct cell populations (Fig. 4D). One population had an identical DNA content to the other three lines (representing 42% of the total), whereas an 'aneuploid' fraction represented the majority (58%) of the population. The 'aneuploid' cells displayed a 1.78-fold increase in total DNA content compared to the less abundant fraction. This is consistent with duplication of an entire diploid genome within near-triploid parental MDA-MB-231 cells, thus yielding a probable near-quintaploid daughter line. To explore this further, clonal populations of parental 231_HM cells were generated by limiting dilution and two daughter lines were characterised. 231_HMcloneB2 displayed a cell cycle distribution identical to that of the minor 231_HM fraction and to the other three cell lines (Fig. 4E). Conversely, 231_HMcloneC7 showed a distribution consistent with a homogenous population of the major near-quintaploid 231_HM population (Fig. 4F). These findings demonstrate that each of the two significant 231_HM populations can exist independently of one another. Both 231_HMcloneB2 and 231_HMcloneC7 lines were spontaneously metastatic to lung, liver and spleen when analysed in an orthotopic resection model *in vivo* (data not shown), suggesting that aggressive metastatic capacity is not specifically associated with DNA ploidy in the 231_HM model.

To define DNA copy number variations (CNVs) in the four MDA-MB-231-derived lines, genomic DNA from untagged

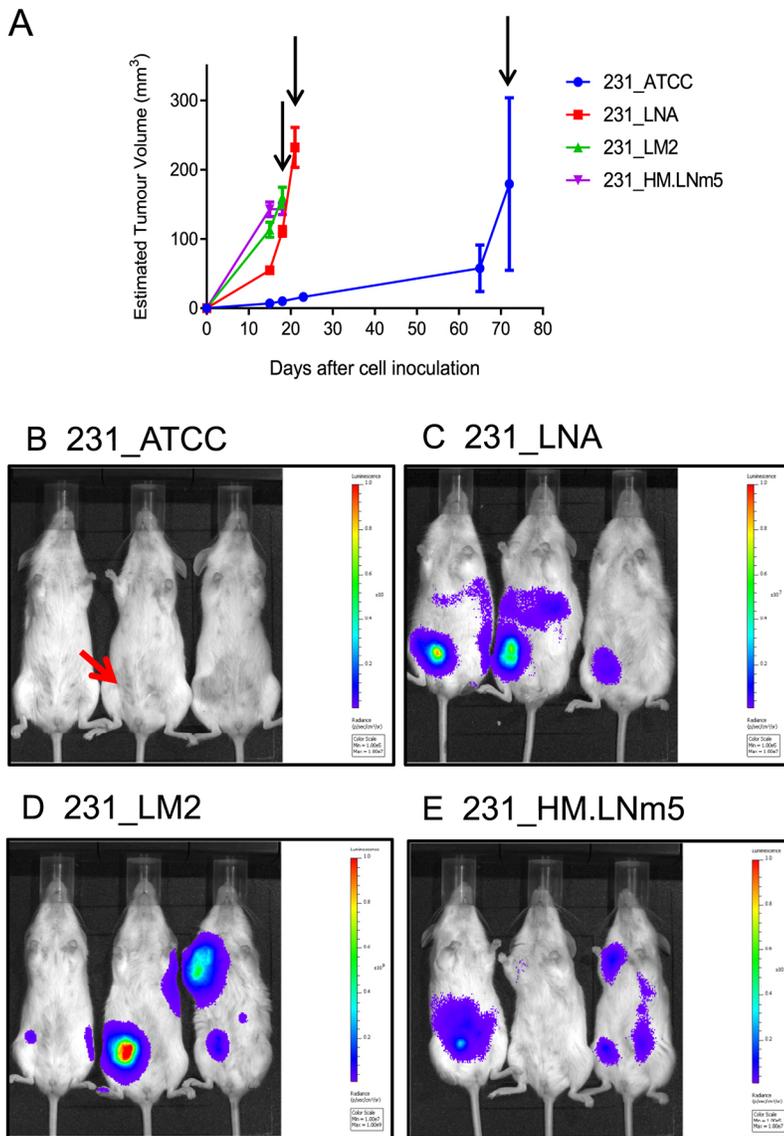


Fig. 1. Comparison of four MDA-MB-231-derived tumour variants *in vivo*. (A) Primary-tumour growth rates were measured using electronic callipers. Mean tumour volume \pm s.e.m. is shown. 231_ATCC ($n=3$), 231_LNA ($n=4$), 231_LM2 ($n=4$), 231_HM.LNm5 ($n=4$). Differences in growth rates were determined using mixed-effects linear regression modelling (Johnstone et al., 2015): 231_ATCC vs 231_LNA ($P=0.002$), 231_ATCC vs 231_LM2 ($P=0.002$), 231_ATCC vs 231_HM.LNm5 ($P=0.001$), 231_LNA vs 231_LM2 ($P=0.006$), 231_LNA vs 231_HM.LNm5 ($P=0.002$), 231_LM2 vs 231_HM.LNm5 ($P=0.168$). 231_LM2 and 231_HM.LNm5 primary tumours were surgically resected at day 18, 231_LNA at day 21 and 231_ATCC at day 72 after inoculation (arrows). (B-E) *In vivo* bioluminescence imaging of breast cancer xenograft models. Luciferase images of live mice were captured 22 days following surgical resection of the primary mammary tumour for all models. Both local and distant tumour recurrence was present in each of the three metastatic models [(C) 231_LNA, (D) 231_LM2, (E) 231_HM.LNm5] but not in mice inoculated with 231_ATCC cells (B). Three mice are shown per model [$n=4$ for each model except for 231_ATCC ($n=3$)]. 231_LM2 cells express Firefly luciferase-2 (*luc2*), radiance scale 1×10^7 (min) to 1×10^9 (max). The other three models express the dimmer Firefly luciferase-1 (*luc1*), radiance scales 1×10^5 (min) to 1×10^7 (max). The site of original primary tumour formation in the right-side inguinal mammary gland is indicated with a red arrow in one mouse as an example (B).

versions of the cell lines (231_ATCC, 231_I, 231_LM2, 231_HM) cultured *in vitro* was applied to SNP arrays (Fig. 4G, Fig. S5). Each line showed a unique pattern of DNA copy number alteration, although common regions of gain/loss and allelic imbalance were also present, including shared homozygous deletion of the *CDKN2A* (encoding p16/INK4A and p14ARF)/*CDKN2B* (encoding p15/INK4B)/*MTAP* locus at *9p21.3* (Fig. 4G, Fig. S5). High level ($\text{CNV } \log_2 \text{ ratio} \geq 1$) DNA amplification was not found in any line (Fig. 4G, Fig. S5). Through analysis of both shared and unique DNA copy number gains/losses across the genome (Table S2), the genetic divergence of each line from 231_ATCC was estimated. 231_LNA was most similar to 231_ATCC (15.2% difference), followed by 231_LM2 (25.8%), with 231_HM being the most divergent (66.1%). However, the latter result should be interpreted with caution since the 231_HM SNP array represents the merging of two different aneuploid populations.

Differential gene expression among MDA-MB-231 variants

To identify the factor(s) underlying the vastly different metastatic capacities of the four congenic models, resected primary tumours

were disaggregated, total RNA recovered from isolated cancer cells (Fig. S1) and digital RNA sequencing (RNA-Seq) conducted (Harrison et al., 2015). Consistent patterns of differential gene expression were found among the four MDA-MB-231 variants (Fig. 5). Gene set enrichment analysis (GSEA) was carried out on the RNA-Seq data to ascertain cellular processes that were deregulated in all three metastatic lines (231_LNA, 231_LM2, 231_HM.LNm5) compared to the non-metastatic 231_ATCC line (Subramanian et al., 2005), and therefore might drive a metastatic phenotype. Remarkably, of the ten most significantly upregulated gene sets in metastatic cells, nine were related to RNA splicing and metabolism, or to ribosome biogenesis (Table 1, Fig. S6). To extend these findings, The Cancer Genome Atlas (TCGA) breast cancer dataset was then interrogated (Koboldt et al., 2012). Comparison of small tumours from stage I patients to primary neoplasms from patients with stage II-IV disease revealed that each of the nine data sets concerning RNA metabolism were also significantly upregulated in tumours from later stage patients and hence could be causally involved in the progression of human breast cancer (Table 1). Interestingly, significant downregulation of genes involved in mammary gland development and morphogenesis was

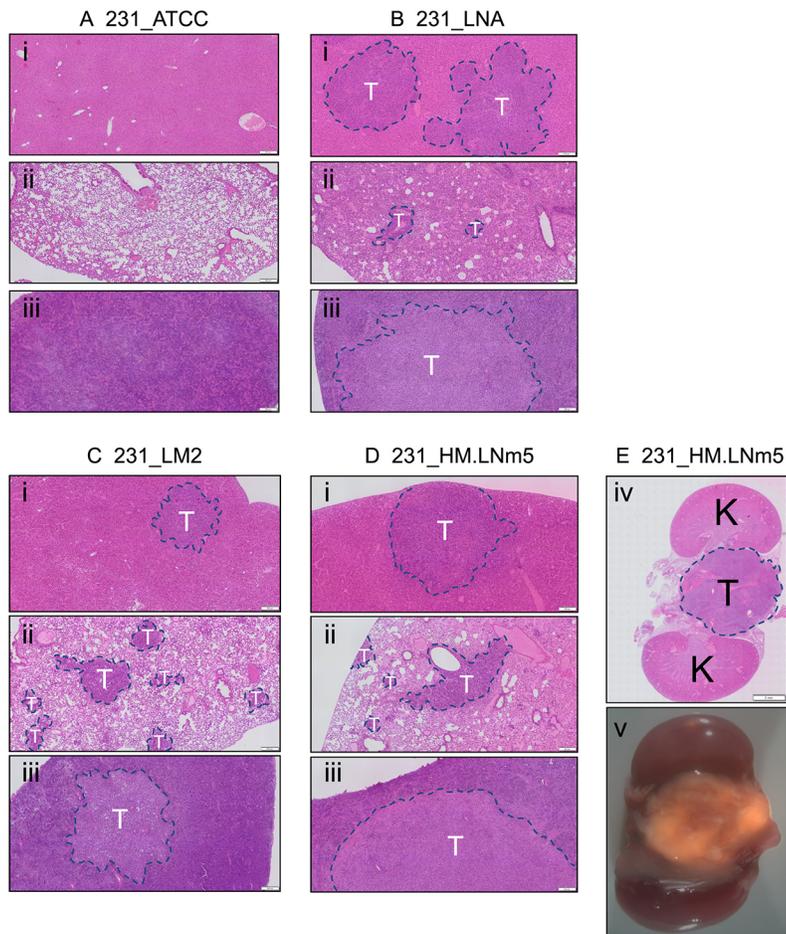


Fig. 2. Spontaneous metastasis of MDA-MB-231 variants to distant organs. Sections (3 µm) were cut from representative FFPE-fixed secondary organs obtained 22 days after primary tumour resection. H&E-stained slides were scanned using an Olympus VS120 instrument and images generated using OlyVIA software (Olympus). (A) 231_ATCC. (B) 231_LNA. (C) 231_LM2. (D,E) 231_HM.LNm5. (i) Liver, (ii) lung, (iii) spleen, (iv) metastatic deposit on the spine between the kidneys, (v) merged brightfield and fluorescent (tdTomato) images of a metastatic deposit on the spine between the kidneys (7× magnification). Metastatic lesions are indicated with a dashed blue line. T, region of tumour; K, kidney. Scale bars: (i-iii) 200 µm, (iv) 2 mm.

also found (Table 1), perhaps indicating that the acquisition of metastatic competency in MDA-MB-231 variants was associated with de-differentiation. One possible explanation for the GSEA results is that increased expression of ribosome biogenesis genes was a reflection of the enhanced requirement for protein synthesis in the faster-proliferating metastatic cells compared to 231_ATCC cells (Orsolich et al., 2016; van Riggelen et al., 2010). However, no evidence was found for increased expression of any gene sets involved in epithelial cell proliferation in the metastatic lines (Table 1). One process that can influence acquisition of metastatic capability in breast cancer is epithelial-to-mesenchymal transition (EMT), whereby a more mesenchymal phenotype can endow cancer cells with the ability to degrade basement membrane and invade into the surrounding tissue (Chaffer et al., 2016; Taube et al., 2010). To test for dysregulation of genes relevant to EMT, GSEA was carried out using ‘core EMT’ genes either up- or downregulated after induction of EMT in human mammary epithelial cells (Taube et al., 2010). The set of epithelial genes downregulated upon induction of EMT were significantly repressed in the three metastatic models compared to 231_ATCC (Table 1). This suggested that, despite MDA-MB-231 possessing a mesenchymal or ‘basal B’ phenotype (Lehmann et al., 2011; Neve et al., 2006), further reductions in epithelial gene expression levels are associated with acquisition of metastatic potential in MDA-MB-231-derived metastatic variants. Significant changes in mesenchymal genes upregulated upon EMT induction were not found (data not shown). To further probe this aspect, the expression and distribution of a key epithelial marker (E-cadherin) and a key mesenchymal marker

(vimentin) were examined. E-cadherin mRNA was not detected in MDA-MB-231-derived tumour cells by RNA-Seq and E-cadherin protein was not present at cell-cell junctions in any of the four lines when evaluated *in vitro* or *in vivo* (Fig. S7B). Vimentin mRNA was abundant in the tumour cells of each model (data not shown), but was not differentially expressed (Fig. S7A). A filamentous cytoplasmic staining pattern for vimentin was present in each of the four cell lines *in vitro* and strong cytoplasmic staining was revealed in each line *in vivo* (Fig. S7C). All four tumour models stained weakly positive when a pan-cytokeratin (CK) antibody (recognising CK1, CK5, CK6 and CK8) was applied (Fig. S8). Furthermore, no differences in either E-cadherin or vimentin staining intensity or distribution were observed in the two clonal 231_HM daughter lines, 231_HMcloneC7 and 231_HMcloneB2 (Fig. S9). Together, these data show that each of the four models possess a strong mesenchymal phenotype.

Dysregulation of individual genes across the four models was also observed and some of these could be grouped into broad categories or families (Table 2). We focused on two patterns of altered gene expression (Nguyen et al., 2009). We first examined transcripts whose expression was deregulated in all metastatic models versus non-metastatic 231_ATCC primary tumour cells. In this context, upregulated genes were designated as metastasis initiation genes and, conversely, downregulated genes were labelled as metastasis suppression genes (Tables S3-S4, Fig. S10). In a similar vein, mRNAs with augmented expression specifically in highly metastatic 231_HM.LNm5 cells compared with the other three models were labelled as metastasis virulence genes, whereas

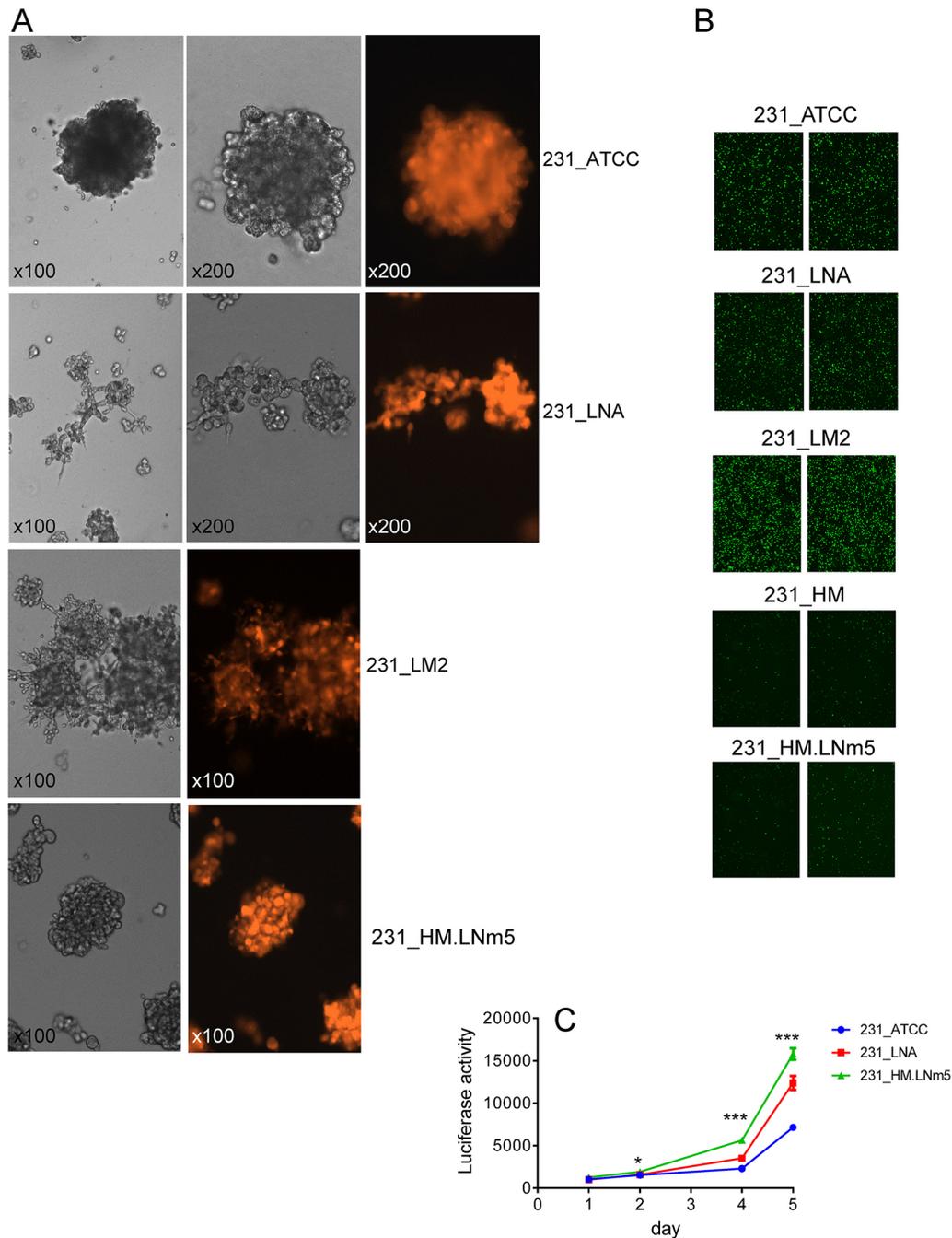
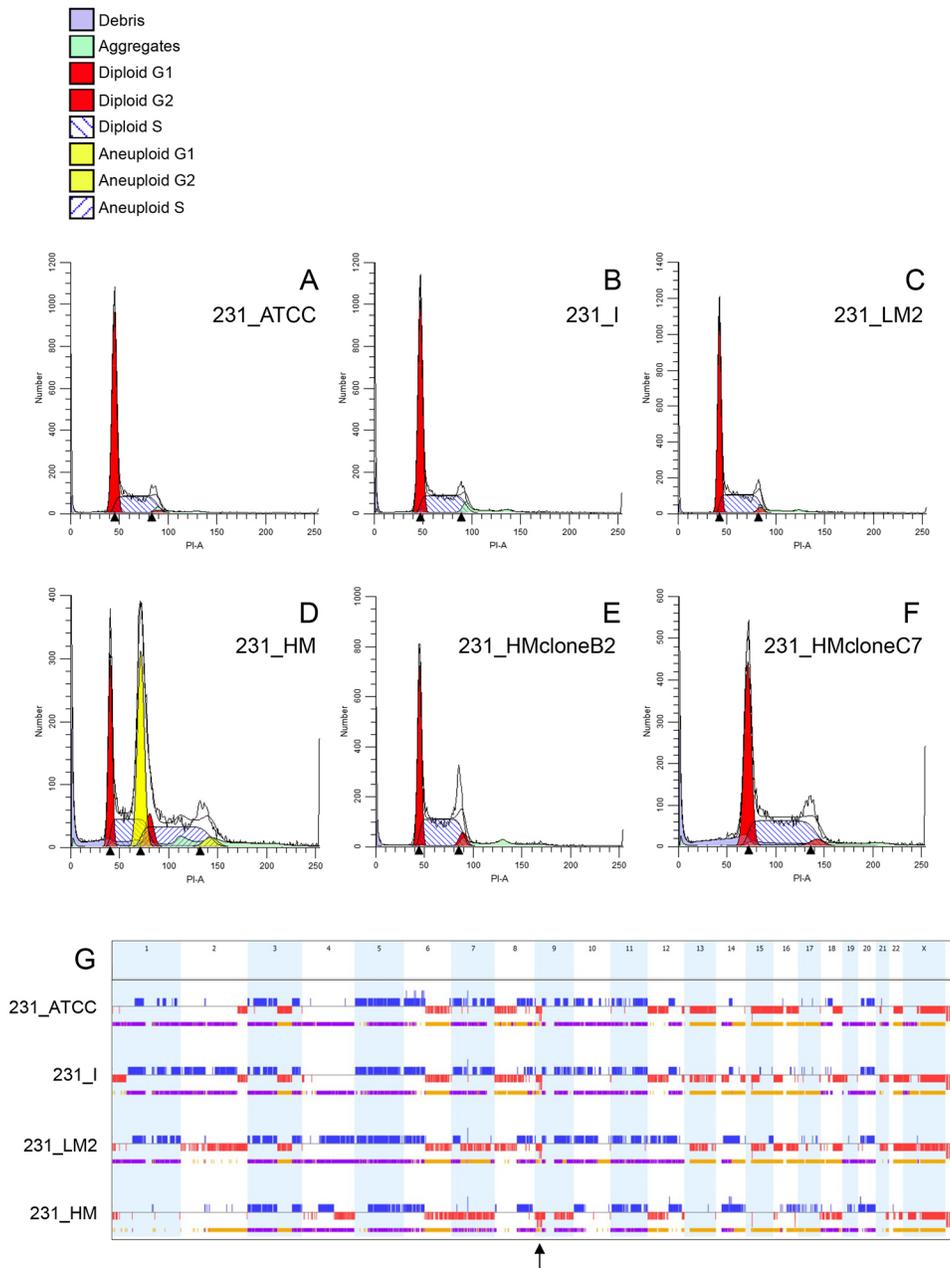


Fig. 3. *In vitro* phenotypes of different MDA-MB-231 variants. Proliferation rate, rather than an invasive growth pattern, is correlated with metastatic ability. (A) Images of cells cultured in 3D. The indicated cell lines were seeded on top of a 50% Cultrex matrix and images captured at the indicated magnifications after 5 days. Brightfield (greyscale) and matched tdTomato fluorescent images are shown. 231_ATCC and 231_HM.LNm5 lines formed non-invasive spheroids, whereas 231_LNA and 231_LM2 formed loosely connected invasive clusters. (B) *In vitro* motility of MDA-MB-231 variants. Transwell migration assays using 10% FBS as a chemoattractant ($n=3$ per line) were completed for the indicated cell lines. Parental MDA-MB-231HM (231_HM) cells were also included for comparison. Two representative fields ($\times 100$ magnification) from two different Transwells per cell line are displayed. (C) 3D *in vitro* proliferation (on 50% Cultrex) was measured over 5 days using the Cell Titer Glo assay (Promega) for the indicated cell lines. Mean \pm s.e.m. ($n=4$) is presented. * $P<0.05$. *** $P<0.001$ by Student's *t*-test.

the corresponding downregulated genes were categorised as metastasis avirulence genes (Tables S3-S4, Fig. S10).

Classes of molecules with confirmed roles in tumour progression were dysregulated (Table 2). The cysteine proteinase cathepsin C [CTSC (Ruffell et al., 2013)] was upregulated in all three metastatic models, whereas several of the type 2 cystatin genes, inhibitors of cysteine proteinases, were downregulated in each (Ai et al., 2006; Cox, 2009). The bone morphogenetic proteins BMP2 and BMP4

were dysregulated. BMP2, reported to promote invasion and progression of MCF-7 breast cancer cells (Clement et al., 2005), was substantially elevated in all three metastatic lines, whereas the breast cancer metastasis suppressor gene BMP4 (Cao et al., 2014) was suppressed in each of the three variants (Table 2). Expression of BMPRIA, which encodes the preferred receptor for both BMP2 and BMP4, was unchanged (data not shown). Interestingly, altered expression of gene clusters was also noted. Metallothioneins are

**Fig. 4. Cell cycle and SNP array analysis**

of four MDA-MB-231 cell line variants. (A-F) The indicated MDA-MB-231-derived cell lines were analysed for cell cycle distribution by flow cytometry. 231_HMcloneB2 (E) and 231_HMcloneC7 (F) are two different clonal daughter lines derived from parental 231_HM cells (D). (G) Genome-wide patterns of DNA copy number gain (blue), loss (red), loss of heterozygosity (yellow) and allelic imbalance (purple) derived from SNP array analysis are depicted linearly for the four indicated cell lines. The chromosome number is indicated at the top. The location of the homozygous deletion at 9p21.3 that is shared by the four cell lines is indicated by an arrow.

involved in the cellular metabolism of metal ions (Dziegiel et al., 2016). Four genes from the metallothionein cluster on human chromosome 16q13 (*MT1A*, *MT1E*, *MT1M*, *MT2A*) were overexpressed in 231_HM.LNm5 tumours only (Table 2), thus placing them in the metastatic virulence category (Table S4). Similarly, three genes from the collagen locus on 21q22 (*COL6A1*, *COL6A2*, *COL18A1*) were specifically repressed in 231_HM.LNm5. However, the type 2 cystatins (*CST1*, *CST2*, *CST4*) and the cystatin pseudogene *CSTP1*, all encoded by a gene cluster on 20p11, were downregulated in the three metastatic models compared to non-metastatic 231_ATCC. Additionally, *CST6*, a related type 2 cystatin gene located on a different chromosome, was strongly repressed in each of the three metastatic lines (Table 2).

Shao and colleagues previously reported increased angiogenesis in MDA-MB-231HM primary xenografts compared to tumours formed by the parental MDA-MB-231 line, as determined by

staining for microvessel density and microangiography (Jin et al., 2012), and our results concur with this (data not shown). To determine whether hypoxia might play a role in augmented neo-angiogenesis in the 231_HM.LNm5 primary tumours, the RNA-Seq data was analysed for hypoxia-induced pro-angiogenic factors. Surprisingly, the average expression level of the key angiogenesis-promoting gene, *VEGFA* (Ferrara, 2002), was lowest in the 231_HM.LNm5 tumour cells, and significantly lower than in non-metastatic 231_ATCC cells (Fig. S11Ai). Likewise, average levels of hypoxia-responsive carbonic anhydrase 9 (*CA9*) were also lowest in 231_HM.LNm5 (Fig. S11Aii). *In vitro* experiments revealed a modest elevation of *VEGFA* mRNA expression in 231_ATCC cells following exposure of cells to 1% oxygen for 6 h or 24 h, and this induction was blunted in 231_HM.LNm5 cells (Fig. S11B). This suggested that other factors in addition to *VEGFA* may promote vigorous angiogenesis in 231_HM.LNm5 primary xenografts.

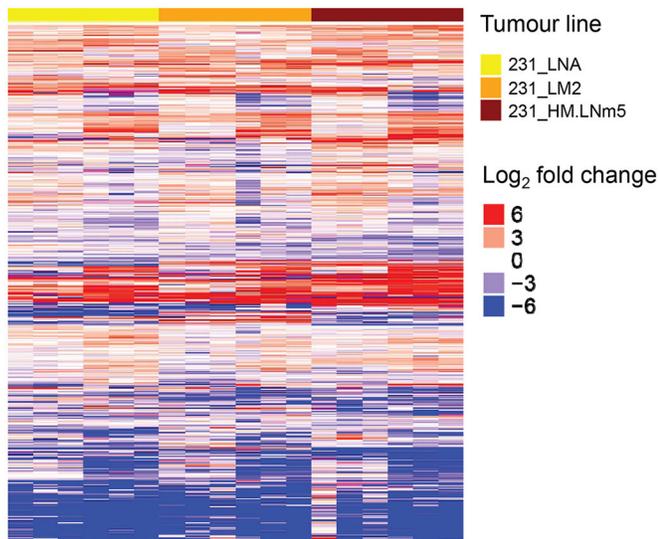


Fig. 5. Heat map of tumour cell gene expression patterns among MDA-MB-231 mammary xenografts. Each row represents a single gene.

Rows were clustered by similarity (Euclidian distance). Only genes with both a \geq twofold change (\log_2 fold change ≤ -1 or ≥ 1) in expression level compared to non-metastatic 231_ATCC tumour cells ($n=2$) and a mean sequencing transcript count of ≥ 30 are shown. 231_LNA ($n=6$), 231_LM2 ($n=6$), 231_HM.LNm5 ($n=6$). The full dataset of differentially expressed genes is provided in Table S3.

Another process that increases blood supply to the tumour is vascular mimicry, whereby cancer cells transdifferentiate into endothelial-like cells to form blood-carrying tubular structures that

can relieve hypoxia and also promote metastasis in mouse models (Seftor et al., 2012; Wagenblast et al., 2015). Three genes that contribute to vascular mimicry, endoglin (*ENG*; a component of the transforming growth factor beta receptor complex on endothelial cells), *TIE1* (encoding a key angiogenic receptor) and the vascular network-promoting anti-coagulant *SERPINE2* (Wagenblast et al., 2015), were all upregulated in metastatic cells (Table 2). Of note, endoglin is able to bind BMP-2 by interacting with the ligand-binding BMP type I receptors ALK3 (BMPRIA) and ALK6 (BMPRII) (Barbara et al., 1999), but whether ectopically upregulated BMP2 and ENG form a functional receptor complex in metastatic MDA-MB-231 variants remains to be elucidated. Thus, vasculogenic mimicry is a possible phenomenon promoting tumour progression, but further evaluation is required to confirm its presence and potential role in haematogenous metastasis of MDA-MB-231 variants. Finally, several genes encoding ion channels were dysregulated among the four tumours (Table 2). Expression of the voltage-gated potassium channels KCNU1, KCNQ3 and KCNAB2 was increased in highly metastatic lines, whereas *SCNNA*, encoding the alpha subunit of the major epithelial sodium channel (non-voltage-gated) and regulated by glucocorticoids (Fietz et al., 2017), was downregulated in all three metastatic tumours and is a candidate metastasis suppressor gene (Table 2). The C-C motif chemokine ligand-2 (*CCL2*) promotes breast cancer metastasis through engagement with its receptor CCR2 on macrophages (Kitamura et al., 2015). *CCL2* mRNA levels were at least fourfold higher in 231_HM.LNm5 primary tumour cells compared to the three other lines (Table 2). In addition, secreted *CCL2* protein was undetectable in 231_ATCC, 231_LNA and 231_LM2 3D cultures *in vitro* but was robustly expressed by 231_HM.LNm5 cells (data not shown). Interestingly, work by Kang

Table 1. Gene set enrichment analysis (GSEA) of RNA-Seq expression data

Gene set	Direction	RNA-Seq FDR	# genes in set (RNA-Seq)	TCGA RNA-Seq FDR	# genes in set (TCGA RNA-Seq)	Total # genes in set
GO_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS	Up	2.24E-06	317	2.29E-12	416	440
GO_RIBONUCLEOPROTEIN_COMPLEX_SUBUNIT_ORGANIZATION	Up	2.89E-06	146	3.91E-07	184	199
GO_RIBOSOME_BIOGENESIS	Up	2.69E-05	216	6.71E-13	294	308
GO_rRNA_METABOLIC_PROCESS	Up	3.18E-05	176	3.36E-12	244	255
GO_RNA_SPLICING_VIA_TRANSESTERIFICATION_REACTIONS	Up	5.61E-05	204	1.16E-09	256	267
GO_RIBONUCLEOPROTEIN_COMPLEX	Up	6.35E-05	503	9.47E-10	674	721
GO_mRNA_METABOLIC_PROCESS	Up	9.42E-05	424	1.69E-06	563	611
GO_MULTI_ORGANISM_METABOLIC_PROCESS	Up	9.80E-05	88	6.48E-07	136	138
GO_CATALYTIC_STEP_2_SPLICEOSOME	Up	1.37E-04	69	5.82E-08	84	89
GO_RNA_PROCESSING	Up	1.39E-04	591	1.91E-09	774	835
GO_POSITIVE_REGULATION_OF_EPITHELIAL_CELL_PROLIFERATION	Down	>0.05	–	–	–	100
GO_REGULATION_OF_EPITHELIAL_CELL_PROLIFERATION	Down	>0.05	–	–	–	190
GO_EPITHELIAL_CELL_PROLIFERATION	Down	>0.05	–	–	–	63
GO_MAMMARY_GLAND_EPITHELIAL_CELL_PROLIFERATION	Down	>0.05	–	–	–	11
GO_POSITIVE_REGULATION_OF_CELL_PROLIFERATION	Down	>0.05	–	–	–	482
GO_REGULATION_OF_CELL_PROLIFERATION	Down	>0.05	–	–	–	906
GO_MAMMARY_GLAND_ALVEOLUS_DEVELOPMENT	Down	1.44E-02	8	–	–	17
GO_MAMMARY_GLAND_LOBULE_DEVELOPMENT	Down	1.44E-02	8	–	–	17
GO_MAMMARY_GLAND_MORPHOGENESIS	Down	1.62E-02	27	–	–	40
GO_MAMMARY_GLAND_EPITHELIUM_DEVELOPMENT	Down	3.59E-02	34	–	–	53
'Core' EMT signature downregulated	Down	1.30E-04	82	–	–	160

The first ten rows (light grey) correspond to the top ten gene sets significantly upregulated [false discovery rate (FDR)<0.05] in each of the three metastatic tumours versus the non-metastatic 231_ATCC tumour. Results for gene sets corresponding to epithelial cell proliferation ($n=6$; dark grey) and mammary gland development ($n=4$; mid-grey), which are not dysregulated (FDR>0.05), as well as a gene set downregulated after EMT (Taube et al., 2010; white) are also shown. GSEA was conducted by interrogating the MSigDB database (<http://software.broadinstitute.org/gsea/msigdb/collections.jsp#C2>) as described in Materials and Methods. The top ten upregulated gene sets were also used to interrogate The Cancer Genome Atlas (TCGA) breast cancer dataset (<https://cancergenome.nih.gov/>). In this analysis primary, TNBC tumours from stage I patients (non-metastatic, $n=19$) were compared to TNBCs from stage III/III/IV patients ($n=97$). The number of dysregulated genes from each gene set is shown for each analysis.

Table 2. Differential gene expression among four MDA-MB-231-derived tumour models

Class	Gene symbol	RNA-Seq data			Adjusted <i>P</i> (ANOVA)	qRT-PCR			<i>P</i> (ANOVA)
		LNA	LM2	HM		LNA	LM2	HM	
Metallothioneins	<i>MT1A</i>	0.40	0.19	3.03	9.92E-11	–	–	–	–
	<i>MT1E</i>	1.08	0.90	3.98	2.37E-09	–	–	–	–
	<i>MT1M</i>	0.83	0.92	10.60	8.67E-07	–	–	–	–
	<i>MT2A</i>	1.69	1.26	4.20	3.56E-08	–	–	–	–
Proteases and inhibitors	<i>CTSC</i>	15.20	6.18	8.44	3.56E-13	8.15	3.37	5.91	5.02E-05
	<i>ADAMTS1</i>	0.29	1.09	69.59	7.40E-11	1.43	1.14	112.65	9.66E-03
	<i>TIMP3</i>	0.23	0.53	0.19	1.38E-06	–	–	–	–
	<i>CST1</i>	0.30	0.018	0.003	8.94E-11	–	–	–	–
	<i>CST2</i>	0.35	0.037	0.006	1.25E-09	–	–	–	–
	<i>CST3</i>	1.54	0.57	0.19	5.55E-15	–	–	–	–
	<i>CST6</i>	0.003	0.15	0.004	8.75E-13	0.003	0.044	0.0005	3.96E-04
	<i>CST4</i>	0.44	0.015	0.005	2.68E-12	–	–	–	–
	<i>CSTP1</i>	0.300	0.015	0.007	4.76E-12	–	–	–	–
Integrins and collagens	<i>ITGB4</i>	0.32	0.66	0.044	3.17E-13	–	–	–	–
	<i>COL18A1</i>	0.45	2.18	0.050	5.14E-17	–	–	–	–
	<i>COL6A1</i>	0.19	1.93	0.062	6.98E-16	–	–	–	–
	<i>COL6A2</i>	0.21	1.68	0.056	6.54E-13	–	–	–	–
Potassium and sodium channels	<i>COL13A1</i>	0.12	2.16	0.030	5.70E-12	–	–	–	–
	<i>KCNU1</i>	0.75	0.84	7.93	2.82E-06	–	–	–	–
	<i>KCNQ3</i>	0.26	0.36	7.85	3.92E-06	–	–	–	–
	<i>KCNAB2</i>	4.41	4.45	17.63	6.65E-06	–	–	–	–
Immune signalling	<i>SCNN1A</i>	0.021	0.067	0.015	2.33E-08	0.004	0.009	0.007	2.98E-02
	<i>CCL2</i>	0.41	0.45	4.36	9.46E-04	–	–	–	–
	<i>IL1RL1</i>	0.37	0.44	8.64	1.04E-07	–	–	–	–
	<i>CXCR4</i>	0.99	1.12	72.00	2.77E-11	–	–	–	–
	<i>PTPN7</i>	0.17	0.21	48.35	6.55E-13	–	–	–	–
	<i>CD33</i>	0.16	0.12	28.01	6.83E-14	–	–	–	–
	<i>CSF2RA</i>	1.10	1.99	7.42	7.76E-10	–	–	–	–
	<i>CD74</i>	1.32	0.58	0.007	1.15E-12	–	–	–	–
	<i>ENG</i>	87.12	38.25	5.34	1.67E-06	1060	538.8	144.4	4.78E-06
Vascular mimicry	<i>TIE1</i>	1.16	0.25	23.69	1.22E-08	4.69	0.78	94.25	8.77E-04
	<i>SERPINE2</i>	6.48	7.74	1.74	5.15E-06	–	–	–	–
	<i>BMP2</i>	12.80	11.76	48.12	3.26E-06	66.76	32.94	339.6	1.30E-05
BMP signalling	<i>BMP4</i>	0.09	0.62	0.07	7.35E-10	0.034	0.39	0.040	1.93E-02
	<i>LINC00707</i>	0.41	0.44	37.90	2.18E-12	–	–	–	–
LINC RNAs	<i>LINC00880</i>	0.94	0.20	4.15	2.45E-07	–	–	–	–

RNA-Seq was conducted using cancer cells isolated from primary mammary tumours established in NSG mice. Data were generated from the following models: 231_ATCC ($n=2$ tumours), 231_LNA ($n=6$), 231_LM2 ($n=6$), 231_HM.LNm5 ($n=6$). Changes in expression of selected genes (shaded grey) were validated by qRT-PCR using independent isolated tumour cell samples ($n=3$ per tumour type). Mean gene expression levels were set to 1 in 231_ATCC tumours for both RNA-Seq and qRT-PCR data. Gene expression ratios in the three metastatic tumours are shown, as are *P*-values (one-way ANOVA). Genes were selected for inclusion based on either fold increase/decrease of expression compared to 231_ATCC tumours, or adjusted *P*-value or biological function of the encoded protein. –, not analysed.

and Lu showed that ectopic expression of CCL2 in 231_LM2 cells enhanced experimental metastasis to lung via recruitment of tumour-promoting lung macrophages that display its cognate receptor CCR2 (Lu and Kang, 2009). Indeed, targeting the CCL2-CCR2 axis is being evaluated for therapy of breast cancer, although recent findings suggest that modulation of this pathway may result in deleterious unintended consequences in patients (Bonapace et al., 2014).

Gene methylation is associated with acquisition of aggressive metastatic capacity

Since robust alterations in gene expression were evident across the four tumour models from the RNA-Seq data, a preliminary investigation into the mechanisms underlying these differences was conducted. Gene CNVs are unlikely to underlie the substantial differences in gene expression observed because high-level amplifications and homozygous deletions were absent from all lines with the exception of deletion at *9p21.3* (*CDKN2A/B*), which was observed in all four cell lines (Fig. 4G). Changes in promoter methylation are associated with gain of

metastatic potential in breast cancer (Fang et al., 2011), so we investigated whether DNA methylation could be an alternative explanation for the marked suppression of selected genes in 231_HM.LNm5 tumour cells by using DNA demethylating agents and histone deacetylase inhibitors (Cameron et al., 1999). Both *BMP4* and *CST6* are metastasis suppression genes downregulated in 231_HM.LNm5 tumour cells relative to non-metastatic 231_ATCC tumour cells (Table 2), and both contain CpG islands spanning their transcription start sites in their 5' proximal regions (Fig. 6A,D). Suppression of *BMP4* (Fig. 6B) and *CST6* (Fig. 6E) levels in 231_HM.LNm5 cells relative to 231_ATCC cells was retained in 2D culture, suggesting that the mechanism(s) underlying downregulation are cancer cell intrinsic. Treatment with the demethylating agent 5-Aza-2'-deoxycytidine (5azadC) resulted in a modest two- to threefold induction of *BMP4* mRNA in 231_HM.LNm5 cells but no induction in 231_ATCC cells (Fig. 6C). The histone deacetylase (HDAC) inhibitor Trichostatin A (TSA) produced only a negligible increase (1.3- to 1.5-fold) in *BMP4* mRNA levels in both cell lines, showing that histone acetylation alone is not responsible for *BMP4* silencing in

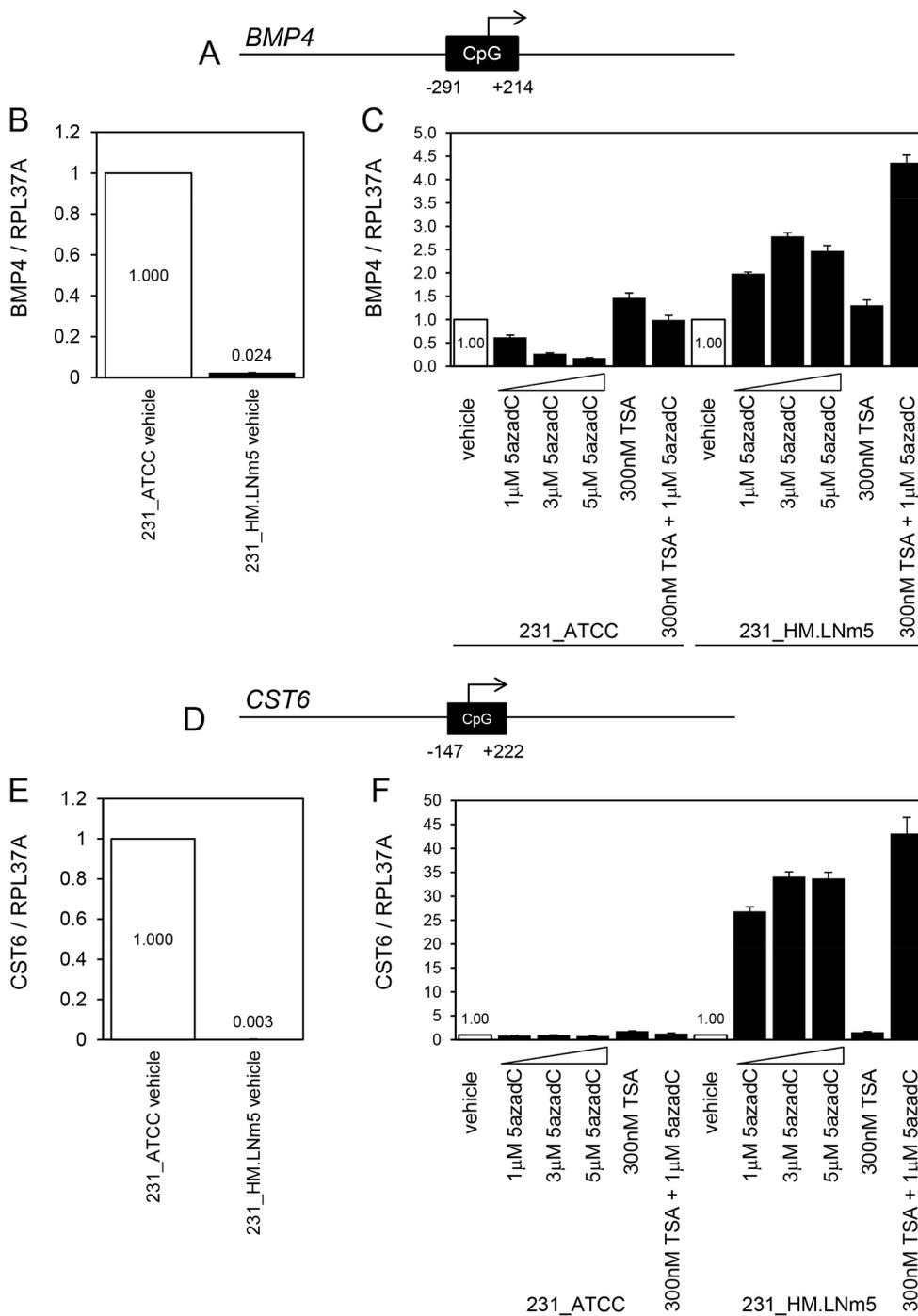


Fig. 6. Analysis of *BMP4* and *CST6* mRNA levels following treatment with demethylating agent and histone deacetylase inhibitor. (A) A CpG island (506 bp) spans the transcription start site of the human *BMP4* gene. (B) Relative *BMP4* mRNA levels in vehicle (DMSO)-treated 231_ATCC and 231_HM.LNm5 cells by TaqMan qRT-PCR. Expression in 231_ATCC was set to 1. Mean \pm s.d. ($n=3$). (C) TaqMan qRT-PCR analysis of *BMP4* mRNA expression levels in 5-Aza-2'-deoxycytidine (5azadC)- and Trichostatin A (TSA)-treated 231_ATCC and 231_HM.LNm5 cells. Expression in vehicle-treated cells was set to 1. Mean \pm s.d. ($n=3$). (D) A CpG island (370 bp) spans the transcription start site of the human *CST6* gene. (E) Relative *CST6* mRNA levels in vehicle (DMSO)-treated 231_ATCC and 231_HM.LNm5 cells by TaqMan qRT-PCR. Expression in 231_ATCC was set to 1. Mean \pm s.d. ($n=3$). (F) TaqMan qRT-PCR analysis of *CST6* mRNA expression levels in 5azadC- and TSA-treated 231_ATCC and 231_HM.LNm5 cells. Expression in vehicle-treated cells was set to 1. Mean \pm s.d. ($n=3$).

231_HM.LNm5 cells. Co-administration of TSA and 5azadC led to a further increase (4.4-fold) in *BMP4* expression specifically in the 231_HM.LNm5 cells. Treatment with 5azadC resulted in a marked 27- to 34-fold increase in *CST6* mRNA levels in 231_HM.LNm5 cells but no increase in 231_ATCC cells (Fig. 6F). As found for *BMP4*, TSA treatment enhanced *CST6* expression only modestly (1.6- to 1.8-fold) in both lines. TSA and 5azadC co-treatment elevated *CST6* mRNA 43-fold in 231_HM.LNm5 cells but again had no effect in 231_ATCC cells.

Together, these data indicate that the key CpG islands of *BMP4* and *CST6* spanning their transcription start sites are not methylated in 231_ATCC cells, and also suggest potential low-level methylation

of *BMP4* and robust methylation of *CST6* in 231_HM.LNm5 cells, although direct assessment of CpG island methylation is required to answer this conclusively. Work by others showed methylation of *CST6* to be associated with worse patient prognosis in breast cancer (Kioulafa et al., 2009), whereas the encoded protein (cystatin M) inhibited breast cancer cell proliferation, motility and metastasis (Zhang et al., 2004). *BMP4* was also shown to be methylated in breast cancer (Hartmann et al., 2009). Interestingly, CpG island methylation was also associated with EMT in a human breast cancer cell line model (Carmona et al., 2014), and could therefore contribute to the downregulation of epithelial genes seen in the three metastatic MDA-MB-231 variants compared with 231_ATCC (Table 1).

DISCUSSION

Associations among proliferation, motility, angiogenesis and spontaneous metastasis in MDA-MB-231 variants

Characterisation of mouse models of TNBC has improved our understanding of the cellular, molecular and physiological processes underlying multi-step metastasis; however, additional models with spontaneous metastatic capacity that are amenable to both gene modification and preclinical testing of experimental therapeutics would be beneficial.

Simultaneous assessment of the four MDA-MB-231-derived lines with different spontaneous metastatic potential showed that growth rate is correlated with metastatic ability in MDA-MB-231 variants, whereas invasive growth in 3D culture and *in vitro* migration are not, which concurs with published data (Fietz et al., 2017). This suggests that the ability of TNBC cells to invade through basement membrane is not the sole determinant of metastatic proclivity *in vivo* (Narod and Sopik, 2018), but rather that primary tumour proliferation rate may be better correlated with intrinsic metastatic potential, as suggested by others based on analysis of MDA-MB-231 variants (Minn et al., 2007).

Despite the increased angiogenesis found in 231_HM.LNm5 primary tumours compared to the other three lines (Jin et al., 2012 and data not shown), average levels of the central hypoxia-induced pro-angiogenic factor VEGFA were unexpectedly lowest in 231_HM.LNm5 and the induction of VEGFA mRNA by hypoxia was blunted in 231_HM.LNm5 cells compared to 231_ATCC cells *in vitro*. Along with low levels of CA9, this suggests that exponentially growing 231_HM.LNm5 tumours may experience less extreme hypoxia than the other tumour types and also that other factors may compensate for low VEGFA to promote angiogenesis. IL-8 protein was strongly expressed by 231_HM.LNm5 cells *in vitro* (data not shown) and was shown previously to compensate for reduced VEGFA levels to sustain angiogenesis in colorectal cancer xenografts (Mizukami et al., 2005). In addition, CCL2, which was upregulated in 231_HM.LNm5 tumour cells, can indirectly promote tumour angiogenesis via recruitment of macrophages (Low-Marchelli et al., 2013). Therefore, the 231_HM.LNm5 model should prove useful for the study of angiogenesis in primary breast cancers and its relationship to haematogenous metastasis.

Differential gene expression in primary tumour cells from MDA-MB-231 orthotopic xenografts

Unexpectedly, GSEA showed that several gene networks related to ribosome biogenesis and RNA metabolism were upregulated both in metastatic MDA-MB-231 cells as well as in whole human primary TNBCs corresponding to a more advanced stage. Although the vast majority of genes in each gene network were upregulated (Table 1), the magnitude of the change for individual genes in the three metastatic lines was generally subtle (data not shown), suggesting coordinate regulation at a systems level. Activity of these pathways is often correlated with cell proliferation in order to ensure enough translational output for cell division (Orsolich et al., 2016; van Riggelen et al., 2010). However, while each of the three metastatic MDA-MB-231 variants grew faster than parental 231_ATCC cells both *in vitro* and *in vivo*, gene sets related to the positive regulation of epithelial cell proliferation were not augmented in metastatic cells *in vivo*. Enhanced ribosome biogenesis and RNA metabolic pathways could also play a direct role in metastatic dissemination independently of effects on cell proliferation. Pursuit of this possibility will require consideration of the activity of the key regulator of translation,

mTORC1, and its downstream effectors p70S6K and 4E-BP1 (Steelman et al., 2016).

Genome-wide expression profiling of the primary tumour cells isolated from orthotopic xenografts enabled the delineation of both previously described and novel genes and gene families associated with pro- or anti-metastatic functions. The ease with which these breast tumour lines can be propagated both *in vitro* and *in vivo* and gene expression altered in a stable manner will allow the elucidation of the precise function(s) of dysregulated genes in the metastatic cascade, particularly for those genes likely to operate in a tumour cell autonomous mode. For example, the potential metastasis suppressor activities of *CST1*, *CST2*, *CST4*, *CST6*, *BMP4* and *SCNN1* could be individually tested by coupling silencing of endogenous expression in non-metastatic 231_ATCC cells with stable restoration of expression in the metastatic 231_LM2 and 231_HM.LNm5 lines. Conversely, the putative pro-metastatic activities of *ADAMTS1*, *CTSC*, *BMP2*, *ENG* and *TIE1* could be tested by ectopic over-expression in 231_ATCC cells paired with stable depletion in the 231_HM.LNm5 line. In addition to the assessment of gene function, the lines should be useful for the testing of experimental therapeutics, particularly specific small molecule inhibitors of the proteinases ADAMTS1 and CTSC and of the transmembrane tyrosine kinase TIE1.

MATERIALS AND METHODS

Cell lines and cell culture

The MDA-MB-231 human breast adenocarcinoma cell line was purchased from ATCC (Cailleau et al., 1974) and designated 231_ATCC. The MDA-MB-231_LM2 variant (231_LM2) (also designated clone 4175) was a kind gift from Joan Massague (Memorial Sloan Kettering Cancer Center, USA) (Minn et al., 2005a). The MDA-MB-231_LNA (231_LNA) line was isolated from a spontaneous lymph node metastasis that arose in a mouse inoculated orthotopically with late-passage MDA-MB-231_I cells (Fig. S1, Table S1; Johnstone et al., 2008). MDA-MB-231HM cells (Chang et al., 2007, 2008) were kindly provided by Zhi Min Shao and Zhou Luo Ou (Breast Cancer Institute, Fudan University, Shanghai, China). The MDA-MB-231HM.LNm5 (231_HM.LNm5) variant was derived as described (Fietz et al., 2017). To obtain clonal sublines of parental MDA-MB-231HM cells, cells were transduced with the pFB_neo_Fluc retroviral vector (a kind gift from Hiroshi Nakagawa, University of Pennsylvania, PA) and selected with G418 (1000 µg/ml) for 2 weeks, followed by seeding of single viable cells using flow cytometry (FACSDiva, Beckton Dickinson, Franklin Lakes, NJ, USA). Luciferase activity positive colonies were expanded and two clones (231_HMcloneC7 and 231_HMcloneB2) were selected for further analysis. All lines were maintained in Phenol-Red-containing Dulbecco's modified Eagle medium (DMEM; Thermo Fisher Scientific, Scoresby, Victoria, Australia) supplemented with 10% (v/v) heat-inactivated foetal bovine serum (FBS; Thermo Fisher Scientific), 15 mM HEPES buffer, 2 mM L-glutamine, 1% (v/v) non-essential amino acids, 5% (v/v) sodium pyruvate, penicillin (100 IU/ml) and streptomycin (100 µg/ml). Cells were maintained at 37°C in 5% CO₂ (v/v) in air and sub-cultured every 4-5 days. For 3D culture, cells were seeded onto a solid basement membrane gel (50% Cultrex in full medium; Trevigen, Gaithersburg, MD, USA). MDA-MB-231-derived cell lines were verified for authenticity using the AmpFLSTR™ Identifier™ PCR Amplification Kit (Thermo Fisher Scientific) at CellBank Australia (<http://www.cellbankaustralia.com>) and confirmed using the GenePrint® 10 System (Promega Corporation, Alexandria, NSW, Australia) at the QIMR Berghofer Medical Research Institute, Australia. According to the International Cell Line Authentication Committee (ICLAC; <http://iclac.org/>), cell lines are considered authentic if >80% of measured alleles [evaluation value (EV)>0.8] matched those of the repository sample (Capes-Davis et al., 2013). Individual EVs for the parental versions of the cell lines were: 231_ATCC (1.00), 231_I (1.00), 231_LM2 (0.80), 231_HM (0.96).

Treatment of cell lines with 5-Aza-2'-deoxycytidine and Trichostatin A

The DNA demethylating agent 5azadC and the histone deacetylase inhibitor TSA (Cayman Chemical, Ann Arbor, MI, USA) were dissolved in DMSO. Cells were treated with increasing doses of 5azadC or vehicle (DMSO) for 4 days, replacing the drug every 24 h. Cells were incubated with TSA alone for 24 h. For co-administration experiments, cells were first incubated with 5azadC (1 μ M) for 3 days, followed by co-incubation with 5azadC (1 μ M) and TSA (300 nM) for 24 h.

Generation of reporter-gene-expressing cell lines

The fluorescent protein-tagged and luciferase-tagged cell lines used in the comparison of MDA-MB-231 variants are summarized in Table S1. Briefly, amphotropic retroviruses encoding reporter genes were generated using 293T-based Phoenix-Ampho packaging cells (Swift et al., 2001). Viral supernatants were filtered through 0.45 μ m membranes and target cells transduced using the spin-infection method as previously described (Johnstone et al., 2008). Stably transduced cells were selected in the appropriate antibiotic followed by sorting for expression of fluorescent protein by flow cytometry (FACSDiva). Luciferase-expressing daughter lines co-expressing eGFP and/or tdTomato proteins (outlined in Table S1) were used in all *in vitro* and *in vivo* assays unless otherwise stated (Shaner et al., 2005). STR analysis was used to authenticate (EVs>0.80) all four reporter-gene-tagged MDA-MB-231 variants (data not shown).

Cell cycle analysis by flow cytometry

Parental 231_ATCC, 231_I, 231_LM2 and 231_HM cells in exponential growth phase were trypsinised and 5×10^5 cells fixed with cold 70% ethanol. Cells were washed twice with cold PBS and then incubated with RNaseA and propidium iodide (PI) in a total volume of 300 μ l. At least 10,000 single cells were analysed by flow cytometry (FACSCanto, Beckton Dickinson, Franklin Lakes, NJ, USA). Flow cytometry data were analysed with ModFit LT v5.0 software (Verity Software House, Topsham, ME, USA) using the manual setting. All coefficient of variation (CV) values were <6.0%.

Protein expression and distribution by immunofluorescence

Cells were cultured on 8-well plastic chamber slides (Nunc[®] Lab-Tek[®] Chamber Slide[™] system, Sigma Aldrich), fixed with 4% paraformaldehyde for 30 min and permeabilised using 0.1% Triton X-100 for 5 min. Slides were blocked with 10% horse serum/1% bovine serum albumin (1 h at room temperature). Mouse anti-human E-cadherin monoclonal antibody (clone 36, 1:100 dilution, BD Transduction Labs) or mouse anti-human vimentin monoclonal antibody (clone V9, 1:500 dilution, eBioscience, Thermo Fisher Scientific) were used with an Alexa-Fluor-568-conjugated anti-mouse secondary antibody (Molecular Probes). Nuclei were visualised using DAPI (Sigma Aldrich) and images generated using a LSM780 inverted confocal microscope (Carl Zeiss, North Ryde, NSW, Australia) with a 10 \times objective and associated ZEN software (Zeiss).

Protein expression and distribution by immunohistochemistry

Orthotopic xenografts were resected from mouse mammary glands and fixed in 10% neutral buffered formalin for 4–8 h. Heat-induced antigen retrieval was conducted for 15 min on 4 μ m sections submerged in citrate buffer (pH 6.0). Sections were blocked with 10% horse serum/1% bovine serum albumin (1 h at room temperature) and then incubated overnight (4°C) in primary antibody. Mouse anti-human E-cadherin monoclonal antibody (clone 36, 1:100 dilution, BD Transduction Labs), mouse anti-human vimentin monoclonal antibody (clone V9, 1:500 dilution, eBioscience, Thermo Fisher Scientific), or mouse anti-human pan-cytokeratin monoclonal antibody (clone PCK-26 ascites, 1:300 dilution, Sigma Aldrich) were used. PCK-26 recognises an epitope on the type II cytokeratins 1, 5, 6 and 8. Sections were incubated with a horseradish peroxidase (HRP)-conjugated anti-mouse secondary antibody (Dako, Agilent Technologies, Mulgrave, Victoria, Australia) for 1 h at room temperature prior to timed incubation with the chromogen 3,3'-diaminobenzidine (Dako). Sections were counterstained with haematoxylin to

visualise nuclei (blue) and slides were scanned using an Aperio Digital Pathology Slide Scanner (Leica Microsystems, Mount Waverley, Victoria, Australia). Images were generated using Aperio ImageScope software (Leica Microsystems).

Isolation of total RNA from primary tumour cells

Resected primary tumours were disaggregated by digestion with collagenase I (Worthington Biochemical Corporation, Lakewood, NJ, USA) and bovine pancreatic DNase I (Roche Diagnostics, North Ryde, NSW, Australia) at 37°C and filtered through a series of sieves prior to sorting of viable tumour cells by flow cytometry (FACSDiva, Beckton Dickinson). Hydroxystilbamidine (Fluoro-Gold, Abcam, Melbourne, Victoria, Australia) was used to exclude non-viable cells. Total RNA was prepared from sorted cells using the Direct-zol RNA mini kit (Zymo Corporation, Irvine, CA, USA) incorporating on-column DNase I digestion in accordance with the manufacturer's instructions.

Quantitative real-time RT-PCR (qRT-PCR)

First-strand cDNA was oligo dT-primed and synthesised from 5 μ g total RNA with M-MLV reverse transcriptase in a 20 μ l reaction volume (Promega). Two-step qPCR was completed (50–250 ng cDNA template/reaction) using either TaqMan gene expression assays (CTSC, ENG, TIE1, BMP2, BMP4, RPL37A) with Fast Universal PCR Master Mix, no AmpErase[™] UNG (Thermo Fisher Scientific) or Fast SYBR[™] Green Master Mix (50–250 ng cDNA template/reaction) and the following oligonucleotide primers (Integrated DNA Technologies, Singapore). ADAMTS1, F 5'-AGCCCAA-GGTTGTAGATG-3', R 5'-GCTTTTACACTGTCCTTG-3'; CST6, F 5'-TCCGAGACACGCACATCATC-3', R 5'-CCCATCTCCATCGTCA-GGAA-3'; SCNN1A, F 5'-TACTGCTACTATAAGCTCCAG-3', R 5'-TGTTGTTGACGGTGTAAATG-3'; RPL37A was used as an internal reference gene. Reactions (10 μ l volume) were conducted for 50 cycles using either Step One Plus or ViiA 7 real-time PCR instruments (Thermo Fisher Scientific).

Expression profiling by RNA-Seq

Total RNA integrity was checked using an RNA6000 Nano chip and 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Only samples with an RNA integrity score (RIN) \geq 8.5 were used. RNA-Seq was completed using the PAT-Seq approach with libraries constructed from 1 μ g total RNA. Sequencing was conducted as described previously (Harrison et al., 2015), except that 150 base chemistry was used on a HiSeq1500 system (Illumina, Scoresby, Victoria, Australia). Raw read filtering, quality checking and alignment against the Ensembl human genome reference (version hg38, release 82) was completed using the Tail-Tools bioinformatics suite (<https://github.com/Victorian-Bioinformatics-Consortium/tail-tools>). The data are available through the NCBI Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/gds/>) with accession number GSE101745.

Bioinformatics analyses

Differentially expressed genes were called by the in-house Tail-Tools pipeline, except in the case of GSEA, where raw gene-wise counts from Tail-Tools were first analysed using the Limma-voom approach (Ritchie et al., 2015). Any visualizations of RNA-Seq data were generated from raw gene-wise counts that were variance stabilized by the 'varistran' R package (<https://github.com/MonashBioinformaticsPlatform/varistran>). GSEA was applied using Camera (Wu and Smyth, 2012), with gene ontology (GO) gene sets taken from the MSigDB Molecular Signatures Database (<http://software.broadinstitute.org/gsea/msigdb/>) (Subramanian et al., 2005). To obtain breast cancer data from TCGA, raw gene expression counts and patient information were downloaded using the GDC data portal (<https://portal.gdc.cancer.gov/>). TNBCs were selected based on clinical immunohistochemical data and divided by stage. TCGA gene expression analyses were conducted using the Limma-voom approach and GSEA was performed as for RNA-Seq data. CpG islands were discovered using the web-based Softberry CpGfinder (www.softberry.com).

Single nucleotide polymorphism arrays

Genomic DNA (500 ng) was analysed using Infinium™ HumanCytoSNP-12 v2.1 300K BeadChip arrays (Illumina). After processing data in GenomeStudio (Illumina), SNP allelic ratios and copy number intensity were imported into Nexus Copy Number v8.0 (BioDiscovery, CA, USA). Data were normalised to diploid regions and segmented using the SNP-FASST2 segmentation algorithm, with a P -value of $<0.5 \times 10^{-8}$ and a minimum number of probes of 10. Copy number gain and loss were called at \log_2 ratios of +0.1 and -0.1, respectively. Allelic imbalance and loss of heterozygosity were called at B allele frequencies of 0.4 and 0.8, respectively. Concordance between cell line pairs was calculated as the percentage of base pairs with a different copy number state divided by the total number of base pairs included in the segmentation analysis.

Histology

Tissues (one large lobe from liver and lung per mouse) were fixed overnight using 10% neutral buffered formalin and then sectioned (3 μ m) for haematoxylin and eosin staining. Slides were scanned using an Aperio ScanScope with ImageScope software (Leica Microsystems, Macquarie Park, NSW, Australia) or an Olympus VS120 microscope and scanner with OlyVIA software (Olympus, Notting Hill, Victoria, Australia).

3D *in vitro* proliferation assays

For 3D assays, cells (2000/well) were seeded onto bovine type I collagen (2.5 mg/ml) gels (PureCol, Advanced BioMatrix, Carlsbad, CA, USA) prepared in white 96-well tissue culture plates. The luciferase-based Cell Titer Glo protocol (Promega) was used in accordance with the manufacturer's instructions and a 1.0 s luminometer integration time.

In vitro analysis of cell migration

Transwell migration and invasion assays were conducted as described previously (Johnstone et al., 2015). Briefly, 2×10^5 cells were seeded into the upper chamber of Fluoroblok (8 μ m membrane pore size) inserts (Corning Life Sciences, Oneonta, NY, USA) in serum-free medium. Cells were allowed to migrate for 6.5 h toward 10% serum-containing medium (700 μ l) in the base of the unit. Inserts were incubated with 4 μ g/ml calcein AM (Enzo Biochem Inc., New York, NY, USA) in PBS prior to imaging and counting of cells.

Monitoring of tumour growth and metastasis *in vivo*

NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) immunodeficient mice (*Mus musculus*), obtained from Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, USA) were maintained in a specific pathogen-free environment and fed *ad libitum*. All procedures involving mice conformed to National Health and Medical Research Council animal ethics guidelines and were approved by the Animal Experimentation and Ethics Committee (AEEC) of the Peter MacCallum Cancer Centre. Mammary tumours were established orthotopically in the right-side inguinal mammary gland of 6- to 8-week-old female mice, and monitored for growth and surgically resected as described previously (Johnstone et al., 2015). Briefly, viable tumour cells resuspended in PBS were mixed with 10-30% Cultrex (Trevigen) and inoculated (1×10^6 cells) into the right-side inguinal mammary gland of female mice and then surgically resected at a size of 200-400 mm³. Tumour size was measured using callipers and tumour volumes were estimated by the modified ellipsoidal formula: volume = $1/2(\text{length} \times \text{width}^2)$ (Tomayko and Reynolds, 1989). Differences in primary tumour growth rates were calculated using mixed-effects linear regression modelling as previously described (Johnstone et al., 2015). Primary tumour growth and metastasis were also monitored by *in vivo* bioluminescence imaging as described previously (Johnstone et al., 2008) using intraperitoneal injection of *in vivo* grade D-luciferin substrate (VivoGlo, Promega) and a Xenogen Lumina II *in vivo* imaging system (Caliper Life Sciences, Perkin Elmer, Hopkinton, MA, USA).

Acknowledgements

We thank the Animal Core Facility, Flow Cytometry Core Facility and Centre for Advanced Histology and Microscopy (CAHM) at the Peter MacCallum Cancer Centre for services provided, and the Adelaide Microarray Centre for running the

Illumina SNP arrays. We also thank Sue Sturrock and Stephen Fox (Anatomical Pathology, Peter Mac) for immunohistochemistry services. Joan Massague (Memorial Sloan Kettering Cancer Center) provided the MDA-MB-231_LM2 cell line, and Z. M. Shao and Z. L. Ou (Breast Cancer Institute, Fudan University, Shanghai) provided the MDA-MB-231HM cell line. We thank Judy Doherty (Peter Mac) for generation of the 231_LNA cell line.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: C.N.J., T.H.B.; Methodology: C.N.J., A.D.P., K.L.G., P.F.H., D.R.P., D.B., T.H.B.; Software: A.D.P., P.F.H., D.R.P., P.L., T.H.B.; Validation: C.N.J., P.F.H., P.L.; Formal analysis: C.N.J., A.D.P., K.L.G., P.F.H., R.L.A.; Investigation: C.N.J., R.L.A.; Resources: C.N.J., P.L., D.B., T.H.B., R.L.A.; Data curation: C.N.J., A.D.P., K.L.G., T.H.B.; Writing - original draft: C.N.J.; Writing - review & editing: C.N.J., A.D.P., K.L.G., M.E., A.G.S., T.H.B., R.L.A.; Visualization: K.L.G., P.L.; Supervision: C.N.J., P.F.H., A.G.S., T.H.B., R.L.A.; Project administration: C.N.J., M.E., A.G.S., T.H.B.; Funding acquisition: C.N.J., M.E., A.G.S., T.H.B., R.L.A.

Funding

This study was supported by a Cass Foundation of Victoria (Australia) Grant # SM/14/5567 (to C.N.J.) and National Health and Medical Research Council (Australia) project grants (APP1128250 and APP1042848) to T.H.B. The Australian Research Council also supported this work.

Data availability

The RNA-Seq data are available through the NCBI Gene Expression Omnibus (GEO) with accession number GSE101745. The SNP array data are available through the NCBI GEO with accession number GSE111705.

Supplementary information

Supplementary information available online at <http://dmm.biologists.org/lookup/doi/10.1242/dmm.032250.supplemental>

References

- Ai, L., Kim, W.-J., Kim, T.-Y., Fields, C. R., Massoll, N. A., Robertson, K. D. and Brown, K. D. (2006). Epigenetic silencing of the tumor suppressor cystatin M occurs during breast cancer progression. *Cancer Res.* **66**, 7899-7909.
- Aslakson, C. J. and Miller, F. R. (1992). Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res.* **52**, 1399-1405.
- Barbara, N. P., Wrana, J. L. and Letarte, M. (1999). Endoglin is an accessory protein that interacts with the signaling receptor complex of multiple members of the transforming growth factor-beta superfamily. *J. Biol. Chem.* **274**, 584-594.
- Bonapace, L., Coissieux, M.-M., Wyckoff, J., Mertz, K. D., Varga, Z., Junt, T. and Bentires-Alj, M. (2014). Cessation of CCL2 inhibition accelerates breast cancer metastasis by promoting angiogenesis. *Nature* **515**, 130-133.
- Cailleau, R., Young, R., Olivé, M. and Reeves, W. J., Jr. (1974). Breast tumor cell lines from pleural effusions. *J. Natl. Cancer Inst.* **53**, 661-674.
- Cameron, E. E., Bachman, K. E., Myöhänen, S., Herman, J. G. and Baylin, S. B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat. Genet.* **21**, 103-107.
- Cao, Y., Slaney, C. Y., Bidwell, B. N., Parker, B. S., Johnstone, C. N., Rautela, J., Eckhardt, B. L. and Anderson, R. L. (2014). BMP4 inhibits breast cancer metastasis by blocking myeloid-derived suppressor cell activity. *Cancer Res.* **74**, 5091-5102.
- Capes-Davis, A., Reid, Y. A., Kline, M. C., Storts, D. R., Strauss, E., Dirks, W. G., Drexler, H. G., MacLeod, R. A. F., Sykes, G., Kohara, A. et al. (2013). Match criteria for human cell line authentication: where do we draw the line? *Int. J. Cancer* **132**, 2510-2519.
- Carey, L. A., Dees, E. C., Sawyer, L., Gatti, L., Moore, D. T., Collichio, F., Ollila, D. W., Sartor, C. I., Graham, M. L. and Perou, C. M. (2007). The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin. Cancer Res.* **13**, 2329-2334.
- Carmona, F. J., Davalos, V., Vidal, E., Gomez, A., Heyn, H., Hashimoto, Y., Vizoso, M., Martinez-Cardus, A., Sayols, S., Ferreira, H. J. et al. (2014). A comprehensive DNA methylation profile of epithelial-to-mesenchymal transition. *Cancer Res.* **74**, 5608-5619.
- Chaffer, C. L., San Juan, B. P., Lim, E. and Weinberg, R. A. (2016). EMT, cell plasticity and metastasis. *Cancer Metastasis Rev.* **35**, 645-654.
- Chang, X.-Z., Li, D.-Q., Hou, Y.-F., Wu, J., Lu, J.-S., Di, G.-H., Jin, W., Ou, Z.-L., Shen, Z.-Z. and Shao, Z.-M. (2007). Identification of the functional role of peroxiredoxin 6 in the progression of breast cancer. *Breast Cancer Res.* **9**, R76.

- Chang, X.-Z., Li, D.-Q., Hou, Y.-F., Wu, J., Lu, J.-S., Di, G.-H., Jin, W., Ou, Z.-L., Shen, Z.-Z. and Shao, Z.-M. (2008). Identification of the functional role of AF1Q in the progression of breast cancer. *Breast Cancer Res. Treat.* **111**, 65-78.
- Chang, A. C.-M., Doherty, J., Huscchtscha, L. I., Redvers, R., Restall, C., Reddel, R. R. and Anderson, R. L. (2015). STC1 expression is associated with tumor growth and metastasis in breast cancer. *Clin. Exp. Metastasis* **32**, 15-27.
- Clement, J. H., Raida, M., Sanger, J., Bicknell, R., Liu, J., Naumann, A., Geyer, A., Waldau, A., Hortschansky, P., Schmidt, A. et al. (2005). Bone morphogenetic protein 2 (BMP-2) induces in vitro invasion and in vivo hormone independent growth of breast carcinoma cells. *Int. J. Oncol.* **27**, 401-407.
- Cox, J. L. (2009). Cystatins and cancer. *Front. Biosci.* **14**, 463-474.
- Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y. et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352.
- Dziegiel, P., Pula, B., Kobierzycki, C., Stasiolek, M. and Podhorska-Okolow, M. (2016). Metallothioneins in normal and cancer cells. *Adv. Anat. Embryol. Cell Biol.* **218**, 1-117.
- Eckhardt, B. L., Parker, B. S., van Laar, R. K., Restall, C. M., Natoli, A. L., Tavaría, M. D., Stanley, K. L., Sloan, E. K., Moseley, J. M. and Anderson, R. L. (2005). Genomic analysis of a spontaneous model of breast cancer metastasis to bone reveals a role for the extracellular matrix. *Mol. Cancer Res.* **3**, 1-13.
- Eckhardt, B. L., Francis, P. A., Parker, B. S. and Anderson, R. L. (2012). Strategies for the discovery and development of therapies for metastatic breast cancer. *Nat. Rev. Drug Discov.* **11**, 479-497.
- Fang, F., Turcan, S., Rimmer, A., Kaufman, A., Giri, D., Morris, L. G., Shen, R., Seshan, V., Mo, Q., Heguy, A. et al. (2011). Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci. Transl. Med.* **3**, 75ra25.
- Ferrara, N. (2002). VEGF and the quest for tumour angiogenesis factors. *Nat. Rev. Cancer* **2**, 795-803.
- Fietz, E. R., Keenan, C. R., López-Campos, G., Tu, Y., Johnstone, C. N., Harris, T. and Stewart, A. G. (2017). Glucocorticoid resistance of migration and gene expression in a daughter MDA-MB-231 breast tumour cell line selected for high metastatic potential. *Sci. Rep.* **7**, 43774.
- Halldórsson, S., Rohatgi, N., Magnúsdóttir, M., Choudhary, K. S., Gudjonsson, T., Knutsen, E., Barkovskaya, A., Hilmarsson, B., Perander, M., Maelandsmo, G. M. et al. (2017). Metabolic re-wiring of isogenic breast epithelial cell lines following epithelial to mesenchymal transition. *Cancer Lett.* **396**, 117-129.
- Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* **144**, 646-674.
- Harrison, P. F., Powell, D. R., Clancy, J. L., Preiss, T., Boag, P. R., Traven, A., Seemann, T. and Beilharz, T. H. (2015). PAT-seq: a method to study the integration of 3'-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA* **21**, 1502-1510.
- Hartmann, O., Spyrtatos, F., Harbeck, N., Dietrich, D., Fassbender, A., Schmitt, M., Eppenberger-Castori, S., Vuoroqueaux, V., Lerebours, F., Welzel, K. et al. (2009). DNA methylation markers predict outcome in node-positive, estrogen receptor-positive breast cancer with adjuvant anthracycline-based chemotherapy. *Clin. Cancer Res.* **15**, 315-323.
- Hennessy, B. T., Gonzalez-Angulo, A.-M., Stenke-Hale, K., Gilcrease, M. Z., Krishnamurthy, S., Lee, J.-S., Fridlyand, J., Sahin, A., Agarwal, R., Joy, C. et al. (2009). Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res.* **69**, 4116-4124.
- Iorns, E., Drews-Elger, K., Ward, T. M., Dean, S., Clarke, J., Berry, D., El Ashry, D. and Lippman, M. (2012). A new mouse model for the study of human breast cancer metastasis. *PLoS ONE* **7**, e47995.
- Jin, W., Chen, B.-B., Li, J.-Y., Zhu, H., Huang, M., Gu, S.-M., Wang, Q.-Q., Chen, J.-Y., Yu, S., Wu, J. et al. (2012). TIEG1 inhibits breast cancer invasion and metastasis by inhibition of epidermal growth factor receptor (EGFR) transcription and the EGFR signaling pathway. *Mol. Cell. Biol.* **32**, 50-63.
- Johnstone, C. N., Mongroo, P. S., Rich, A. S., Schupp, M., Bowser, M. J., Delemos, A. S., Tobias, J. W., Liu, Y., Hannigan, G. E. and Rustgi, A. K. (2008). Parvin-beta inhibits breast cancer tumorigenicity and promotes CDK9-mediated peroxisome proliferator-activated receptor gamma 1 phosphorylation. *Mol. Cell. Biol.* **28**, 687-704.
- Johnstone, C. N., Smith, Y. E., Cao, Y., Burrows, A. D., Cross, R. S. N., Ling, X., Redvers, R. P., Doherty, J. P., Eckhardt, B. L., Natoli, A. L. et al. (2015). Functional and molecular characterisation of EO771.LMB tumours, a new C57BL/6-mouse-derived model of spontaneously metastatic mammary cancer. *Dis. Model. Mech.* **8**, 237-251.
- Kioulafra, M., Balkouranidou, I., Sotiropoulou, G., Kakkamanis, L., Mavroudis, D., Georgoulas, V. and Lianidou, E. S. (2009). Methylation of cystatin M promoter is associated with unfavorable prognosis in operable breast cancer. *Int. J. Cancer* **125**, 2887-2892.
- Kitamura, T., Qian, B.-Z., Soong, D., Cassetta, L., Noy, R., Sugano, G., Kato, Y., Li, J. and Pollard, J. W. (2015). CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. *J. Exp. Med.* **212**, 1043-1059.
- Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-veizer, J., McMichael, J. F., Fulton, L. L., Dooling, D. J., Ding, L., Mardis, E. R. et al. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70.
- Le, C. P., Nowell, C. J., Kim-Fuchs, C., Botteri, E., Hiller, J. G., Ismail, H., Pimentel, M. A., Chai, M. G., Karnezis, T., Rotmensz, N. et al. (2016). Chronic stress in mice remodels lymph vasculature to promote tumour cell dissemination. *Nat. Commun.* **7**, 10634.
- Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y. and Pietenpol, J. A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750-2767.
- Lehmann, B. D., Jovanović, B., Chen, X., Estrada, M. V., Johnson, K. N., Shyr, Y., Moses, H. L., Sanders, M. E. and Pietenpol, J. A. (2016). Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS ONE* **11**, e0157368.
- Liedtke, C., Mazouni, C., Hess, K. R., André, F., Tordai, A., Mejia, J. A., Symmans, W. F., Gonzalez-Angulo, A. M., Hennessy, B., Green, M. et al. (2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J. Clin. Oncol.* **26**, 1275-1281.
- Low-Marchelli, J. M., Ardi, V. C., Vizcarra, E. A., van Rooijen, N., Quigley, J. P. and Yang, J. (2013). Twist1 induces CCL2 and recruits macrophages to promote angiogenesis. *Cancer Res.* **73**, 662-671.
- Lu, X. and Kang, Y. (2009). Chemokine (C-C motif) ligand 2 engages CCR2+ stromal cells of monocytic origin to promote breast cancer metastasis to lung and bone. *J. Biol. Chem.* **284**, 29087-29096.
- Minn, A. J., Gupta, G. P., Siegel, P. M., Bos, P. D., Shu, W., Giri, D. D., Viale, A., Olshen, A. B., Gerald, W. L. and Massagué, J. (2005a). Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518-524.
- Minn, A. J., Kang, Y., Serganova, I., Gupta, G. P., Giri, D. D., Doubrovin, M., Ponomarev, V., Gerald, W. L., Blasberg, R. and Massagué, J. (2005b). Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **115**, 44-55.
- Minn, A. J., Gupta, G. P., Padua, D., Bos, P., Nguyen, D. X., Nuyten, D., Kreike, B., Zhang, Y., Wang, Y., Ishwaran, H. et al. (2007). Lung metastasis genes couple breast tumor size and metastatic spread. *Proc. Natl. Acad. Sci. USA* **104**, 6740-6745.
- Mizukami, Y., Jo, W.-S., Duerr, E.-M., Gala, M., Li, J., Zhang, X., Zimmer, M. A., Iliopoulos, O., Zukerberg, L. R., Kohgo, Y. et al. (2005). Induction of interleukin-8 preserves the angiogenic response in HIF-1alpha-deficient colon cancer cells. *Nat. Med.* **11**, 992-997.
- Mongroo, P. S., Johnstone, C. N., Naruszewicz, I., Leung-Hageteijn, C., Sung, R. K., Carnio, L., Rustgi, A. K. and Hannigan, G. E. (2004). Beta-parvin inhibits integrin-linked kinase signaling and is downregulated in breast cancer. *Oncogene* **23**, 8959-8970.
- Morris, V. L., Tuck, A. B., Wilson, S. M., Percy, D. and Chambers, A. F. (1993). Tumor progression and metastasis in murine D2 hyperplastic alveolar nodule mammary tumor cell lines. *Clin. Exp. Metastasis* **11**, 103-112.
- Narod, S. A. and Sopik, V. (2018). Is invasion a necessary step for metastases in breast cancer? *Breast Cancer Res. Treat.* **169**, 9-23.
- Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F. et al. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-527.
- Nguyen, D. X., Bos, P. D. and Massagué, J. (2009). Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* **9**, 274-284.
- Nielsen, T. O., Hsu, F. D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., Hernandez-Boussard, T., Livasy, C., Cowan, D., Dressler, L. et al. (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin. Cancer Res.* **10**, 5367-5374.
- Orsolich, I., Jurada, D., Pullen, N., Oren, M., Eliopoulos, A. G. and Volarevic, S. (2016). The relationship between the nucleus and cancer: current evidence and emerging paradigms. *Semin. Cancer Biol.* **37-38**, 36-50.
- Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A. et al. (2000). Molecular portraits of human breast tumours. *Nature* **406**, 747-752.
- Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., He, X. and Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68.
- Prat, A., Adamo, B., Cheang, M. C. U., Anders, C. K., Carey, L. A. and Perou, C. M. (2013). Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *Oncologist* **18**, 123-133.
- Puchalapalli, M., Zeng, X., Mu, L., Anderson, A., Hix Glickman, L., Zhang, M., Sayyad, M. R., Mosticone Wangenstein, S., Clevenger, C. V. and Koblinski, J. E. (2016). NSG mice provide a better spontaneous model of breast cancer metastasis than athymic (Nude) mice. *PLoS ONE* **11**, e0163521.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47.

- Ruffell, B., Affara, N. I., Cottone, L., Junankar, S., Johansson, M., DeNardo, D. G., Korets, L., Reinheckel, T., Sloane, B. F., Bogyo, M. et al. (2013). Cathepsin C is a tissue-specific regulator of squamous carcinogenesis. *Genes Dev.* **27**, 2086-2098.
- Santner, S. J., Dawson, P. J., Tait, L., Soule, H. D., Eliason, J., Mohamed, A. N., Wolman, S. R., Heppner, G. H. and Miller, F. R. (2001). Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast Cancer Res. Treat.* **65**, 101-110.
- Satya-Prakash, K. L., Pathak, S., Hsu, T. C., Olivé, M. and Cailleau, R. (1981). Cytogenetic analysis on eight human breast tumor cell lines: high frequencies of 1q, 11q and HeLa-like marker chromosomes. *Cancer Genet. Cytogenet.* **3**, 61-73.
- Seftor, R. E. B., Hess, A. R., Seftor, E. A., Kirschmann, D. A., Hardy, K. M., Margaryan, N. V. and Hendrix, M. J. C. (2012). Tumor cell vasculogenic mimicry: from controversy to therapeutic promise. *Am. J. Pathol.* **181**, 1115-1125.
- Shaner, N. C., Steinbach, P. A. and Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905-909.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S. et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869-10874.
- Steelman, L. S., Martelli, A. M., Cocco, L., Libra, M., Nicoletti, F., Abrams, S. L. and McCubrey, J. A. (2016). The therapeutic potential of mTOR inhibitors in breast cancer. *Br. J. Clin. Pharmacol.* **82**, 1189-1212.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545-15550.
- Swift, S., Lorens, J., Achacoso, P. and Nolan, G. P. (2001). Rapid Production of Retroviruses for Efficient Gene Delivery to Mammalian Cells Using 293T Cell-Based Systems. *Curr. Protoc. Immunol.* Chapter 10: Unit 10.17C.
- Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., Hartwell, K., Onder, T. T., Gupta, P. B., Evans, K. W. et al. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. USA* **107**, 15449-15454.
- Tomayko, M. M. and Reynolds, C. P. (1989). Determination of subcutaneous tumor size in athymic (nude) mice. *Cancer Chemother. Pharmacol.* **24**, 148-154.
- Turner, N. C. and Reis-Filho, J. S. (2013). Tackling the diversity of triple-negative breast cancer. *Clin. Cancer Res.* **19**, 6380-6388.
- van Riggelen, J., Yetil, A. and Felsher, D. W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat. Rev. Cancer* **10**, 301-309.
- Wagenblast, E., Soto, M., Gutiérrez-Ángel, S., Hartl, C. A., Gable, A. L., Maceli, A. R., Erard, N., Williams, A. M., Kim, S. Y., Dickopf, S. et al. (2015). A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* **520**, 358-362.
- Weaver, V. M., Howlett, A. R., Langton-Webster, B., Petersen, O. W. and Bissell, M. J. (1995). The development of a functionally relevant cell culture model of progressive human breast cancer. *Semin. Cancer Biol.* **6**, 175-184.
- Wu, D. and Smyth, G. K. (2012). Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133.
- Zhang, J., Shridhar, R., Dai, Q., Song, J., Barlow, S. C., Yin, L., Sloane, B. F., Miller, F. R., Meschonat, C., Li, B. D. L. et al. (2004). Cystatin m: a novel candidate tumor suppressor gene for breast cancer. *Cancer Res.* **64**, 6957-6964.