2015

A cell–ECM screening method to predict breast cancer metastasis

Lauren E. Barney
University of Massachusetts Amherst

E. C. Dandley
North Carolina State University

Lauren Jansen
University of Massachusetts Amherst

Nicholas G. Reich
University of Massachusetts Amherst

A. M. Mercurio
University of Massachusetts Medical School

See next page for additional authors

Recommended Citation

Barney, Lauren E.; Dandley, E. C.; Jansen, Lauren; Reich, Nicholas G.; Mercurio, A. M.; and Peyton, Shelly, "A cell–ECM screening method to predict breast cancer metastasis" (2015). *Integrative Biology*. 844.
https://doi.org/10.1039/c4ib00218k

A cell–ECM screening method to predict breast cancer metastasis†

L. E. Barney,[a] E. C. Dandley,[b] L. E. Jansen,[a] N. G. Reich,[c] A. M. Mercurio[d] and S. R. Peyton*[a]

Breast cancer preferentially spreads to the bone, brain, liver, and lung. The clinical patterns of this tissue-specific spread (tropism) cannot be explained by blood flow alone, yet our understanding of what mediates tropism to these physically and chemically diverse tissues is limited. While the microenvironment has been recognized as a critical factor in governing metastatic colonization, the role of the extracellular matrix (ECM) in mediating tropism has not been thoroughly explored. We created a simple biomaterial platform with systematic control over the ECM protein density and composition to determine if integrin binding governs how metastatic cells differentiate between secondary tissue sites. Instead of examining individual behaviors, we compiled large patterns of phenotypes associated with adhesion to and migration on these controlled ECMs. In combining this novel analysis with a simple biomaterial platform, we created an *in vitro* fingerprint that is predictive of *in vivo* metastasis. This rapid biomaterial screen also provided information on how β_1, α_2, and α_6 integrins might mediate metastasis in patients, providing insights beyond a purely genetic analysis. We propose that this approach of screening many cell–ECM interactions, across many different heterogeneous cell lines, is predictive of *in vivo* behavior, and is much simpler, faster, and more economical than complex 3D environments or mouse models. We also propose that when specifically applied toward the question of tissue tropism in breast cancer, it can be used to provide insight into certain integrin subunits as therapeutic targets.

Insight, innovation, integration

We developed a high-throughput method to rapidly screen cell adhesion, motility, and growth factor responses on biomaterial surfaces. This approach is analogous to systems biology, relying on cell phenotypes in lieu of genetics. We used this technique to reveal patterns of phenotypes associated with breast cancer metastasis to possible tissue sites (bone, brain, lung). By comparing the phenotypic patterns between cell lines that metastasize to only one tissue site with heterogeneous cell lines, we provide the first method to connect *in vitro* phenotype to *in vivo* fate. This method is successful without genetic analysis, yet it also predicts outcomes related to integrin gene expression, potentially identifying new targets for tissue-specific metastasis.

Introduction

Breast cancer is the most common cancer in women, and metastasis is responsible for 90% of all cancer deaths. The microenvironment is a critical regulator of metastasis,[1] and *in vivo* studies have provided insight into many microenvironment-mediated mechanisms.[2–6] However, these studies lack features of human physiology, contain uncontrolled variables, and are often not comparable across different mouse models. Thus, it is difficult to parse the varying contributions of each factor, limiting the broad applicability of these results. In contrast, *in vitro* models can be either oversimplified,[7] or highly complex, expensive, low-throughput, and limited to highly specialized laboratories.[8] Because metastasis remains both largely incurable and poorly understood, there is a need for quick, cost-efficient *in vitro* model systems with enough complexity to recapitulate certain aspects of *in vivo* biology, while maintaining affordability and efficiency.

Two-dimensional *in vitro* disease models are appropriately low cost and simple, however, it is now generally appreciated

[a] *Department of Chemical Engineering, University of Massachusetts, Amherst, MA 01003, USA. E-mail: speyton@ecs.umass.edu*
[b] *Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695, USA*
[c] *Division of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA 01003, USA*
[d] *Department of Cancer Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4ib00218k

that two-dimensional cell behaviors are usually not conserved in a three-dimensional context. One recent example of this is work by Meyer *et al.*, which showed that only two-dimensional growth-factor induced membrane protrusion, but not any other measured 2D motility parameter, accurately predicts 3D motility.[9] This suggests that individual two-dimensional measurements will also not be predictive of even more complex *in vivo* cell phenotypes. We suggest that a better approach may be to avoid measurements of single, likely non-predictive metrics, and instead, measure patterns of many phenotypes across several conditions and several cell sources.[10]

Breast cancer metastasis is a clear candidate for this type of approach because of its striking, yet unexplained, clinical patterns of metastatic spread (tropism) to the bone, brain, liver, and lung, but not to tissues such as the skin, heart, uterus, or spleen.[11,12] This tropism is hypothesized to depend upon an unknown relationship between metastatic cells (the seeds) and hospitable microenvironments (the soils).[12] Because the tissues often colonized by breast cancer cells each have a distinct ECM, we posit that integrin binding to the ECM is one feature that plays a critical role in the early stages of tissue-specific colonization and fate of extravasated cancer cells. It is known that secondary site colonization requires activation of integrin-mediated signaling,[10,13–16] and several individual integrins have been implicated in breast cancer metastasis *in vivo*.[17–20] However, this research is limited to metastasis at a single tissue site, or to the effect of a single integrin, which is not representative of the variety of cell–matrix interactions simultaneously occurring *in vivo* during cell adhesion to a secondary tissue site post-extravasation.

To this end, we created a novel biomaterial platform comprised of complex ECMs that can present any combination of full-length proteins with high fidelity, reproducibility, and is permanent during the cell culture period. We used this biomaterial to quantify how different metastatic breast cancer cell lines differentiate between secondary sites *via* integrin binding. We compiled sixty-six distinct cell phenotypes associated with cell adhesion and motility, both with and without epidermal growth factor (EGF) stimulation. By collating all the responses to each biomaterial ECM surface together, we created a phenotypic *fingerprint* of bone, brain, and lung metastasis, which is both capable of distinguishing between genetically related tropic cell subpopulations, as well as predictive of the *in vivo* metastasis of several other more heterogeneous cell lines, thereby predicting *in vivo* fate with a quick *in vitro* screen. Taken together, we propose this comprehensive analysis of cell–matrix interactions as a tool to predict *in vivo* fate, as well as understand the roles of integrins in tropism, thus providing insight toward integrins as druggable targets for metastatic disease.

Results

Rapid biomaterial screening of cell–ECM interactions

The ECMs present in secondary tissues often recipient of breast cancer colonization are each strikingly complex and distinct,[21] suggesting that cell–matrix interactions play a role in mediating metastasis. Our goal here was to create a simple *in vitro* biomaterial system to investigate the role of integrin binding in breast cancer tropism. These biomaterial surfaces are glass coverslips, modified *via* straightforward silane chemistry. They present covalently coupled ECM proteins, and they remain stable throughout the cell culture period (at least 72 hours). This method can be used to couple any combination of full-length proteins or peptides of interest, providing control over cell–matrix interactions in a highly robust (Fig. S1, ESI†), scalable, and facile process. It is particularly well suited to observe integrin-mediated phenotypes of adherent cell lines, and allows for functional investigation of real-time binding to the ECM.

We used this approach to create three ECM microenvironments containing combinations of ECM proteins inspired by the biochemical cues found at the *in vivo* tissues to which breast cancer commonly spreads (Table 1). ECM at these tissues has significant patient-to-patient variability,[21] and can be remodelled by both tumor and stromal cells over time.[5,6] Our approach was to determine if integrin binding alone, at early time points (directly after extravasation), was sufficient to direct or predict tissue-specific spread. Therefore, we created three biomaterial surfaces containing combinations of ECM proteins that isolate distinct integrin heterodimers, independently of the *in vivo* heterogeneity, inspired by the ECM protein content of healthy tissues at these sites.[6,22–26]

Table 1 Composition of tissue-inspired biomaterial ECMs

	ECM 1 (bone)	ECM 2 (brain)	ECM 3 (lung)
In vivo ECM density	High	Low	Moderate
In vivo ECM proteins	Collagen I Fibronectin[22] Osteopontin[23]	Laminin fibronectin collagen IV tenascins proteoglycans[24] Hyaluronic acid[25]	Fibronectin[6] Collagens laminins elastin tenascin C proteoglycans[26]
In vitro ECM density	5 μg cm^{-2}	1 μg cm^{-2}	2 μg cm^{-2}
In vitro ECM proteins	99% collagen I 1% osteopontin	50% fibronectin 25% vitronectin 20% tenascin C 5% laminin	33% laminin 33% collagen IV 15% collagen I 15% fibronectin 4% tenascin C

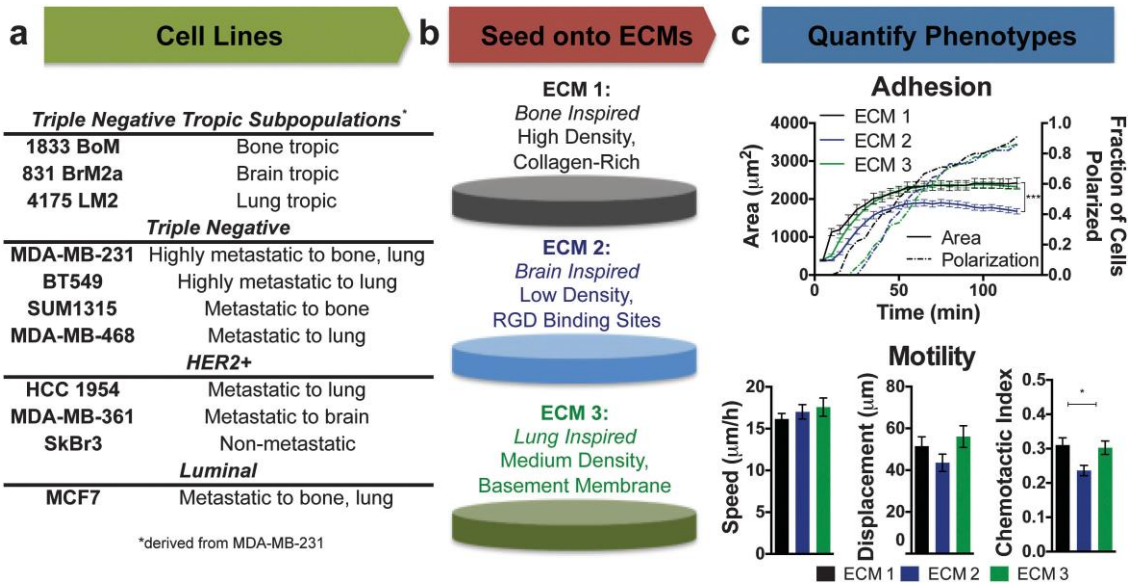

Fig. 1 Biomaterial platform for integrin-mediated phenotyping. (a) Breast cancer cell lines with their known *in vivo* metastatic tropisms.[2–4,28–33] (b) Three distinct ECM microenvironments regulate integrin binding. (c) Adhesion and motility phenotypes of the MDA-MB-231 cell line. Black: ECM 1; blue: ECM 2; green: ECM 3.

As a first attempt to capture the heterogeneity of breast cancer, we screened phenotypes associated with adhesion and motility across a large panel of human cell lines that span the clinical subtypes[27] and metastatic specificity to the bone, brain, and lung sites[2–4,28–33] (Fig. 1a). We first measured spreading and polarization of cells during initial adhesion to the ECMs, quantifying time-dependent cell area, spreading rate, and polarization of the population (Fig. 1c top and Fig. S2a–c, ESI†). Next, we observed long-term motility, and quantified cell migration speed, displacement, and chemotactic index (ratio of displacement to path length, quantifies migration straightness) (Fig. 1c bottom, Fig. S2d–f, ESI†). Altogether, 550 individual cell observations were made per cell line on average (Fig. 1c and Fig. S2, ESI†). In combination with an ELISA characterization of the protein coupling (Fig. S1, ESI†), these measurements validate the biomaterial platform as being able to elicit differential cell line responses *via* integrin binding alone (Fig. S2, ESI†).

Phenotypic screen predicts tissue tropism *in vivo*

We hypothesized that an *in vitro* analysis of integrin-mediated phenotypes would be capable of differentiating cell lines with different tropisms *in vivo*, reflecting the functional binding interactions required for successful metastatic outgrowth. We began by phenotyping three highly tropic subpopulations of the MDA-MB-231 parental cell line, which were selected from *in vivo* metastases by Massagué and colleagues.[2–4] These geneti- cally distinct MDA-MB-231 subpopulations each display strong tropism to either the bone (“bone tropic”),[2] brain (“brain tropic”),[4] or lung (“lung tropic”)[3] in mice, and have been used to identify genetic determinants of tissue-specific metastasis. We found that these tissue-specific cell lines each responded to the ECMs in unique ways (Fig. S3, ESI†).

We immediately noticed that for each different measurement we took, the tropic cell lines were sensitive to the ECMs in some cases, while insensitive in others. As an example, when we quantified cell speed, the bone and brain tropic cell lines had very different cell speeds on each ECM surface, whereas the lung tropic cells had similar cell speeds on all three ECMs. To quantifiably demonstrate this, we applied a statistical tool called the coefficient of variation (CoV) to each set of measurements for a given cell line. This measurement quantifies dispersion of a data set, and is computed by dividing the standard deviation of a set of measurements by their mean. In its application here, higher CoV values identify adhesion or motility measurements where a cell line is highly sensitive to the different ECM surfaces (*i.e.*, the relative standard deviation of the measurements across each ECM is at least 10%). Because these cell lines were created to exhibit strong *in vivo* metastasis to one site, we first wanted to validate that this *in vivo* selectivity was reflected in the *in vitro* cell–ECM screen. This CoV analysis showed that each of the tropic cell subpopulations was significantly more sensitive to the ECMs than the parental cell line in each of the adhesion and motility phenotypes we quantified (Fig. S4a, ESI†), validating this approach.

Given that tropic cell lines have heightened sensitivity to growth factors,[34] we then quantified the change in each adhesion and motility metric in response to EGF stimulation, and found that the patterns of adhesion and motility across the ECMs under normal and EGF-stimulated conditions varied significantly between the tropic cell lines (Fig. S3, ESI†). The effect of EGF stimulation was especially striking in the tropic cell adhesion. EGF slowed spreading and polarization of the bone tropic cell line, while it increased the spreading rate of the brain and lung tropic cell lines in an ECM-dependent fashion (Fig. S3a–f, ESI†). Growth factor sequestration and growth

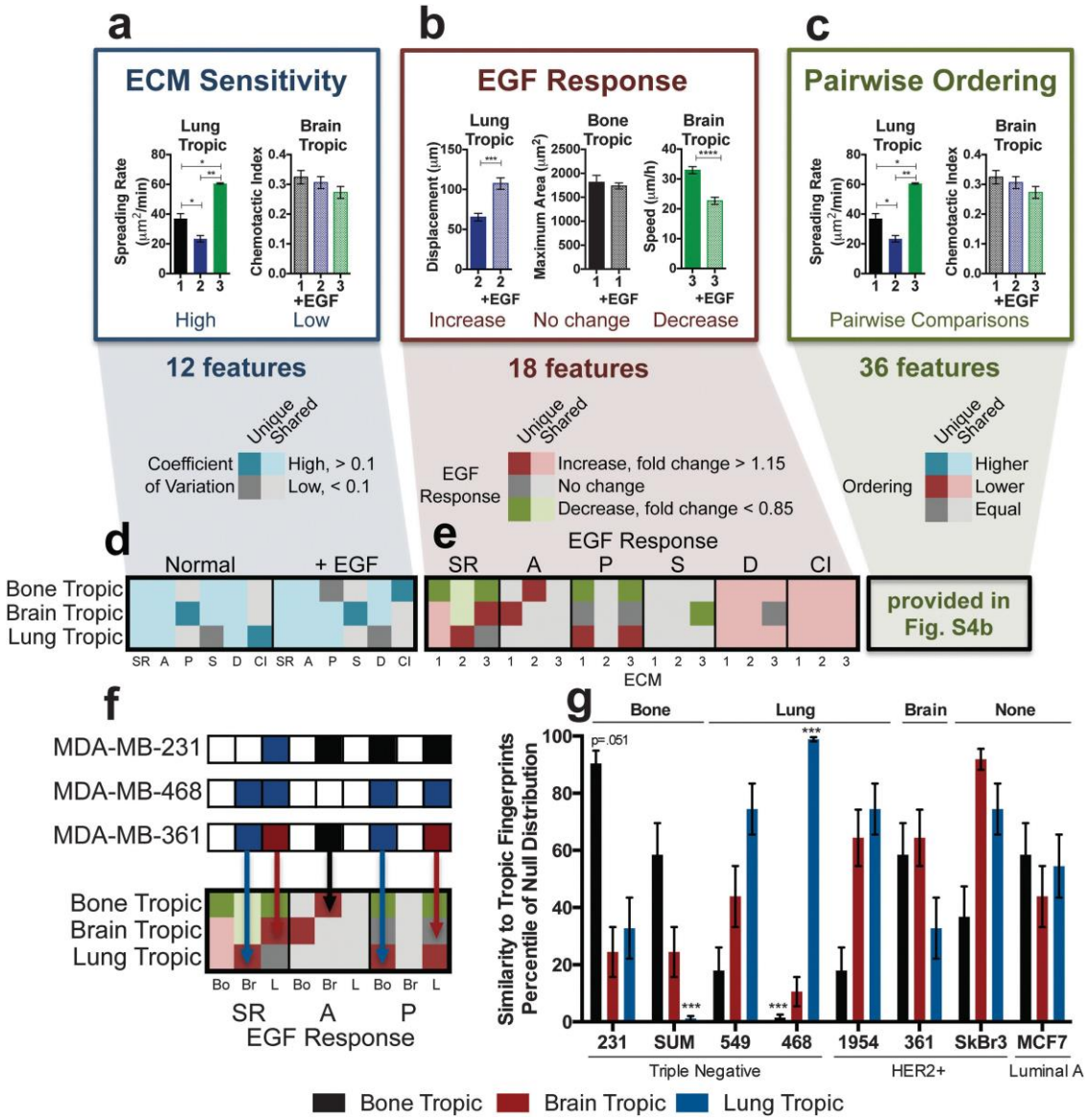

Fig. 2 Phenotypic fingerprint predicts bone, brain, and lung metastasis. Development of phenotypic fingerprints of bone, brain, and lung tropism from patterns of integrin-mediated behavior. The (a) CoV, (b) EGF response, and (c) measurement ordering features were measured for each tropic cell line to create a fingerprint. (d, e) Phenotypic fingerprints of tropic behavior (rows), with unique identifiers distinguished in dark colors. (d) Coefficient of variation for both the normal and EGF-stimulated conditions. Blue: high CoV ($\geq$0.1); grey: low CoV ($<$0.1). (e) EGF response on each ECM. Red: increase (fold change $\geq$ 1.15); grey: no change; green: decrease (fold change $\leq$ 0.85). Pairwise ordering is provided in Fig. S4b (ESI†). Abbreviations: 1: ECM 1; 2: ECM 2; 3: ECM 3; SR: spreading rate; A: area; P: polarization; S: speed; D: displacement; CI: chemotactic index. (f) Fingerprint of heterogeneous cell lines overlaid onto tropic fingerprints allows for quantification of similarity to each tropic subpopulation. Colored boxes in overlaid cell lines identify features shared with each tropic cell line. Black: bone tropic, red: brain tropic, blue: lung tropic. (g) *In vitro* tropism of the MDA-MB-231, SUM1315 MO2, BT549, MDA-MB-468, HCC 1954, MDA-MB-361, SkBr3, and MCF7 cell lines. Black: bone tropic, red: brain tropic, blue: lung tropic. Bars represent the percentile of the null distribution where each heterogeneous cell line lies, with error bars displaying the range of the respective percentile. High percentiles indicate cell lines that are highly phenotypically tropic to a tissue site. Statistics shown above bars indicate that the cell line is significantly higher or lower than the null distribution for the indicated tropism. Top labels: known *in vivo* tropism. Bottom labels: clinical subtype designation.

factor signalling both depend on the composition of the ECM,[35,36] which, in addition to the differences in the tropic subpopulations, is likely one reason for the differences in EGF responses.

With this collection of measurements, we sought to identify all the phenotypes unique to cells metastasizing to the bone, brain, or lung by compiling them *en masse* into a heat map. All the adhesion and motility measurements were organized across three separate classes of observable responses to the ECMs (coefficient of variation, response to EGF, and pairwise ordering across ECMs, Fig. 2a–c).

As the first feature of these collective tropic phenotypes, measurements in which the cell line was either highly sensitive or insensitive to changes in the ECM were identified (Fig. 2a).

For example, the lung tropic line has a high spreading rate CoV, whereas the brain tropic cell line has a low chemotactic index CoV under EGF stimulation (Fig. 2a). We then quantified whether each phenotype measured increased, decreased, or remained unchanged in response to EGF stimulation on each of the three ECMs (Fig. 2b). As examples, the lung tropic cells increased their displacement with EGF stimulation, the bone tropic cell line spreading area was unchanged, and the brain tropic cell speed decreased upon EGF stimulation (Fig. 2b). Finally, a pairwise statistical test was used to compare each phenotype measured across each of the ECM surfaces, resulting in an ''ECM ordering'' (Fig. 2c). As an example, the lung tropic cell line has the statistically highest spreading rate on ECM 3, followed by ECM 1, and is lowest on ECM 2 (Fig. 2c). In contrast, the chemotactic index of the brain tropic cell line was statistically equivalent on all ECMs.

We computed each of these features for all phenotypes and compiled this into a heat map, creating a row for each tropic cell line (Fig. 2d, e and Fig. S4b, ESI†). The color of each box identifies the value of the phenotype for the respective cell line. For the CoV measurement, highly sensitive measurements (CoV $\lhd$ 0.1) are identified with a blue box, while insensitive measurements (CoV $\rhd$ 0.1) are identified with a grey box (Fig. 2d). In Fig. 2e, EGF changes are determined *via* fold change from the normal measurement. An increase (fold change $\lhd$ 1.15) is red, a decrease (fold change $\rhd$ 0.85) is green, and no change (0.85 $\rhd$ fold change $\rhd$ 1.15) is grey (Fig. 2e). In Fig. S4b (ESI†), for each pairwise comparison, a blue box indicates that the first measurement is greater than the second, a grey box indicates that they are equal, and a red box indicates that the second measurement is greater than the first. Altogether, greater than 1000 individual cell observations per tropic cell line were quantified and compiled together to create sixty-six features of integrin-mediated phenotypes for each tropic cell subpopulation (Fig. 2d, e and Fig. S4b, ESI,† rows).

Looking down each column in Fig. 2d, e and Fig. S4b (ESI†), we then identified behaviors that were *specific* to only one of the three tropic cell lines. These instances are identified by a dark shaded box of the appropriate color. Looking across each row, thirteen measured phenotypes are unique to bone tropism, a separate thirteen features are unique to lung tropism, and fifteen to brain tropism (Fig. 2d, e and Fig. S4b, ESI†). Although no singular adhesion or motility-associated phenotype was predictive of breast cancer tropism (not shown), this collective analysis of many cell responses to ECMs combined to create phenotypic *fingerprints* of bone, brain, and lung metastasis (Fig. 2d, e and Fig. S4b, ESI†).

We repeated all these measurements using more heterogeneous cell lines with known, literature-reported metastasis *in vivo* (Fig. 1a), and overlaid each cell line's pattern of behaviors onto the three fingerprints of tissue-specific metastasis we created (Fig. 2f). The measurements from each cell line can be compared to the tissue-specific fingerprints generated from the tropic cell lines, resulting in a fractional similarity value that represents the amount of unique phenotypic features shared with the cell lines that specifically metastasize to either the bone, brain, or lung (Fig. S5a, ESI†). As a control, a null distribution of 10 000 cell lines with random phenotypes was generated *in silico* to represent ''random metastasis'' (Fig. S5a–d, ESI†). We determined how similar the patterns were between each of the heterogeneous cell lines and the bone, brain, and lung tropic fingerprints (Fig. S5a, ESI†), and represented this as a percentile of the respective null distribution (Fig. 2g).

With this *in vitro* method, we predicted the *in vivo* metastasis of seven out of the eight heterogeneous cell lines tested. First, the SUM1315 MO2 cell line, which metastasizes highly to bone,[37] most highly matches the bone fingerprint and is clearly anti-brain and anti-lung tropic. Both the MDA-MB-468[33] and BT549[28] cell lines metastasize to the lung *in vivo*, and are clearly lung tropic by our fingerprint. The HER2+ cell line HCC 1954 is lung metastatic[32] and phenotypically lung tropic. The HER2+ MDA-MB-361 cell line was derived from a brain metastasis, metastasizes to the brain in mice,[38] and is comparatively brain tropic by our fingerprint. The parental MDA-MB-231 is spontaneously metastatic to many sites upon orthotopic implantation into the mammary fat pad[39] and intracardiac injection,[4,40] and metastasizes to the bone more than to the brain or lung through the latter technique.[40] Our fingerprint identified this cell line as highly bone tropic (0.62 fractional similarity, 90th percentile of the randomly generated bone tropism null distribution), significantly greater than the similarity to the brain and lung fingerprints, reflecting this feature of *in vivo* behavior (Fig. 2g and Fig. S5a, ESI†). We focused on this particular aspect of MDA-MB-231 behavior, as this is the parental cell line from which the bone, brain, and lung tropic cell lines were derived by Massagué. As a non-tropic control, the MCF7 cells are only moderately metastatic, and not tissue-specific,[29,30] and were not tropic to any one tissue based on our fingerprint. Finally, the HER2+ SkBr3 cells were identified as brain tropic in our approach, but they are not highly tumorigenic or metastatic to the brain or other sites *in vivo*.[31] HER2-overexpressing tumors often spread to the brain in humans,[11] which may explain the prediction we obtained. However, we would argue that this example shows that this approach, although highly predictive, is not 100% accurate, and is only a predictor of tissue selectivity, not metastatic capability.

When looking more closely, we discovered that the unique features making up each fingerprint are primarily comprised of differential responses to EGF stimulation. In fact, all of the features in the bone tropic fingerprint are related to EGF stimulation, 80% for the brain fingerprint, and 62% for the lung fingerprint. Concurrently, we found that the cell lines with higher EGFR mRNA expression displayed stronger matching to just one of the tropic fingerprints than the cell lines with lower EGFR expression (Fig. 2g and Fig. S5e, ESI†). Using published proteomic data,[41] we discovered a strong correlation between basal EGFR expression in our cell lines and the CoV of the *in vitro* tropism (Fig. S5f, ESI†), confirming this qualitative trend. Although outside the scope of this study, we speculate that a more exhaustive screen of additional growth factor responses, or across different ECM protein combinations, could

lend even more specificity for this fingerprinting approach, potentially resolving the inconsistency we observed in the SkBr3 cell line.

In sum, we compiled sixty-six features of integrin-mediated phenotypes observed in bone, brain, and lung tropic cell lines into an *in vitro* fingerprint, which predicted the *in vivo* metastasis of other more heterogeneous cell lines. We stress that this result was impossible to achieve using any single adhesion or motility measurement, and instead required quantification of patterns of behavior. Because only some singular features of two-dimensional cell behavior can predict responses in a more realistic three-dimensional environment,[9] this requirement was perhaps not surprising. The divergence between two and three-dimensional phenotypes necessitated this new method of analysis to accurately connect measurements made in a two-dimensional context to *in vivo* outcomes.

Targeting integrin binding on *in vitro* ECMs reveals the need for subtype-specific analysis

Integrins have been explored as cancer therapeutics in pre-clinical and clinical trials, but many of these drugs have only shown limited success.[42] We hypothesized that this lack of clinical efficacy may be in part from differential responses across the heterogeneity of breast cancer subtypes. Because we are controlling for integrin-binding on our ECM protein-coupled surfaces, our predictive phenotyping platform provides a unique opportunity to investigate integrins as therapeutic targets for metastatic disease. We focused on triple negative and HER2+ cancer, which both have particularly poor prognosis. As proof of concept, we chose MDA-MB-231 as a representative triple negative cell line, and used the SkBr3s to examine HER2+ breast cancer. We focused on integrin subunits with various affinities to the proteins on each of our ECMs: β_1, which is involved in binding to many proteins across all the *in vitro* ECMs; α_2, which binds primarily to collagens but also to laminin (collagen is present on ECMs 1 and 3; laminin is present on ECMs 2 and 3); and α_6, which specifically binds laminin (present on ECMs 2 and 3).[43]

We performed the same adhesion and motility experiments from Fig. 1c in the presence of function-affecting antibodies to these three integrins. Targeting the function of each integrin reduced both cell adhesion and migration of the MDA-MB-231 cell line on the three ECMs. When cell adhesion measurements were plotted against cell migration measurements for each antibody of interest (as well as without antibody and with EGF stimulation), the responses clustered together (Fig. 3a and b). This revealed that the MDA-MB-231 cells were overall more sensitive to the antibody treatments and EGF stimulation than they were to the three ECMs. Each antibody had a different potency, resulting in a strong correlation between two measurements of cell adhesion (spreading rate and maximum area) and motility (migration speed and displacement, Spearman correlations and *p*-values shown in Fig. 3a and b).

This same clustering of responses around each ECM was not observed in the HER2+ SkBr3 cell line. Most surprisingly, treatment with the β_1 and α_2 integrin antibodies actually caused an increase in ECM sensitivity in these cells. Specifically, the SkBr3 cells migrated faster and farther on the collagen-rich ECM 1 (Fig. 3c and d, arrows) in the presence of both of these function-affecting antibodies. Looking more closely at videos of the cell behaviors, we determined that the increase in motility was caused by complete detachment of individual SkBr3 cells, which would then re-adhere at different locations, essentially hopping along the surface of the biomaterial (Fig. 3e, f and Fig. S6a, Movie S1, ESI†). This behavior resulted in a small population of cells that had abnormally fast migration speeds and long displacements (Fig. 3e, f and Fig. S6a, ESI†). This phenotype was observed most often on ECM 1 in the presence of the β_1 and α_2 integrin antibodies, and rarely in the three other conditions (Fig. S6b, ESI†). This was never observed in the MDA-MB-231 cells.

This behavior did not resemble traditional adherent cell motility, and we suspected that these weakly-to-non adherent cells might be highly invasive in a 3D context, perhaps resembling an amoeboid-type motility.[44] When these same β_1 and α_2 integrin antibodies were administered to SkBr3 cells that were seeded onto ECM 1 overlaid with a 3D collagen gel, we observed increased upward 3D invasion by three to four fold compared to the control (Fig. 3g and Fig. S6c, ESI†). The SkBr3 cell line has lower integrin protein expression of each of these integrin subunits in comparison to the MDA-MB-231 cell line,[45,46] which is potentially responsible for the detachment and invasion phenotypes. However, there were no differences in β_1 integrin mRNA expression (Fig. S7a, ESI†), suggesting that a proteomic, but not a genomic analysis could have predicted this result. We propose that the motility and invasion of some cancer cell lines can actually be increased when targeting integrins,[47] potentially limiting the efficacy of this class of therapeutics in some patients. The striking differences in antibody responses of the triple negative MDA-MB-231 and HER2+ SkBr3 cell lines highlight the heterogeneity in response to integrin therapeutics observed across different breast cancer clinical subtypes, potentially explaining the limited efficacy of these drugs thus far.

Integrin binding and expression dictates tropism

The β_1, α_2, and α_6 integrin subunits each have a reported role in cancer: high expression of both β_1 (ref. 48) and α_6 (ref. 49) integrins drives tumorigenicity and metastasis, while α_2 integrin is a tumor suppressor.[50] Although all three integrins appear important in overall patient prognosis, none have been directly connected with tissue-specific metastasis. Given the results in Fig. 2, where tropic cells shared similar integrin-mediated phenotypes, and Fig. 3, where the triple negative MDA-MB-231 and HER2+ SkBr3 cell lines had divergent responses to integrin targeting, we hypothesized that targeting β_1, α_2, and α_6 integrins could shift the observed *in vitro* tropism of these cell lines.

We compiled the adhesion and motility measurements for the MDA-MB-231 and SkBr3 cell lines that were performed in the presence of integrin antibodies into our predictive fingerprint. This treatment with integrin antibodies, which lowers the binding ability with these integrin subunits, shifted the tropism predictions of both cell lines (Fig. 4a and b). In the MDA-MB-231 cells,

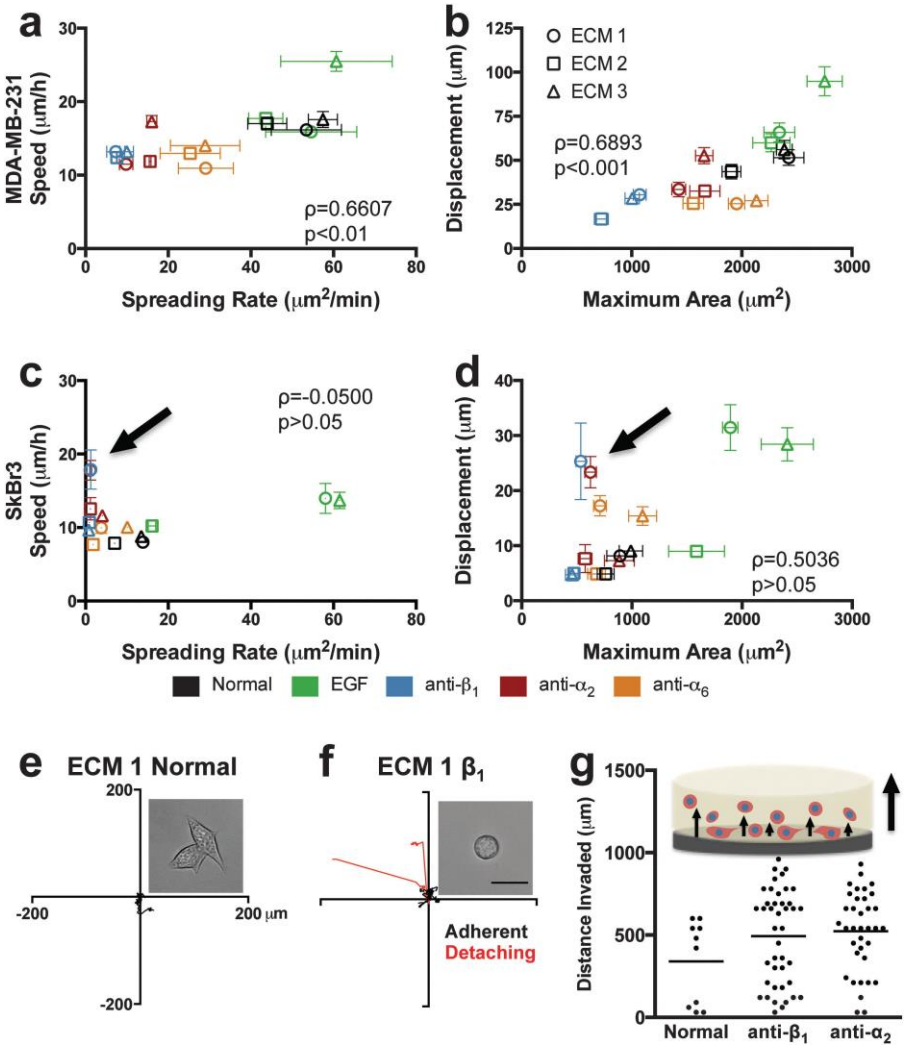

Fig. 3 Correlations between adhesion and migration responses identify potent integrin antibodies *in vitro*. Pairwise comparisons between adhesion and migration measurements in the (a, b) MDA-MB-231 and (c, d) SkBr3 cell lines across normal, EGF-stimulated, and integrin antibody conditions. Arrows highlight conditions where integrin antibodies increased migration metrics. Spearman correlations are indicated on each plot with two-tailed *p*-values. Circle: ECM 1; square: ECM 2; triangle: ECM 3; black: normal; green: EGF; blue: anti-b_1 integrin; red: anti-a_2 integrin; orange: anti-a_6 integrin. (e, f) SkBr3 migration mechanisms are displayed *via* 10 random cell paths under (e) normal and (f) anti-b_1 conditions. Red paths identify cells detaching and adhering elsewhere on the surface. Inset: representative images of cell morphology. Scale bar is 25 μm. (g) Individual cells that invaded into an overlaid 3D collagen gel from the ECM 1 surface after 48 hours. Bar indicates mean distance invaded of all invading cells. Inset: schematic of cells invading upward from the ECM surface into an overlaid gel.

treatment with any of the integrin antibodies shifted the tropism from bone to brain (Fig. 4a). In other words, blocking binding *via* b_1, a_2, or a_6 integrins made the MDA-MB-231 cells less similar to the bone tropic fingerprint, and instead, they more resembled the brain tropic cell subpopulation (compare the tropism category containing the highest black bar with the tropism category containing the highest colored bars for each integrin). The SkBr3 cells shifted from brain to bone tropic upon b_1 integrin inhibition, they became equally brain and lung tropic when a_2 integrin was blocked, and a_6 integrin targeting had no effect (Fig. 4b). Just as we found in Fig. 3, the MDA-MB-231 cell line was far less sensitive to differences in ECM during integrin targeting than the SkBr3 cell line. Interestingly, the only response that was shared between the MDA-MB-231 and SkBr3 cell lines was the "null effect" of targeting a_2 and a_6 integrins on lung tropism. Overall, these results suggest that targetingany one of these integrins may effectively prevent bone metastasis in the triple negative cell line, while b_1 integrin may be an efficacious therapeutic target to prevent brain and lung (but not bone) metastasis in the HER2+ cell line.

For this approach to have significant impact, it must be compared with clinical patient outcomes. We analyzed gene expression data from 630 breast cancer patients with known metastatic outcomes (combined from GSE 2034, GSE 2603, GSE 5327, GSE 12276; includes 214 bone metastases, 35 brain metastases, 101 lung metastases). We first determined whether expression of these integrin genes was associated with tissue-specific metastasis in the clinic (Fig. 4c–e), and then compared

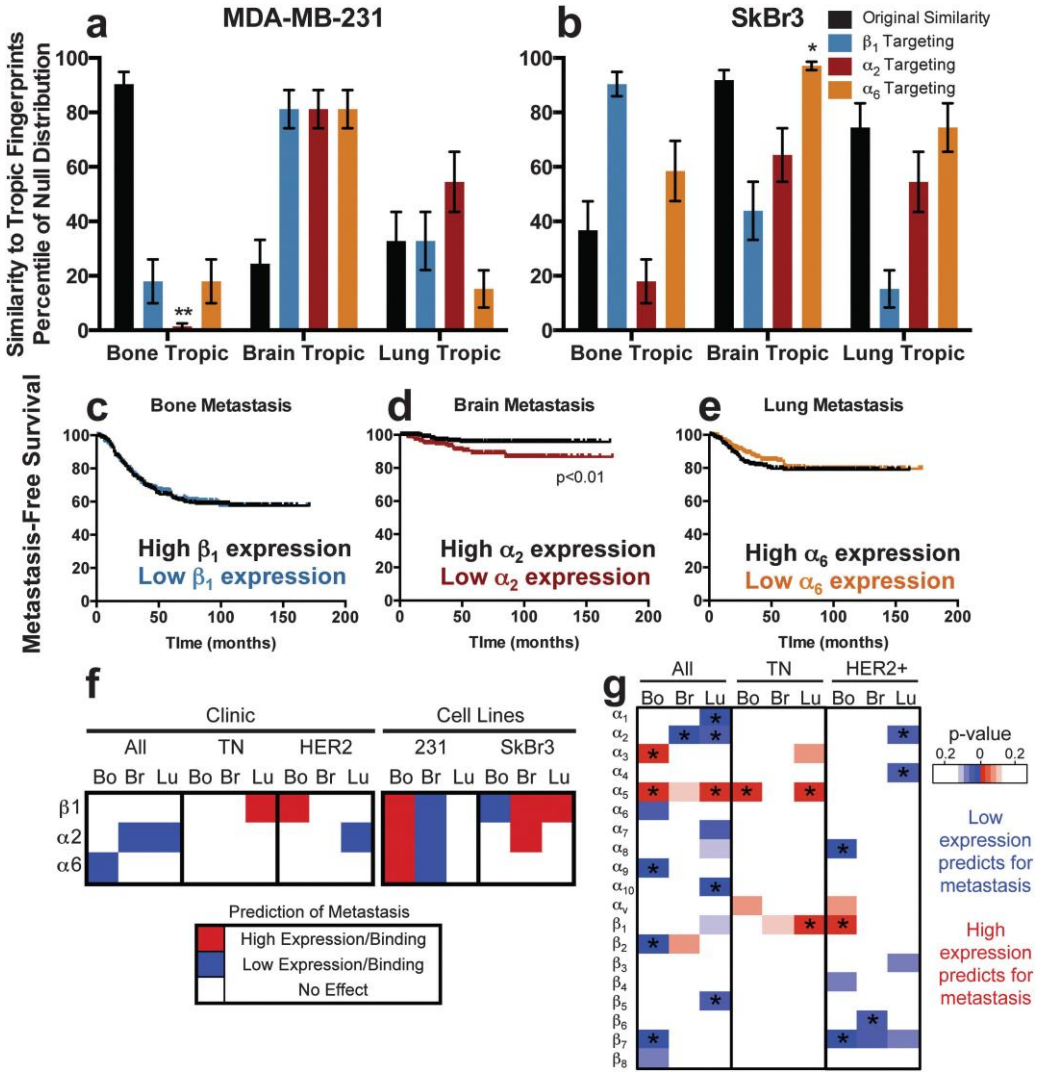

Fig. 4 Integrin binding and gene expression predict breast cancer metastasis. Integrin antibodies shift tropic fingerprinting of (a) MDA-MB-231 and (b) SkBr3 cell lines. Compare the tropism containing the highest black bar (normal) with the category containing the highest value for each colored bar (blue: b_1, red: a_2, orange: a_6 integrin). (c–e) Magnitude of integrin gene expression dictates (c) bone, (d) brain, and (e) lung metastasis clinically (analysis of GSE 2034, GSE 2603, GSE 5327, GSE 12276). (f) Left, effect of b_1, a_2, and a_6 integrin gene expression on clinical metastasis in all patients, only triple negative patients, and only HER2+ patients. Right, effect of b_1, a_2, and a_6 integrin antibodies on tropic fingerprint in the MDA-MB-231 (triple negative) and SkBr3 (HER2+) cell lines. Red: high expression or binding predicts for increased metastasis; blue: low expression or binding predicts for increased metastasis; white: no effect. (g) Heat map displaying integrin genes that significantly predict for tissue-specific metastasis in patients. Blue: low expression of the gene predicts for increased risk of metastasis; red: high expression of the gene predicts for increased risk of metastasis; white: no effect on metastasis. Asterisks indicate statistically significant relationships, but other clearly visible trends are included for completeness. Abbreviations: Bo: bone metastasis; Br: brain metastasis; Lu: lung metastasis.

this with the results predicted from our *in vitro* fingerprinting approach when targeting integrin binding (Fig. 4a and b). As one example, the gene expression data sets revealed that low a_2 integrin expression in the primary tumor correlates with higher rates of brain metastasis, and, similarly, decreasing a_2 integrin binding significantly increased the MDA-MB-231 brain tropic phenotype *in vitro* (compare Fig. 4a, increase in brain tropism from black to red, with Fig. 4d, increased risk of brain metastasis from black to red). As another example, a_6 integrin has no effect on lung metastasis, in both clinical patient outcomes and the tropism of both cell lines (Fig. 4a and b). In this case, when these same patients are classified into clinical subtypes using published annotations,[4,51] and survival analysis is repeated in these smaller, specific patient populations, the subtype-specific trends in tropism match the behavior of corresponding the cell line we examined (Fig. 4f). We want to stress that this latter result was only one of two shared outcomes we found between the triple negative and HER2+ cell lines, highlighting the need for individual analysis of subtypes, as demonstrated in Fig. 3.

Interestingly, when this clinical analysis is expanded to examine all integrin subunits, many are associated with metastasis to the brain, bone, or lung, but these same genes are not conserved within the triple negative and HER2+ subtypes

(Fig. 4g and Fig. S7b, ESI†). Strong examples are α_1, α_3, and β_2 integrins, where altered expression of these genes is correlated with tropism to a specific tissue across breast cancer broadly, but they are not correlated with tropism within these two subtypes. Instead, different integrin subunit genes are correlated with tissue-specific metastasis within each of these subtypes. As two examples, low α_8 integrin expression is correlated with bone metastasis in the HER2+ subtype, and high β_1 integrin expression is correlated with lung metastasis in the triple negative subtype.

When we compared our *in vitro* phenotyping approach to this existing gene data set, we found that our *in vitro* tropism responses imperfectly correlated with these genetic indicators. High β_1 expression is associated with poor prognosis,[48] and in our work, targeting β_1 integrin makes the MDA-MB-231 cell line much less bone tropic, matching this result (Fig. 4a). However, the SkBr3 cell line significantly increased in bone tropism upon targeting β_1 integrin, (Fig. 4b), and our clinical analysis surprisingly showed that the magnitude of β_1 integrin gene expression is not predictive of bone metastasis in patients (Fig. 4c). However, the divergent responses of the two candidate cell lines, from different clinical subtypes, combine to match the clinical pattern. Although our β_1 integrin targeting results were not predictive of these clinical outcomes, we emphasize that this highlights the divergence expected between a genetic-focused approach and a protein-based, functional approach, as we have taken here.

In sum, we have used a simple integrin-mediated phenotyping approach to predict breast cancer metastasis in a large panel of cell lines, and we implicate β_1, α_2, and α_6 integrin binding in tissue-specific spread distinctly across several disease subtypes, a phenomenon that is prevalent across many integrin gene-tissue combinations in clinical patient populations (Fig. 4g and Fig. S7b, ESI†). Importantly, our integrin targeting results provide many insights not apparent in gene expression data, highlighting the need for functional, protein-centric screens of cell behaviors. We emphasize the utility of this phenotyping approach as quick, bench-top screening tool that can be used to predict *in vivo* outcomes. Here, we have illustrated this functionality by predicting metastatic outcomes and identifying biomarkers potentially overlooked when analyzing gene data sets from heterogeneous clinical patient populations.

Discussion

Extracellular matrix (ECM) properties, such as tissue stiffness,[7] local growth factors,[52] stromal cells,[6,53] and ECM proteins,[5,10,54] can each individually promote metastasis. However, we are only beginning to understand the role of each of these properties in tissue-specific metastatic colonization.[5,6,52,54] To truly understand the microenvironmental factors that mediate metastasis, each factor must be systematically isolated from other cues, and investigated functionally in a controlled system that closely represents the *in vivo* microenvironment. While others have used adhesion phenotyping to differentiate metastatic and non-metastatic cells,[10] and tropic subpopulations,[34] these studies have either used end-point analyses or have included multiple convoluted microenvironmental factors. The biomaterial platform we designed overcomes these limitations by mimicking some of the biochemical complexity present in *in vivo* ECMs, while limiting microenvironmental cues to only those ECM proteins presented by the system, with growth factors supplemented as desired. This can then be used for observing differences in adhesion and motility resulting from differential binding to these multi-protein ECMs.

We propose this system as a potential bridge between the overwhelming complexity of *in vivo* observations and simple *in vitro* cell biology, without the need for expensive or lab-specific 3D models. The first challenge we noted was that genomics is the increasingly popular approach for identifying cancer biomarkers due to its ease and increasingly low cost, facilitating large-scale patient-specific analysis. However, gene expression does not necessarily correlate with protein expression or function. Therefore, we attempted to connect gene-centric patient data with a protein-centric cell screening approach. Breast cancer was a suitably complex target to study, as it is a notoriously heterogeneous disease.[11] This heterogeneity may be why integrins were not consistently strong genetic biomarkers across large, diverse patient populations (Fig. 4). Our protein-centered approach mirrored this inconsistency between the SkBr3 and MDA-MB-231 cell lines (Fig. 3 and 4), and we quantified significant differences in cell adhesion and motility between the three MDA-MB-231 tropic subpopulations (Fig. S3, ESI†), which have minimal differences in integrin gene expression (Fig. S7a and c, ESI†).[2–4]

The immediate conclusion is that differences in gene expression of integrin subunits alone is not sufficient to mediate metastasis to different tissues. This remains an open question, as gene expression of several integrins was strongly correlated with tropic metastasis in patients (Fig. 4g), but most of these integrins have not yet been functionally examined *in vivo*. For those that have, the *in vivo* reports we found did not match the clinical correlations in our analysis of the gene data sets. For example, $\alpha_3\beta_1$ integrin binding is known to mediate lung metastasis in rats and mice.[18,19] However, high α_3 integrin gene expression showed only a non-significant correlation with lung metastasis, and only when we isolated the analysis to triple negative patients. Across all subtypes of patients, high α_3 integrin gene expression actually more highly correlated with bone metastasis (Fig. 4g). As another example, $\alpha_v\beta_3$ integrin has been linked to bone metastasis in rats,[17] however, β_3 integrin gene expression in these data sets does not predict for bone metastasis, and α_v integrin gene expression is only more strongly associated with bone metastasis in the triple negative and HER2+ subtypes than across all patients (Fig. 4g). It is important to note that clinical tumor samples contain stroma, are often contaminated with immune cells, fibroblasts, and epithelial cells,[55] and even the best dissection techniques only produce tumor contents near 70%,[3] which could skew genetic analysis. We stress that based on this conflict between our data and the genetic analysis we performed, alongside obvious conflicts between these gene data sets and literature reports,

successful risk assessments of tropic metastasis require an integrated proteomic and genetic approach.

Our results suggest that tropism depends more upon how cells are able to use their integrins to bind to the ECM, rather than variations in gene expression, even in genetically similar cell lines. We propose that this type of functional screen, across many heterogeneous cell lines, and many phenotypes, is independent of these conflicts observed between genetic and proteomic analyses. This is demonstrated by the fact that the bone, brain, and lung tropic phenotypes we obtained were well-conserved across cell lines that share the same preferred metastatic site, but that have very different integrin gene and protein expression (Fig. 2g). As an example, the MDA-MB-231 and SUM1315 MO2 cells are both bone metastatic in mice,[2,37] and both matched our bone fingerprint, even though the MDA-MB-231 cell line has significantly higher surface expression of α_2 and α_5 integrins, and lower expression of $\alpha_v\beta_3$, than the SUM1315 MO2 cell line.[56] Others have shown that metastasis depends upon activation, rather than simply expression, of integrins.[57–59] It is known that splicing variation of a single integrin can initiate cancer stem cell plasticity and likely impact tropism.[16] Taken together, this suggests that understanding the true role of integrins in mediating tropism requires a functional analysis of how metastatic cells interact with the tissue site, as we have taken here, rather than a limited view of only gene expression profiles.

One study took a similar approach to ours and examined cancer cell behavior on different biomaterial surfaces, but focused on stiffness as the driving force.[7] They found that measurements of cell area were not predictive of metastasis, but there was a weak correlation between both migration and proliferation with metastatic site preference.[7] Upon our own closer examination, their data shows that tropic subpopulation cell areas were more sensitive to the biomaterial environments than the non-tropic subpopulations, consistent with our results (Fig. S4a, ESI†). This suggests that applying our approach to other microenvironment cues, such as material stiffness, would provide further insight into the biophysical regulation of tropism and add another dimension to the fingerprint reported here.

Integrins are attractive therapeutic targets for metastasis because they mediate adhesion to the tissue microenvironment, can confer resistance to treatments, and drive disease progression and stemness.[16,60,61] However, integrin-targeted therapeutics have experienced limited clinical success,[42] and some potentially dangerous outcomes have been reported.[62,63] Others have shown that inhibiting β_1 integrin can reduce primary tumor growth, but this subsequently promotes lung metastasis by switching the migration mode of triple negative 4T1 mouse breast cancer cells from collective to single cell migration, observed both *via* knockdown and antibody targeting.[63] Similar to these results, we saw both efficacious and dangerous cell adhesion and motility phenotypes when we targeted integrin binding on our ECMs. β_1 and α_2 integrin antibody treatment on ECM 1 prevented spreading, but increased both 2D motility and 3D invasion of the SkBr3 cell line (Fig. 3c–g). In the highly metastatic MDA-MB-231 cell line, our β_1, α_2, and α_6 integrin antibodies each reduced adhesion and motility (Fig. 3a and b), but also dangerously increased the phenotypic similarity to brain tropism (Fig. 4a). These dangerous phenotypes may result from plasticity and adaptability of metastatic cells when integrins are targeted,[62,64] which is potentially responsible, in part, for the lack of clinical success of integrin therapeutics.

These results highlight the known heterogeneity challenge in breast cancer, and there is a ground swell toward personalized therapeutics. As proof of concept toward using our platform for discovery of treatments for metastasis in this heterogeneous disease, we examined cell lines from two clinical subtypes with poor prognosis, and found that they displayed striking differences in response to integrin antibodies. In our fingerprinting results, these cell lines only shared two of nine possible responses to the integrin antibodies, suggesting subtype-specific roles for β_1, α_2, and α_6 integrins in bone, brain, and lung metastasis. The MDA-MB-231 tropism was equally affected by each of these integrin antibodies, while the SkBr3 cell line was much more sensitive to β_1 integrin targeting than to targeting α_2 or α_6 integrins. Supporting the subtype-specific responses we observed, there is *in vivo* evidence for distinct roles of the same integrin in tumorigenesis and metastasis in different types of breast cancer.[64] In PyV-MT tumors, β_1 integrin is required for tumor initiation. However, in mice that also express active *erbB2* (HER2), β_1 integrin is only necessary to mediate lung metastasis, while tumorigenesis is independent of β_1 integrin.[64] In our work, β_1 integrin targeting had no effect on the MDA-MB-231 lung tropism, but significantly decreased lung tropism in the HER2+ SkBr3 cell line. However, of the eleven cases where integrin gene expression predicts for tropic metastasis in the clinic, only two were conserved for both the triple negative and HER2+ patient subpopulations (Fig. 4). This indicates that integrins are not conserved biomarkers across the entire heterogeneous patient population, necessitating subtype- or even tumor-specific screening to identify successful integrin therapeutics. We suggest that this type of functional screening would be capable of identifying efficacious tumor-specific therapeutics, providing additional insight in combination with current genetics-based approaches.

Conclusions

Here, we developed a simple, yet robust *in vitro* biomaterial platform that allowed us to quantify many different cell phenotypes associated with adhesion and motility in order to take a comprehensive view of how integrin–ECM interactions regulate bone, brain, and lung metastasis. Importantly, these results would not have been realized by simply looking at individual measurements, and required us to instead quantify patterns of measurements. This compiled pattern of measurements created a phenotypic *fingerprint*, which is dependent only on integrin binding to the ECM, and can predict integrin-mediated metastatic spread in cell lines. Our results suggest that targeting integrins across the heterogeneous breast cancer clinical subtypes is not appropriate (Fig. 3 and 4), potentially explaining the

lack of success of integrin-targeted therapeutics thus far. Importantly, our results are largely independent of integrin gene or protein expression, highlighting the need for this type of functional approach, particularly when comparing among different breast cancer subtypes. We emphasize the utility of this biomaterial platform to screen integrin–ECM interactions rapidly and reproducibly, and by collectively quantifying patterns of phenotypes *en masse*, the ability to predict *in vivo* fate with a simple *in vitro* approach.

Materials and methods

Cell culture

Human breast cancer cell lines MDA-MB-231, BT549, MCF7, and SkBr3 were generous gifts from Shannon Hughes at the Massachusetts Institute of Technology. Highly metastatic MDA-MB-231 variants, isolated from *in vivo* selection, were kindly provided by Joan Massagué.[2–4] These cell lines preferentially metastasize to the bone (1833 BoM), brain (831 BrM2a) or lung (4175 LM2). All were routinely cultured in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum (FBS), 1% penicillin–streptomycin (P/S), 1% L-glutamine, and 1% non-essential amino acids. The MDA-MB-361, SUM1315 MO2, MDA-MB-468, and HCC 1954 cell lines were provided by Mario Niepel at Harvard Medical School. SUM1315 MO2 cells were cultured in Ham's F-12 medium supplemented with 5 mg ml^{-1} insulin, 10 ng ml^{-1} EGF, 10 mM HEPES, 10% FBS, and 1% P/S. MDA-MB-361 cells were cultured in Leibovitz's L-15 medium with 20% FBS and 1% P/S. MDA-MB-468 cells were cultured in Leibovitz's L-15 medium with 10% FBS and 1% P/S. HCC 1954 cells were cultured in RPMI medium with 10% FBS and 1% P/S. With the exception of the MDA-MB-361 and MDA-MB-468 cells, which were cultured without supplemental CO_2, all cell lines were cultured at 37 °C and 5% CO_2. All cell culture supplies were purchased from Life Technologies (Carlsbad, CA).

Preparation of ECMs

Glass coverslips (15 mm and 18 mm diameter, Fisher Scientific, Agawam, MA, USA) were oxygen plasma treated (Harrick Plasma, Ithaca, NY, USA), and silanized through vapor phase deposition of (3-aminopropyl)triethoxysilane (Sigma-Aldrich, St. Louis, MO, USA) at 90 °C for a minimum of 18 hours.[65] The coverslips were rinsed sequentially in toluene (Fisher Scientific), 95% ethanol (Pharmco-AAPER, Brookfield, CT, USA), and water, and dried at 90 °C for one hour. They were then functionalized with 10 g L^{-1} *N*,*N*-disuccinimidyl carbonate (Sigma-Aldrich) and 5% v/v diisopropylethylamine (Sigma-Aldrich) in acetone (Fisher Scientific) for two hours.[66] Coverslips were rinsed three times in acetone and air-dried. ECM protein cocktails were then covalently bound to the glass coverslips through reactive amines,[67] using cocktails that were inspired by the ECM of secondary sites as follows: ECM 1 (bone): 5 mg cm^{-2} of 99% collagen I and 1% osteopontin; ECM 2 (brain): 1 mg cm^{-2} of 50% fibronectin, 25% vitronectin, 20% tenascin C, and 5% laminin; and ECM 3 (lung): 2 mg cm^{-2} of 33% laminin, 33% collagen IV, 15% collagen I, 15% fibronectin, and 4% tenascin C (all weight%). Rat-tail collagen I and natural mouse laminin were purchased from Life Technologies; human tenascin C, human vitronectin, and human osteopontin from R&D Systems (Minneapolis, MN, USA); human collagen IV from Neuromics (Edina, MN, USA); and human plasma fibronectin from EMD Millipore (Billerica, MA, USA). Coverslips were incubated with proteins at room temperature for three hours, then with 10 mg cm^{-2} MA(PEG)$_{24}$ (Thermo Scientific, Rockford, IL, USA) for two hours to block non-specific protein adsorption on any remaining surface area.

An ELISA was used to quantify coupling of collagen III (Fig. S1, ESI,† 0, 1, 10 and 50 mg cm^{-2}; Fibrogen, San Francisco, CA, USA). Surfaces were blocked with 10 mg ml^{-1} BSA (Fisher Scientific) for 1 hour, reacted with 1 : 200 primary antibody (Santa Cruz Biotechnology, Dallas, TX, USA) for 1.5 hours, and then 1 : 200 HRP-conjugated secondary antibody (Abcam, Cambridge, MA, USA) for 1 hour at room temperature, rinsing four times with PBS in between each step. Coverslips were incubated with 0.1 M sodium acetate (pH 5.5) containing 1 mg ml^{-1} 3,3′,5,5′-tetramethylbenzidine (Sigma-Aldrich) and 0.05 wt% hydrogen peroxide (Fisher Scientific) and the reaction proceeded for 30 minutes, then was stopped with 1 M H_2SO_4 (Sigma-Aldrich). The absorbance at 450 nm was read immediately (Biotech ELx800, Winooski, VT, USA).

Cell adhesion and polarization

Cells were seeded at 4000 cells per cm^2 in growth medium, medium supplemented with 40 ng ml^{-1} epidermal growth factor (EGF, R&D Systems) immediately, or with antibody pretreatment, which occurred for 30 minutes prior to seeding. Anti-b_1 integrin (clone P5D2, R&D Systems) was used at 0.83 mg ml^{-1}, and anti-a_2 and anti-a_6 integrins were used at 3.3 mg ml^{-1} (a_2: clone P5E6, a_6: clone NK1-GoH3, both from Millipore). Cell adhesion was captured through imaging cells beginning 10 minutes after seeding in an environment-controlled Zeiss Axio Observer Z1 microscope (Carl Zeiss, Oberkochen, Germany) using an AxioCam MRm camera and an EC Plan-Neofluar 20× 0.4 NA air objective. Images were taken using AxioVision (Carl Zeiss) at five-minute intervals for a minimum of 2 hours until cells had reached steady-state. ImageJ (National Institutes of Health, Bethesda, MD, USA) was used to trace cell areas using the built-in measurement function. Spreading rate was defined as the slope of the linear portion of the area *versus* time graph during initial adhesion. Individual cells were marked when they had polarized, and the fraction of cells that had polarized at 2 hours post-seeding was quantified for comparison across conditions. Cells that contacted other cells, underwent division or apoptosis, or were not fully in frame were excluded. $N \geq 2$ independent biological replicates, $N \geq 40$ cells per condition.

Cell migration

Cells were seeded at 4000 cells per cm^2 and given 18 hours to adhere in growth medium. Seeded cells were treated with a live-cell fluorescent dye (CMFDA, Life Technologies), and then

provided fresh medium or medium supplemented with EGF and/or integrin antibodies (as described above) 4 hours prior to microscopy. Brightfield and fluorescent images were taken at 15 minute intervals for 12 hours using an EC Plan-Neofluar 10x 0.3 NA air objective (Carl Zeiss). Cells were tracked using Imaris (Bitplane, St. Paul, MN, USA) to generate individual cell paths. Individual cell speeds were determined by calculating a speed at every 15 minute time interval, then averaging these over the entire 12 hours. Displacement was defined as the net change in position of the cell over the total time. Chemotactic index is the ratio of a cell's net displacement to path length. Individual cell average speeds, net displacements, and chemotactic indices were then averaged to determine a mean population value. Cells that contacted other cells, underwent division or apoptosis, or were not fully in frame for the entire 12 hours were excluded. $N \geq 2$ independent biological replicates, $N \geq 90$ cells per condition.

Fingerprinting

We quantified adhesion and migration phenotypes for the three tropic subpopulations, both under normal and EGF-stimulated conditions. To create phenotypic fingerprints, first, we calculated the CoV (standard deviation divided by average, calculated for final mean values for one measurement across the three ECMs) for each tropic cell line to determine their ECM-sensitivity in each parameter (a CoV greater than 0.1 was considered ECM-sensitive; Fig. 2a and d). Second, we quantified the fold-change in these metrics in response to EGF stimulation on each ECM (fold-changes greater than 1.15 or less than 0.85 were defined as significant; Fig. 2b and e). Finally, we did all pairwise statistical comparisons across ECMs (*via* a one-way ANOVA, described below), and quantified ordering of values (*i.e.*, for normal speed measurements, was ECM 1 greater than, equal to, or less than ECM 2; Fig. 2c and Fig. S4b, ESI†). We then compiled this data to create a fingerprint of the 66 phenotypes associated with a specific tissue preference (Fig. 4d and e and Fig. S4b, ESI†). Upon comparison of each phenotype across the three tropic subpopulations, only those phenotypes which were identified in just one tropic cell line were retained, and all phenotypes shared in two or three of the cell lines were ignored (Fig. 4b–e, and Fig. S4b, ESI,† dark colors). To validate these fingerprints, we calculated these same criteria for other cell lines with known *in vivo* metastasis. We then overlaid these patterns onto each tropic fingerprint, and quantified the fractional similarity between each cell line and the unique phenotypes within these fingerprints (Fig. 2g and Fig. S5a, ESI†).

Null distributions of bone, brain, and lung tropisms were generated *via* a custom code written in MATLAB R2012a (The MathWorks, Inc., Natick, MA). 10 000 random cell line phenotypes were generated and were compared with the unique features of the bone, brain, and lung fingerprints to obtain tropism similarity values. The distributions were ordered, percentiles assigned, and then the heterogeneous cell lines were compared with each of these null distributions, matching their similarity with the percentile associated with this same similarity in the appropriate null distribution. Because our distributions were comprised of 10 000 cells with only 13 or 15 possible discrete values, each percentile for a cell line had a range, illustrated *via* the error bars in Fig. 2g. Cell lines in the top or bottom 5% of the ordered distributions were considered to be significantly higher or lower than the null distribution. Specifically, the lowest and highest 500 randomly generated cells had $p < 0.05$.

Collagen gel invasion

Cells were seeded at 600 cells per cm^2 in normal growth medium and given 12 hours to adhere. Collagen gels (2 mg ml^{-1}) were made by mixing 5% v/v 1 M NaOH (Fisher Scientific), growth medium and type I collagen (Life Technologies) on ice. The medium was removed from the seeded coverslips, and gel solutions were overlaid onto the cells. Gelation proceeded for 30 minutes at 37 °C and 5% CO_2, then 250 μl growth medium was added to each well. EGF and integrin antibodies were included in the gel solution and the supernatant medium as described above. For accurate comparisons, an equal number of randomly selected positions were observed in all conditions. $N = 2$ independent biological replicates, each with three technical replicates, and 25 positions were imaged per technical replicate.

Gene expression analysis

Gene expression data from primary breast tumors (GSE 2034, GSE 2603, GSE 5327, and GSE 12276) was retrieved from the NCBI Gene Expression Omnibus (GEO) database. Each was RMA normalized using the R platform (Vienna, Austria).[68] Datasets were cross-mapped based upon common probes, and batch effects were corrected with an Empirical Bayes method using the ComBat algorithm in R.[69] Genes represented by more than one probe were collapsed to the probe with the highest mean value. For metastasis-free survival analysis, patients with known metastasis and clinical subtype information were classified into groups of high and low gene expression based upon median expression. Published survival and clinical subtype information was used.[4,51] Survival was analyzed in Prism v6.0b (GraphPad Software, La Jolla, CA, USA), and significance was evaluated using a log-rank (Mantel–Cox) test. To analyze integrin expression in the tropic cell lines, GSE 2603 and GSE 12237 were retrieved from the NCBI GEO database, RMA normalized, combined using the ComBat algo- rithm as previously described, and data corresponding to the cell subpopulations used here (1833, BrM2a, and 4175) were extracted. Individual integrin genes were extracted, values were collapsed to the probe with the highest expression, and a heat map was generated in R.

Reverse-transcriptase PCR

Cells were seeded at 40 000 cells per cm^2 and allowed to adhere to coverslips or tissue culture plastic for a minimum of 18 hours. Coverslips or wells were washed with PBS prior to detaching cells with 0.05% trypsin–EDTA (Life Technologies). Total RNA was extracted using the Genelute Mammalian Total RNA kit (Sigma)

followed by cDNA synthesis using the RevertAid reverse transcriptase protocol, with the exception of using RNasin 40 U μL^{-1}, (Promega, Madison, WI) as the RNase inhibitor. The amplification was carried out with Jumpstart Taq polymerase according to the manufacturer's instructions using a BioRad MJ Mini Personal thermal cycler (Hercules, CA). Primer sequences (Integrated DNA Technology, Coralville, IA) were as follows:

GAPDH (NM_001256799.1) forward 5′-CACTGACACGTTGGCAGTGG-3′

reverse 5′-CATGGAGAAGGCTGGGGCTC-3′

Integrin β_1: (NM_002211.3) forward 5′-CTGGGCTTTACGGAGGAAGT-3′

reverse 5′-GTCTACCAACACGCCCTTCA-3′

Integrin α_2: (NM_002203.3) forward 5′-CTGGTGTTAGCGCTCAGTCA-3′

reverse 5′-CCAGGGTGAACCAACCAGTA-3′

Integrin α_6: (NM_001079818.1) forward 5′-CGAGGACAAGCGGCTGTT-3′

reverse 5′-TGACCCCCATCCACTGATCT-3′

EGFR: (NM_005228.3) forward 5′-CTTCGGGGAGCAGCGATG-3′

reverse 5′-CAGCTCCTTCAGTCCGGTTT-3′

Each reaction was analyzed on a 1% agarose gel with ethidium bromide staining, then visualized under ultraviolet light using the IN Genius Syngene Bioimaging platform (Frederick, MD). GAPDH was used as a housekeeping gene. $N \geq 2$ independent biological replicates.

Statistical analysis and correlations

Statistical analysis was performed using Prism v6.0b. Data are reported as mean ± standard error. Statistical significance was evaluated using a one-way analysis of variance, followed by a Tukey's post-test for pairwise comparisons. Significance between proportions (*i.e.*, fraction of cells polarized) was evaluated with a Fisher's exact test with two-tailed *p*-values. To determine significance between slopes (*i.e.*, spreading rate), an analysis of covariance was used, with a one-way analysis of variance for multiple comparisons, defining the sample size as the degrees of freedom plus one. Spearman correlations were calculated from mean values paired by condition, and significance was determined using two-tailed *p*-values. $p < 0.05$ was considered statistically significant. $p < 0.05$ is denoted with *, <0.01 with **, <0.001 with ***, and <0.0001 with ****; $p \geq 0.05$ is considered not significant ('ns').

Acknowledgements

We are grateful to Shannon Hughes, Doug Lauffenburger, Jeffrey Blanchard, and Aaron Meyer for helpful intellectual discussions, Wei Chen for guidance with silane chemistry, Thomas McCarthy and the UMass MRSEC for use of equipment, and Isaac Han, Matthew Crotty, Elyse Hartnett, and Patrick Colleton for technical assistance. We thank Joan Massagué, Shannon Hughes, and Mario Niepel for generously providing cell lines. This work was funded by an NSF-NCI award DMR-1234852 to SRP and NGR, the University of Massachusetts, Amherst, and by a grant from the NIH (1DP2CA186573-01). SRP is a Pew Biomedical Scholar supported by the Pew Charitable Trusts. LEB was partially supported by National Research Service Award T32 GM008515 from the National Institutes of Health. ECD was partially supported by an HHMI undergraduate research fellowship. AMM was supported by NIH Grant CA168464.

Notes and references

1 J. A. Joyce and J. W. Pollard, *Nat. Rev. Cancer*, 2008, 9, 239–252.

2 Y. Kang, P. M. Siegel, W. Shu, M. Drobnjak, S. M. Kakonen, C. Cordon-Cardo, T. A. Guise and J. Massagué, *Cancer Cell*, 2003, 3, 537–549.

3 A. J. Minn, G. P. Gupta, P. M. Siegel, P. D. Bos, W. Shu, D. D. Giri, A. Viale, A. B. Olshen, W. L. Gerald and J. Massagué, *Nature*, 2005, 436, 518–524.

4 P. D. Bos, X. H. F. Zhang, C. Nadal, W. Shu, R. R. Gomis, D. X. Nguyen, A. J. Minn, M. J. van de Vijver, W. L. Gerald, J. A. Foekens and J. Massagué, *Nature*, 2009, 459, 1005–1009.

5 T. Oskarsson, S. Acharyya, X. H.-F. Zhang, S. Vanharanta, S. F. Tavazoie, P. G. Morris, R. J. Downey, K. Manova-Todorova, E. Brogi and J. Massagué, *Nat. Med.*, 2011, 17, 867–874.

6 R. N. Kaplan, R. D. Riba, S. Zacharoulis, A. H. Bramley, L. Vincent, C. Costa, D. D. MacDonald, D. K. Jin, K. Shido, S. A. Kerns, Z. Zhu, D. Hicklin, Y. Wu, J. L. Port, N. Altorki, E. R. Port, D. Ruggero, S. V Shmelkov, K. K. Jensen, S. Rafii and D. Lyden, *Nature*, 2005, 438, 820–827.

7 A. Kostic, C. D. Lynch and M. P. Sheetz, *PLoS One*, 2009, 4, e6361.

8 M. H. Zaman, *Nat. Rev. Cancer*, 2013, 13, 596–603.

9 A. S. Meyer, S. K. Hughes-Alford, J. E. Kay, A. Castillo, A. Wells, F. B. Gertler and D. A. Lauffenburger, *J. Cell Biol.*, 2012, 197, 721–729.

10 N. E. Reticker-Flynn, D. F. B. Malta, M. M. Winslow, J. M. Lamar, M. J. Xu, G. H. Underhill, R. O. Hynes, T. E. Jacks and S. N. Bhatia, *Nat. Commun.*, 2012, 3, 1122.

11 H. Kennecke, R. Yerushalmi, R. Woods, M. C. U. Cheang, D. Voduc, C. H. Speers, T. O. Nielsen and K. Gelmon, *J. Clin. Oncol.*, 2010, 28, 3271–3277.

12 S. Paget, *Lancet*, 1889, 133, 571–573.

13 T. Shibue and R. A. Weinberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, 106, 10290–10295.

14 D. Barkan, H. Kleinman, J. L. Simmons, H. Asmussen, A. K. Kamaraju, M. J. Hoenorhoff, Z. Liu, S. V Costes, E. H. Cho, S. Lockett, C. Khanna, A. F. Chambers and J. E. Green, *Cancer Res.*, 2008, 68, 6241–6250.

15 M. Abdel-Ghany, H.-C. Cheng, R. C. Elble and B. U. Pauli, *J. Biol. Chem.*, 2002, 277, 34391–34400.

16 H. L. Goel, T. Gritsko, B. Pursell, C. Chang, L. D. Shultz, D. L. Greiner, J. H. Norum, R. Toftgard, L. M. Shaw and A. M. Mercurio, *Cell Rep.*, 2014, 7, 747–761.

17 S. Takayama, S. Ishii, T. Ikeda, S. Masamura, M. Doi and M. Kitajima, *Anticancer Res.*, 2005, 25, 79–83.

18 H. Wang, W. Fu, J. H. Im, Z. Zhou, S. A. Santoro, V. Iyer, C. M. DiPersio, Q.-C. Yu, V. Quaranta, A. Al-Mehdi and R. J. Muschel, *J. Cell Biol.*, 2004, 164, 935–941.

19 B. Zhou, K. Gibson-Corley, M. Herndon, Y. Sun, E. Gustafson-Wagner, M. Teoh-Fitzgerald, F. Domann, M. Henry and C. Stipp, *Mol. Cancer Res.*, 2014, 12, 143–154.

20 B. Dumont, V. Castronovo, O. Peulen, N. Blétard, P. Clézardin, P. Delvenne, E. De Pauw, A. Turtoi and A. Bellahcène, *J. Proteome Res.*, 2012, 11, 2247–2260.

21 A. Naba, K. R. Clauser, S. Hoersch, H. Liu, S. Carr and R. O. Hynes, *Mol. Cell. Proteomics*, 2012, 11, M111.014647.

22 A. Malara, M. Currao, C. Gruppi, G. Celesti, G. Viarengo, C. Buracchi, L. Laghi, D. L. Kaplan and A. Balduini, *Stem Cells*, 2013, 32, 926–937.

23 S. Stier, Y. Ko, R. Forkert, C. Lutz, T. Neuhaus, E. Grünewald, T. Cheng, D. Dombkowski, L. M. Calvi, S. R. Rittling and D. T. Scadden, *J. Exp. Med.*, 2005, 201, 1781–1791.

24 E. Ruoslahti, *Glycobiology*, 1996, 6, 489–492.

25 A. Bignami, M. Hosley and D. Dahl, *Anat. Embryol.*, 1993, 188, 419–433.

26 P. R. A. Johnson, J. K. Burgess, P. A. Underwood, W. Au, M. H. Poniris, M. Tamm, Q. Ge, M. Roth and J. L. Black, *J. Allergy Clin. Immunol.*, 2004, 113, 690–696.

27 R. M. Neve, K. Chin, J. Fridlyand, J. Yeh, F. L. Baehner, T. Fevr, L. Clark, N. Bayani, J.-P. Coppe, F. Tong, T. Speed, P. T. Spellman, S. DeVries, A. Lapuk, N. J. Wang, W.-L. Kuo, J. L. Stilwell, D. Pinkel, D. G. Albertson, F. M. Waldman, F. McCormick, R. B. Dickson, M. D. Johnson, M. Lippman, S. Ethier, A. Gazdar and J. W. Gray, *Cancer Cell*, 2006, 10, 515–527.

28 A. Banerjee, Z.-S. Wu, P. Qian, J. Kang, V. Pandey, D.-X. Liu, T. Zhu and P. E. Lobie, *Breast Cancer Res.*, 2011, 13, R112.

29 N. Rucci, E. Ricevuto, C. Ficorella, M. Longo, M. Perez, C. Di Giacinto, A. Funari, A. Teti and S. Migliaccio, *Bone*, 2004, 34, 697–709.

30 S. M. Shafie and L. A. Liotta, *Cancer Lett.*, 1980, 11, 81–87.

31 M. Tarragona, M. Pavlovic, A. Arnal-Estapé, J. Urosevic, M. Morales, M. Guiu, E. Planet, E. González-Suárez and R. R. Gomis, *J. Biol. Chem.*, 2012, 287, 21346–21355.

32 B. Clinchy, A. Gazdar, R. Rabinovsky, E. Yefenof, B. Gordon and E. S. Vitetta, *Breast Cancer Res. Treat.*, 2000, 61, 217–228.

33 C. Sheridan, H. Kishimoto, R. K. Fuchs, S. Mehrotra, P. Bhat-Nakshatri, C. H. Turner, R. Goulet, S. Badve and H. Nakshatri, *Breast Cancer Res.*, 2006, 8, R59.

34 T. Yoneda, P. J. Williams, T. Hiraga, M. Niewolna and R. Nishimura, *J. Bone Miner. Res.*, 2001, 16, 1486–1495.

35 E. Bade and B. Nitzgen, *In Vitro Cell. Dev. Biol.: Anim.*, 1985, 21, 245–248.

36 M. M. Martino, P. S. Briquez, E. Güç, F. Tortelli, W. W. Kilarski, S. Metzger, J. J. Rice, G. A. Kuhn, R. Müller, M. A. Swartz and J. A. Hubbell, *Science*, 2014, 343, 885–889.

37 C. Kuperwasser, S. Dessain, B. E. Bierbaum, D. Garnet, K. Sperandio, G. P. Gauvin, S. P. Naber, R. A. Weinberg and M. Rosenblatt, *Cancer Res.*, 2005, 65, 6130–6138.

38 R. Zhang, I. Fidler and J. Price, *Invasion Metastasis*, 1991, 11, 204–215.

39 C. M. Ghajar, H. Peinado, H. Mori, I. R. Matei, K. J. Evason, H. Brazier, D. Almeida, A. Koller, K. a Hajjar, D. Y. R. Stainier, E. I. Chen, D. Lyden and M. J. Bissell, *Nat. Cell Biol.*, 2013, 15, 807–817.

40 D. E. Jenkins, Y. S. Hornig, Y. Oei, J. Dusich and T. Purchio, *Breast Cancer Res.*, 2005, 7, R444–R454.

41 M. Niepel, M. Hafner, E. A. Pace, M. Chung, D. H. Chai, L. Zhou, B. Schoeberl and P. K. Sorger, *Sci. Signaling*, 2013, 6, ra84.

42 J. S. Desgrosellier and D. A. Cheresh, *Nat. Rev. Cancer*, 2010, 10, 9–22.

43 R. O. Hynes, *Cell*, 2002, 110, 673–687.

44 O. T. Fackler and R. Grosse, *J. Cell Biol.*, 2008, 181, 879–884.

45 C. C. Park, H. Zhang, M. Pallavicini, J. W. Gray, F. Baehner, C. J. Park and M. J. Bissell, *Cancer Res.*, 2006, 66, 1526–1535.

46 M. Morini, M. Mottolese, N. Ferrari, F. Ghiorzo, S. Buglioni, R. Mortarini, D. M. Noonan, P. G. Natali and A. Albini, *Int. J. Cancer*, 2000, 87, 336–342.

47 M. H. Zaman, L. M. Trapani, A. L. Sieminski, D. MacKellar, H. Gong, R. D. Kamm, A. Wells, D. A. Lauffenburger and P. Matsudaira, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, 103, 10889–10894.

48 P. B. dos Santos, J. S. Zanetti, A. Ribeiro-Silva and E. I. C. Beltrão, *Diagn. Pathol.*, 2012, 7, 104.

49 R. Mukhopadhyay, R. Theriault and J. Price, *Clin. Exp. Metastasis*, 1999, 17, 325–332.

50 N. E. Ramirez, Z. Zhang, A. Madamanchi, K. L. Boyd, L. D. O'Rear, A. Nashabi, Z. Li, W. D. Dupont, A. Zijlstra and M. M. Zutter, *J. Clin. Invest.*, 2011, 121, 226–237.

51 X. H.-F. Zhang, Q. Wang, W. Gerald, C. A. Hudis, L. Norton, M. Smid, J. A. Foekens and J. Massagué, *Cancer Cell*, 2009, 16, 67–78.

52 W. Kozlow and T. A. Guise, *J. Mammary Gland Biol. Neoplasia*, 2005, 10, 169–180.

53 G. S. Karagiannis, T. Poutahidis, S. E. Erdman, R. Kirsch, R. H. Riddell and E. P. Diamandis, *Mol. Cancer Res.*, 2012, 10, 1403–1418.

54 T. E. Kruger, A. H. Miller, A. K. Godwin and J. Wang, *Crit. Rev. Oncol./Hematol.*, 2014, 89, 330–341.

55 C. L. Chaffer, N. D. Marjanovic, T. Lee, G. Bell, C. G. Kleer, F. Reinhardt, A. C. D'Alessio, R. A. Young and R. A. Weinberg, *Cell*, 2013, 154, 61–74.

56 A. V Taubenberger, V. M. Quent, L. Thibaudeau, J. A. Clements and D. W. Hutmacher, *J. Bone Miner. Res.*, 2013, 28, 1399–1411.

57 Y.-C. Lee, J.-K. Jin, C.-J. Cheng, C.-F. Huang, J. H. Song, M. Huang, W. S. Brown, S. Zhang, L.-Y. Yu-Lee, E. T. Yeh, B. W. McIntyre, C. J. Logothetis, G. E. Gallick and S.-H. Lin, *Mol. Cancer Res.*, 2013, 11, 405–417.

58 B. Felding-Habermann, T. E. O'Toole, J. W. Smith, E. Fransvea, Z. M. Ruggeri, M. H. Ginsberg, P. E. Hughes, N. Pampori, S. J. Shattil, A. Saven and B. M. Mueller, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, 98, 1853–1858.

59 H. Kato, Z. Liao, J. V Mitsios, H.-Y. Wang, E. I. Deryugina, J. A. Varner, J. P. Quigley and S. J. Shattil, *PLoS One*, 2012, 7, e46576.
60 L. Seguin, S. Kato, A. Franovic, M. F. Camargo, J. Lesperance, K. C. Elliott, M. Yebra, A. Mielgo, A. M. Lowy, H. Husain, T. Cascone, L. Diao, J. Wang, I. I. Wistuba, J. V. Heymach, S. M. Lippman, J. S. Desgrosellier, S. Anand, S. M. Weis and D. A. Cheresh, *Nat. Cell Biol.*, 2014, 16, 457–468.
61 C. C. Park, H. J. Zhang, E. S. Yao, C. J. Park and M. J. Bissell, *Cancer Res.*, 2008, 68, 4398–4405.
62 J. G. Parvani, A. J. Galliher-Beckley, B. J. Schiemann and W. P. Schiemann, *Mol. Biol. Cell*, 2013, 24, 3449–3459.
63 H. H. Truong, J. Xiong, V. P. S. Ghotra, E. Nirmala, L. Haazen, S. E. Le Dévédec, H. E. Balciolu, S. He, B. E. Snaar-Jagalska, E. Vreugdenhil, J. H. N. Meerman, B. van de Water and E. H. J. Danen, *Sci. Signaling*, 2014, 7, ra15.
64 L. Huck, S. M. Pontier, D. M. Zuo and W. J. Muller, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, 107, 15559–15564.
65 M. Zhu, M. Z. Lerum and W. Chen, *Langmuir*, 2012, 28, 416–423.
66 C. Gauchet, G. R. Labadie and C. D. Poulter, *J. Am. Chem. Soc.*, 2006, 128, 9274–9275.
67 M. Morpurgo, E. A. Bayer and M. Wilchek, *J. Biochem. Biophys. Methods*, 1999, 38, 17–28.
68 R Core Team, 2014.
69 W. E. Johnson, C. Li and A. Rabinovic, *Biostatistics*, 2007, 8, 118–127.