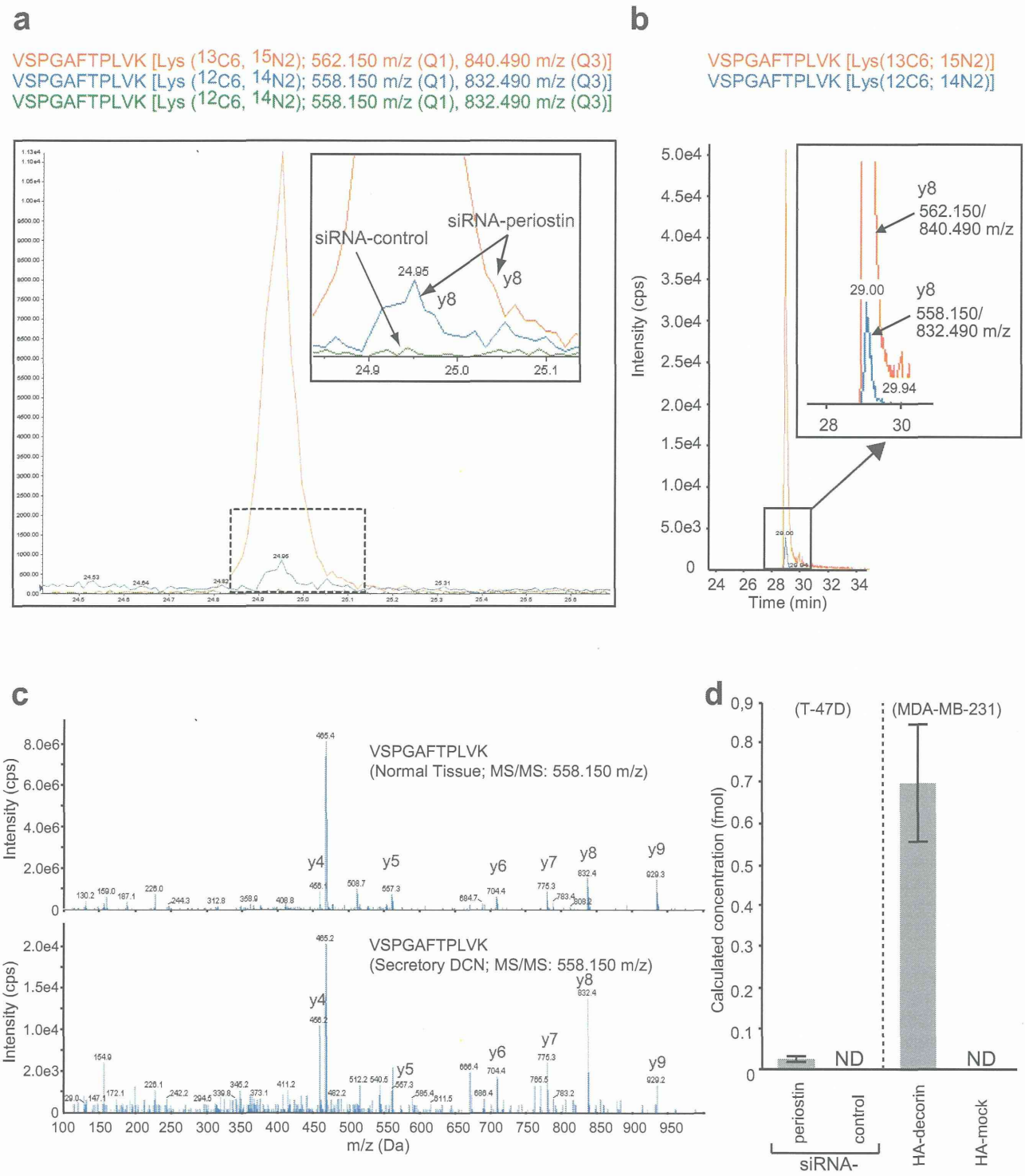

Figure 4 | The detection of secreted decorin in the culture medium using multiple reaction monitoring (MRM). (a) MRM chromatograms of VSPGAFTPLVK fragments and their standard (AQUA: red line) analogues. The peptides in culture medium from siRNA-periostin–treated cells (blue line) or siRNA-control–treated cells (green line) were analyzed using the MRM method. The doubly charged precursor mass was chosen as the Q1 mass, and the y8 fragment ion was chosen as the Q3 mass. Insets contain magnified views. (b) MRM chromatograms for VSPGAFTPLVK fragments and their standard (AQUA) analogues. MRM transitions for the endogenous (blue line) and standard (red line) peptides were monitored. For VSPGAFTPLVK, the doubly charged precursor mass was chosen as the Q1 mass, and the y8 fragment ions were chosen as Q3 mass. Insets contain magnified views. (c) MRM-triggered MS/MS product ion spectra obtained by nanoflow LC/MS/MS, comparing normal tissue digested with trypsin (upper) with decorin secreted from decorin-overexpressing MDA-MB-231 cells (lower). The spectrum of the peptide clearly shows y-ion fragments. (d) Calculated concentration of decorin in cell-culture medium of siRNA-periostin–treated T47D cells and decorin-overexpressing MDA-MB-231 cells (n = 3).

Knockdown of periostin and overexpression of decorin prevent cells motility and invasion. The data described above demonstrate that decorin is secreted from periostin-knockdown BT-20 cells, as well as from decorin-overexpressing BT-20 and MDA-MB-231 cells. We next investigated whether these cells could inhibit cell motility and invasion. To this end, we treated BT-20 cells with siRNA-periostin for 48 h. Each sample was subjected to SDS-PAGE, followed by immunoblot analysis with anti-decorin or anti-periostin antibody. β-actin was used as a loading control (Figure 5a). We then compared the proliferation of BT-20 cells treated with siRNA-periostin or siRNA-control. Cell number was measured using the water-soluble tetrazolium-1 (WST-1) assay. Relative fluorescence units (RFU) indicate the relative amount of proliferation. Column graphs show the means ± SEM of results from six samples. Periostin knockdown did not affect cell proliferation, as judged by the WST-1 assay (Supplementary Figure S6). Cell motility was measured in a wound-healing assay by time-lapse microscopy (Figure 5b, 5e, and 5g, left panel). Phase contrast was shown the images at start time (0 h) and 24 h. The dotted lines indicated cells at the start time, and white lines show the tips of migrated cells after 24 h. Bar graphs show the proportion of cell motility and means ± SEM from three samples (Figure 5b, 5e and 5g, right panel). Knockdown of periostin in cells led to a significant decrease in cell motility relative to cells treated with siRNA-control ($P = 0.016$, Student's T-test; $n = 6$) (Figure 5b). BT-20 cells transfected with HA-decorin exhibited significantly reduced motility relative to cells transfected with HA-mock vector ($P = 0.0003$, Student's T-test; $n = 3$) (Figure 5d and 5e). Similarly, cell invasion was measured using the CytoSelect 96-Well Collagen I Cell Invasion Assay (Cell Biolabs, San Diego, CA, USA). Invasive cells pass through the basement membrane layer, whereas noninvasive cells stay in the upper chamber. After removal of non-invasive cells, invading cells were stained and counted. Column graphs show means ± SEM of results from five samples. Cell invasion was also inhibited by knockdown of periostin ($P = 0.016$, Student's T-test; $n = 4$) (Figure 5c). In MDA-MB-231 cells, expression of HA-decorin had no effect on proliferation (Supplementary Figure S7). Expression of HA-decorin in these cells was confirmed by immunoblot analysis (Figure 5f). HA-decorin–expressing cells exhibited significantly reduced motility and invasion relative to cells transfected with HA-mock vector (motility; $P = 0.001$, Student's T-test; $n = 3$, invasion; $P = 0.032$, Student's T-test; $n = 6$) (Figure 5g and 5h). These data demonstrate that secreted decorin plays a functional role in promoting cancer cell motility and invasion.

Discussion

In this study, we characterized the interaction between cellular decorin and periostin not only in phyllodes tumors but also in BT-20 breast cancer cells. We detected secreted decorin in the culture medium of periostin-knockdown T-47D cells (Figure 4a and Supplementary Figure S4). Likewise, when decorin was overexpressed in MDA-MB-231 cells, in which decorin is not normally expressed, decorin was detected in the culture medium (Figure 4b and Supplementary Figure S5). These decorin-secreting cells inhibited cell motility and invasion more effectively than control cells (Figure 5).

Our findings demonstrate that periostin is more abundant in phyllodes tumors than in normal tissues (Figure 1, 2a, and 2b), and that it forms a complex with decorin (Figure 2e, 2f, 3b, and 3c). Previous work showed that decorin can delay tumor growth by blocking TGF-β[29], inhibiting inducers of angiogenesis such as VEGF[23], or interacting with E-cadherin[38]. On the other hand, knockdown or neutralization of endogenous periostin results in inhibition of cell migration and invasion[39,40], although the mechanism remains unclear. In this report, we describe a novel function of periostin: tethering of free decorin in the cytoplasm of cancer cells, thereby preventing release of decorin to the extracellular space. Our data obtained using both phyllodes tumors and breast cancer cells raise the possibility that knockdown of periostin in cancer cells may cause an effect similar to that of decorin produced by fibroblasts and myofibroblasts. Accordingly, we propose a model in which secretion of decorin is attributed to an inappropriate balance between the levels of decorin and periostin (Figure 5i).

As a component of the extracellular matrix, decorin prevents migration and invasion. Consistent with this, stromal decorin expression adjacent to malignant cells in invasive breast cancer tumors is significantly weaker than that in pure ductal carcinoma in situ (DCIS)[41]. In a previous study of breast cancer[42], an adenoviral vector containing a decorin transgene retarded primary tumor growth by 67% and greatly reduced pulmonary metastasis. Because secreted decorin inhibits cell motility and invasion, the results of our study suggest the importance of periostin as a potential therapeutic target in cancer cells that express both decorin and periostin. Our mechanistic studies demonstrated that siRNA knockdown of periostin abolishes the interaction with decorin, thereby increasing the level of decorin secreted from cancer cells.

Methods

Information about immunoblotting analysis, immunoprecipitation, knockdown of gene expression, total RNA extraction, RT-PCR, immunofluorescence and cell proliferation assay can be found in the Supplementary Methods.

Ethics statement. All human experiments were performed in accordance with the guidelines approved by the Ethics Committee of Tokyo Medical and Dental University (TMDU). The Institutional Review Board of TMDU approved the study, and written informed consent was obtained from each patient before surgery.

Patients and tissue samples. Tissue specimens analyzed in this study were obtained from 35 patients with phyllodes tumors and 37 patients with fibroadenomas. All patients underwent surgical resection in the Department of Breast Surgery at Tokyo Medical and Dental University, Japan, between March 2003 and August 2012. The clinical characteristics of the tumors we examined are summarized in Table 1. All specimens were formalin-fixed and paraffin-embedded (FFPE). Three tissue samples used for mass spectrometry were snap-frozen in liquid nitrogen and preserved at −80°C.

Cell lines. BT-20, T-47D, MDA-MB-231, and HeLa S3 cells were generously provided by the Japanese Foundation for Cancer Research. The media for each cell line are summarized in Supplementary Table S3. Cells were maintained at 37°C in a humidified atmosphere containing 5% CO_2. All cell lines were authenticated in December 2012.

iTRAQ labeling. Tumor and normal tissues were lysed in T-PER and centrifuged at 100,000 *g* at 4°C for 1 hour. Albumin and IgG were removed from the supernatants using the ProteoSeek™ Albumin/IgG Removal Kit (Thermo Scientific, Waltham, MA, USA), followed by concentration of the sample using Nanosep centrifugal devices (Pall, Ann Arbor, MI, USA). For mass spectrometry, 100 μg of protein lysate was reduced in 25 mM TCEP and 0.05% SDS for 60 min at 60°C, alkylated with methyl methanethiosulfonate for 10 min at RT, and then digested with trypsin at 37°C for 12–16 h. Desalted tryptic peptides were labeled with isobaric tags for relative quantification using iTRAQ reagents (AB SCIEX, Framingham, MA, USA). Briefly, peptides were dried and resuspended in 20 μl iTRAQ dissolution buffer. Trypsin-digested peptides isolated from tumor and normal tissue of three patients with phyllodes tumors were labeled with 2-plex iTRAQ tags; isobaric tags with m/z of 114 were added to normal tissue, and isobaric tags with m/z of 117 were added to tumor tissue (Supplementary Figure S1: cases 1–3). Samples were mixed, passed through a column with elution buffer (KCl: 10, 25, 50, 100, 175, and 350 mM), desalted on MonoSpin C18 columns (GL Science, Tokyo, Japan), and finally prepared for mass spectrometry. Six separate fractions were analyzed by mass spectrometry in at least three trials. The samples were separated by nanoflow liquid chromatography (300 nL/min) on a nano LC Dina-A system (KYA TECH Corp., Tokyo, Japan) in line with a Q-TRAP 5500 instrument (AB SCIEX) using a 75-min gradient of 5–100% acetonitrile in 0.1% formic acid.

Mass-spectrometry analysis of immunoprecipitates. Samples immunoprecipitated with antibodies against decorin or periostin were subjected to SDS-PAGE, and the gels were stained using the Silver Quest Staining Kit (Invitrogen). The stained gel bands were cut out and treated with dithiothreitol (DTT, Nacalai Tesque, Kyoto, Japan) dissolved in ammonium hydrogen carbonate (Nacalai Tesque), followed by treatment with iodoacetamide (Wako, Osaka, Japan). After the gels were dried, 20 μl of 0.05 pmol/μl trypsin (AB SCIEX) solution was applied to each gel piece and incubated for 12–16 h at 37°C to digest proteins. Digested peptides were extracted by washing the gel pieces twice with 50% trifluoroacetic acid (TFA, Wako), followed by

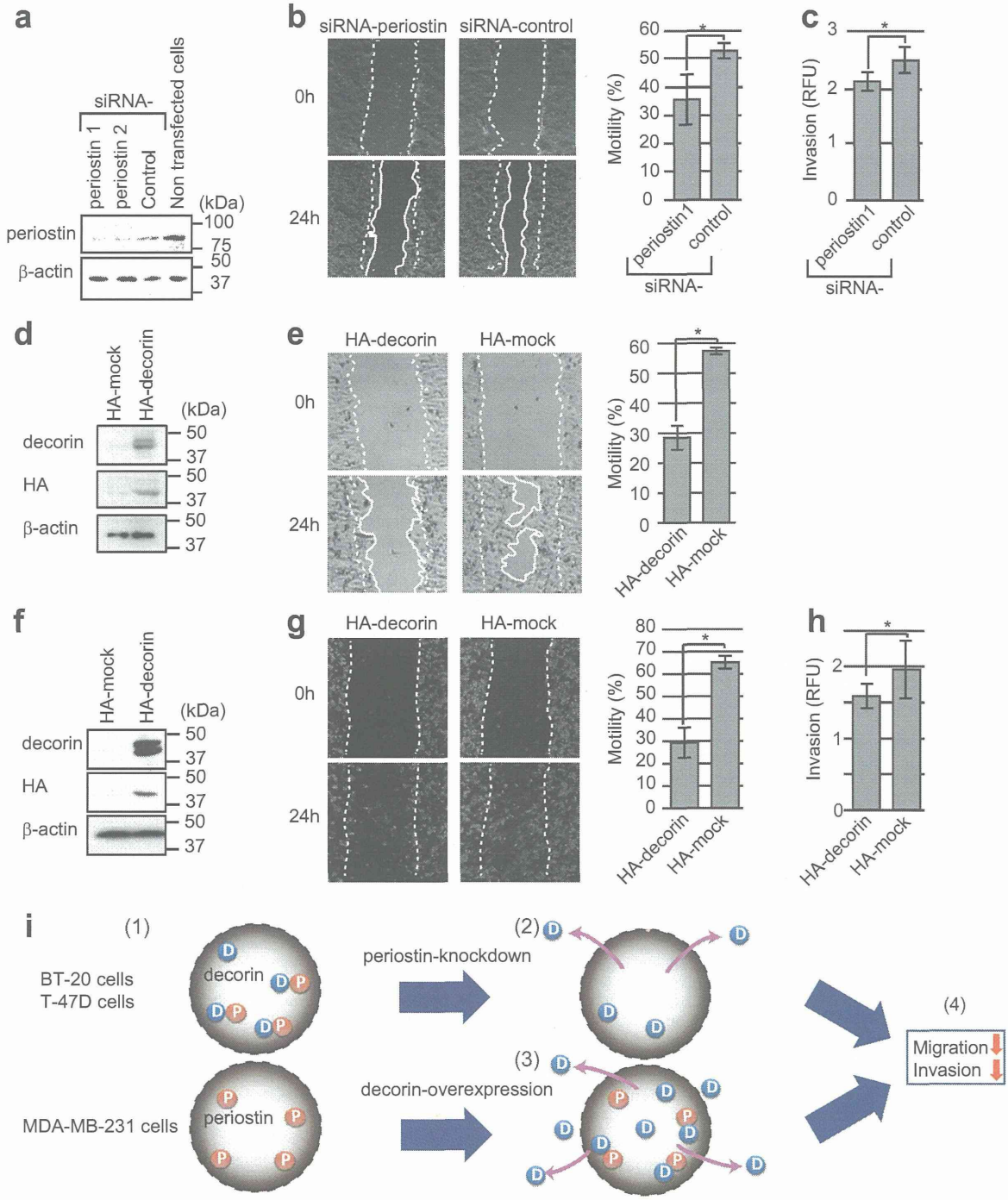

Figure 5 | Cell motility and invasion assay following knockdown of periostin or overexpression of decorin. (a) BT-20 cells were treated with siRNA-periostin for 48 h. Each sample was subjected to SDS-PAGE, followed by immunoblot analysis with anti-decorin or anti-periostin. β-actin was used as a loading control. (b) BT-20 cells were treated with siRNA-periostin for 48 h and then evaluated by the wound-healing assay. (c) BT-20 cells were treated with siRNA-periostin for 48 h, and then subjected to the cell invasion assay. (d) BT-20 cells expressing HA-decorin or HA-mock were subjected to SDS-PAGE followed by immunoblot analysis with anti-decorin or anti-HA antibodies. β-actin was used as a loading control. (e) BT-20 cells were transfected with HA-decorin or HA-mock expression vector for 24 h, and then subjected to the wound-healing assay. Column graphs show the means ± SEM of results from three samples. (f) MDA-MB-231 cells expressing HA-decorin or HA-mock were subjected to SDS-PAGE followed by immunoblot analysis with anti-decorin or anti-HA antibodies. β-actin was used as a loading control. (g) MDA-MB-231 cells were transfected with HA-decorin or HA-mock expression vector for 24 h, and then subjected to the wound-healing assay. (h) MDA-MB-231 cells were transfected with a HA-decorin expression vector for 24 h, and then subjected to the cell invasion assay. (b, c, e, g and h) P values were determined using Student's t-test. Asterisks indicate statistically significant differences ($P < 0.05$). (i) Potential roles of decorin and periostin in phyllodes tumor or breast cancer cell lines (BT-20, T-47D, and MDA-MB-231). (1) Decorin interacts with periostin in phyllodes tumor tissues and BT-20 cells. (2) Secreted decorin is detected in the culture medium of periostin-knockdown T-47D cells. (3) Transient expression of decorin in MDA-MB-231 cells leads to secretion of decorin into the culture medium. (4) Extracellular decorin significantly decreases cell motility and invasion.

Table 1 | Clinical characteristics of the tumors. Mean ages of patients with phyllodes tumor and fibroadenoma were 39.3 and 30.1 years, respectively. All patients were female

		phyllodes tumor	fibroadenoma
		n = 35	n = 37
Sex	Female	35	37
	Male	0	0
Age, y	–30	7	19
	31–40	12	13
	41–50	12	5
	51–	4	0
Operation	partial resection	25	36
	mastectomy	11	1
Grade	benign	16	-
	boarderline	14	-
	malignant	5	-
Tumor size(mm)		74.8(±60.5)	37.2(±29.6)
Cutaneous sympton	+	4	1
	-	31	36
Recurrence	+	6	0
	-	29	37
Mortality	dead	0	0
	alive	35	37

washing with 80% TFA. The purified peptide samples were injected onto a reversed-phase C18 column (HiQ sil C18W-3P, 3 μm, 120 Å; KYA TECH Corp.) and separated by nanoflow liquid chromatography (300 nL/min) on a nano LC Dina-A system (KYA TECH Corp.) in line with a Q-TRAP 5500 instrument (AB SCIEX) using a 75-min gradient of 5–100% acetonitrile in 0.1% formic acid.

Internal standardization with standard peptides. Standard peptides, VSPGAFTPLVK ($^{13}C_6$, $^{15}N_2$) and DLPPDTTLLDLQNNK ($^{13}C_6$, $^{15}N_2$), were purchased from Thermo Fisher Scientific (Ulm, Germany). The peptides were delivered in 5% (v/v) acetonitrile at a concentration of 5 pmol/μl.

Multiple reaction monitoring (MRM) analysis. Conditioned medium (CM) was concentrated 100-fold using a Vivaspin 20 (Sartorius Stedim, Göttingen, Germany). The protein concentration in CM was determined using the Bradford protein assay (Bio-Rad, Hercules, CA, USA). Peptides obtained from in-solution digestion of raw CM were analyzed by multiple reaction monitoring (MRM). The doubly charged precursor ion was chosen as the Q1 mass, and the most intense fragment ion from the precursor was chosen as the Q3 mass. The optimized instrument parameters and selected MRM transitions were tested by analyzing the endogenous protein and standard peptides. The samples were separated by nanoflow liquid chromatography (300 nl/min) on a nano LC Dina-A system in line with a Q-TRAP 5500 instrument using a 45-min gradient of 5–100% acetonitrile in 0.1% formic acid. Absolute quantitation was performed using the MultiQuant Software (AB SCIEX). To improve the accuracy of the quantitation, the "heavy" peptide was added into each biological sample at a fixed amount to act as the internal standard for these samples. The actual concentration of peptide in the biological sample was computed from the ratio of the endogenous peptide (light) to the added internal standard (heavy). The concentration C of the targeted endogenous peptide was calculated as

$$C_{peptide} = C_{labeledpeptide} \times \text{Area ratio} \quad (\text{peak area}_{peptide}/\text{peak area}_{labeledpeptide})$$

Immunohistochemical staining. Immunohistochemical staining was carried out by the streptavidin–biotin method using the Histofine SAB-PO kit (Nichirei Co., Tokyo, Japan). Sections (4 μm thick) were cut from each FFPE tissue block. After deparaffinization and rehydration, antigen retrieval treatment was carried out in a temperature-controllable microwave processor (MI-77; Azumaya Co., Tokyo, Japan) at 98°C for 20 min (decorin) or 30 min (periostin) in 10 mM sodium citrate buffer (pH 6.0). Endogenous peroxidase activity was blocked by incubating the sample in a solution of 3% hydrogen peroxide in absolute methanol for 15 min. Nonspecific binding was blocked by treating the slides with 5% EzBlock (including 10% normal goat serum) for 10 min at RT. For detection of decorin, sections were incubated with anti-decorin antibodies (1 : 1000 dilution), and then beam-irradiated with the MW processor at 27°C for 15 min. The Histofine SAB-PO kit was used for visualization. For detection of periostin, sections were incubated for 90 min at RT with anti-periostin antibodies (1 : 2000 dilution). The Histofine Simple Stain MAX PO (MULTI) kit (Nichirei Corp.) was used for visualization. Color development was carried out with DAB (0.02% 3,3′-diaminobenzidine tetrahydrochloride; Nichirei Corp) for 10 min at RT. The sections were then counterstained with 1% Mayer's hematoxylin.

Immunohistochemical evaluation. Immunostaining of decorin and periostin was analyzed under a light microscope. Evaluations of stromal decorin and periostin expression were performed around normal gland and tumor tissue. Digital images were analyzed semiquantitatively. The intensities of decorin and periostin signals were determined using the ImageJ software, according to the method described by Augoff et al[43]. Briefly, random areas at the periphery of lesions were captured as digital images (680 × 512 pixels) with a digital camera. For each digital image, the signal from 10 representative areas was digitized in grayscale ranging from 0 (white) to 255 (black), and these data were used to generate a histogram. Nuclei were omitted from this analysis. Stroma in the negative control samples (i.e., without primary antibody) was used as an internal control. The intensity of the decorin signal was standardized by subtracting the mean intensity of the internal control.

Wound-healing assay. Confluent cell monolayers were wounded (lightly scratched) with a pipet tip. After careful washing to remove detached cells, the cells were cultured for 24 h. Phase-contrast images were taken every 30 min for 24 h. The width of the wound was monitored using an FW4000-TZ time-lapse microscope (Leica).

Cell invasion assay. Cell invasion was measured using the CytoSelect 96-Well Collagen I Cell Invasion Assay (Cell Biolabs). Cells (5×10^5) were seeded in serum-free media onto polycarbonate membrane inserts (8 μm thick) whose upper surfaces were coated with a uniform layer of dried Bovine Type I Collagen matrix (Cell Biolabs). Inserts were then submerged in media containing 10% fetal bovine serum (FBS), and the cells were cultured for 24 h. Invading cells were stained with cell stain solution, followed by measurement using a Fluoroskan Ascent plate reader (Thermo Scientific) at 560 nm.

Statistical analysis. The Wilcoxon signed-rank test was used for comparisons of normal and tumor tissue within the phyllodes patients. Quantitative decorin and periostin stromal expression for comparison of phyllodes and fibroadenoma were analyzed using the Mann-Whitney U test. Student's t-test was used to evaluate the results of proliferation, migration, and invasion assays. P values < 0.05 were considered to indicate statistically significant differences. All statistical analyses were performed using the SPSS software (IBM, Armonk, NY, USA).

1. Gupta, G. P. & Massague, J. Cancer metastasis: Building a framework. *Cell* **127**, 679–695; DOI:10.1016/j.cell.2006.11.001 (2006).
2. Joyce, J. A. & Pollard, J. W. Microenvironmental regulation of metastasis. *Nature Rev. Cancer* **9**, 239–252; DOI:10.1038/nrc2618 (2009).
3. Shibue, T. & Weinberg, R. A. Metastatic colonization: Settlement, adaptation and propagation of tumor cells in a foreign tissue environment. *Semin. Cancer Biol.* **21**, 99–106; DOI:10.1016/j.semcancer.2010.12.003 (2011).
4. Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**, 85–U95; DOI:10.1038/nature10694 (2012).
5. Merline, R. *et al.* Signaling by the Matrix Proteoglycan Decorin Controls Inflammation and Cancer Through PDCD4 and MicroRNA-21. *Science Signaling* **4**, ra75; DOI:10.1126/scisignal.2001868 (2011).
6. Wu, S. *et al.* Silencing of periostin inhibits nicotine-mediated tumor cell growth and epithelial-mesenchymal transition in lung cancer cells. *Mol. Med. Rep.* **7**, 875–880; DOI:10.3892/mmr.2013.1267 (2013).
7. Ouyang, G. *et al.* Upregulated expression of periostin by hypoxia in non-small-cell lung cancer cells promotes cell survival via the Akt/PKB pathway. *Cancer Lett.* **281**, 213–219; DOI:10.1016/j.canlet.2009.02.030 (2009).
8. Bao, S. *et al.* Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. *Cancer Cell* **5**, 329–339; DOI:10.1016/S1535-6108(04)00081-9 (2004).
9. Tilman, G., Mattiussi, M., Brasseur, F., van Baren, N. & Decottignies, A. Human periostin gene expression in normal tissues, tumors and melanoma: evidences for periostin production by both stromal and melanoma cells. *Mol. Cancer* **6**, 80; DOI:10.1186/1476-4598-6-80 (2007).
10. Baril, P. *et al.* Periostin promotes invasiveness and resistance of pancreatic cancer cells to hypoxia-induced cell death: role of the beta(4) integrin and the PI3k pathway. *Oncogene* **26**, 2082–2094; DOI:10.1038/sj.onc.1210009 (2007).
11. Fluge, O., Bruland, O., Akslen, L., Lillehaug, J. & Varhaug, J. Gene expression in poorly differentiated papillary thyroid carcinomas. *Thyroid* **16**, 161–175 (2006).
12. Ismail, R. *et al.* Differential gene expression between normal and tumor-derived ovarian epithelial cells. *Cancer Res.* **60**, 6744–6749 (2000).
13. Shao, R. *et al.* Acquired expression of periostin by human breast cancers promotes tumor angiogenesis through up-regulation of vascular endothelial growth factor receptor 2 expression. *Mol. Cell. Biol.* **24**, 3992–4003; DOI:10.1128/MCB.24.9.3992-4003.2004 (2004).
14. Tischler, V. *et al.* Periostin is up-regulated in high grade and high stage prostate cancer. *BMC Cancer* **10**, 273; DOI:10.1186/1471-2407-10-273 (2010).
15. Siriwardena, B. S. M. S. *et al.* Periostin is frequently overexpressed and enhances invasion and angiogenesis in oral cancer. *Br. J. Cancer* **95**, 1396–1403; DOI:10.1038/sj.bjc.6603431 (2006).

16. Li, B., Wang, L. & Chi, B. Upregulation of periostin prevents P53-mediated apoptosis in SGC-7901 gastric cancer cells. *Mol. Biol. Rep.* **40**, 1677–1683; DOI:10.1007/s11033-012-2218-3 (2013).
17. Gillan, L. *et al.* Periostin secreted by epithelial ovarian carcinoma is a ligand for alpha(v)/beta(3) and alpha(v)beta(5) integrins and promotes cell motility. *Cancer Res.* **62**, 5358–5364 (2002).
18. Kikuchi, Y. *et al.* The niche component periostin is produced by cancer-associated fibroblasts, supporting growth of gastric cancer through ERK activation. *Am. J. Pathol.* **184**, 859–870. DOI: 10.1016/j.ajpath.2013.11.012 (2014).
19. Kikuchi, Y. *et al.* Periostin is expressed in pericryptal fibroblasts and cancer-associated fibroblasts in the colon. *J. Histochem Cytochem.* **56**, 753–764. DOI: 10.1369/jhc.2008.951061 (2008).
20. Leygue, E. *et al.* Lumican and decorin are differentially expressed in human breast carcinoma. *J. Pathol.* **192**, 313–320 (2000).
21. Ishiwata, T. *et al.* Role of lumican in cancer cells and adjacent stromal tissues in human pancreatic cancer. *Oncol. Rep.* **18**, 537–543 (2007).
22. Edwards, I. J. Proteoglycans in prostate cancer. *Nat. Rev. Urol.* **9**, 196–206; DOI:10.1007/s11033-012-2218-3 (2012).
23. Grant, D. *et al.* Decorin suppresses tumor cell-mediated angiogenesis. *Oncogene* **21**, 4765–4777; DOI:10.1038/sj.onc.1205595 (2002).
24. Keene, D. *et al.* Acceleracted publication - Decorin binds near the C terminus of type I collagen. *J. Biol. Chem.* **275**, 21801–21804; DOI:10.1074/jbc.C000278200 (2000).
25. Neame, P., Kay, C., McQuillan, D., Beales, M. & Hassell, J. Independent modulation of collagen fibrillogenesis by decorin and lumican. *Cell. Mol. Life Sci.* **57**, 859–863; DOI:10.1007/s000180050048 (2000).
26. Geng, Y., McQuillan, D. & Roughley, P. J. SLRP interaction can protect collagen fibrils from cleavage by collagenases. *Matrix Biol.* **25**, 484–491; DOI:10.1016/j.matbio.2006.08.259 (2006).
27. Santra, M., Reed, C. & Iozzo, R. Decorin binds to a narrow region of the epidermal growth factor (EGF) receptor, partially overlapping but distinct from the EGF-binding epitope. *J. Biol. Chem.* **277**, 35671–35681; DOI:10.1074/jbc.M205317200 (2002).
28. Wu, H. *et al.* Regulation of Intracellular Decorin via Proteasome Degradation in Rat Mesangial Cells. *J. Cell. Biochem.* **111**, 1010–1019; DOI:10.1002/jcb.22789 (2010).
29. Yamaguchi, Y., Mann, D. M. & Ruoslahti, E. Negative Regulation of Transforming Growth-Factor-Beta by the Proteoglycan Decorin. *Nature* **346**, 281–284; DOI:10.1038/346281a0 (1990).
30. Neill, T. *et al.* Decorin Antagonizes the Angiogenic Network: concurrent inhibition of Met, hypoxia inducible factor 1α, vascular endothelial growth factor A, and induction of thrombospondin-1 and TIMP3. *J. Biol. Chem.* **287**, 5492–5506; DOI:10.1074/jbc.M111.283499 (2012).
31. Honda, E. & Munakata, H. Purification and characterization of decorin from the culture media of MRC-5 cells. *Int. J. Biochem. Cell Biol.* **36**, 1635–1644; DOI:10.1016/j.biocel.2004.01.023 (2004).
32. Goldoni, S. *et al.* An antimetastatic role for decorin in breast cancer. *Am. J. Pathol.* **173**, 844–855; DOI:10.2353/ajpath.2008.080275 (2008).
33. Dumont, N. *et al.* Breast Fibroblasts Modulate Early Dissemination, Tumorigenesis, and Metastasis through Alteration of Extracellular Matrix Characteristics. *Neoplasia* **15**, 249–262 (2013).
34. Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nature Rev. Cancer* **6**, 392–401; DOI:10.1038/nrc1877 (2006).
35. Ji, J. *et al.* Comparative proteomic analysis of rat hepatic stellate cell activation: A comprehensive view and suppressed immune response. *Hepatology* **56**, 332–349; DOI: 10.1002/hep.25650. (2012).
36. Bao, S. *et al.* Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. *Cancer Cell* **5**, 329–339; DOI:10.1016/S1535-6108(04)00081-9 (2004).
37. Soltermann, A. *et al.* Prognostic Significance of Epithelial-Mesenchymal and Mesenchymal-Epithelial Transition Protein Expression in Non-Small Cell Lung Cancer. *Clin. Cancer Res.* **14**, 7430–7437; DOI:10.1158/1078-0432.CCR-08-0935 (2008).
38. Bi, X. *et al.* Decorin-mediated inhibition of colorectal cancer growth and migration is associated with E-cadherin in vitro and in mice. *Carcinogenesis* **33**, 326–330; DOI:10.1093/carcin/bgr293 (2012).
39. Michaylira, C. Z. *et al.* Periostin, a Cell Adhesion Molecule, Facilitates Invasion in the Tumor Microenvironment and Annotates a Novel Tumor-Invasive Signature in Esophageal Cancer. *Cancer Res.* **70**, 5281–5292; DOI:10.1158/0008-5472.CAN-10-0704 (2010).
40. Zhu, M. *et al.* Neutralizing Monoclonal Antibody to Periostin Inhibits Ovarian Tumor Growth and Metastasis. *Mol. Cancer Ther.* **10**, 1500–1508; DOI:10.1158/1535-7163.MCT-11-0046 (2011).
41. Oda, G. *et al.* Significance of stromal decorin expression during the progression of breast cancer. *Oncol. Rep.* **28**, 2003–2008; DOI:10.3892/or.2012.2040 (2012).
42. Reed, C. *et al.* Decorin prevents metastatic spreading of breast cancer. *Oncogene* **24**, 1104–1110; DOI:10.1038/sj.onc.1208329 (2005).
43. Augoff, K. *et al.* Expression of Decorin in Esophageal Cancer in Relation to the Expression of Three Isoforms of Transforming Growth Factor-Beta (TGF-1,-2, and-3) and Matrix Metalloproteinase-2 Activity. *Cancer Invest.* **27**, 443–452; DOI:10.1080/07357900802527221 (2009).

Acknowledgments

We thank all members of the Department of Molecular Genetics of Tokyo Medical and Dental University (TMDU) for valuable input on the manuscript.

Author contributions

All the authors contributing to this work have made the following declarations: Y.M., A.N. and T.I. conceived and designed the experiments. A.N. and T.I. performed the experiments. A.N. and T.I. analyzed the data. T.I. contributed reagents and materials tools. Y.M., A.N. and T.I. wrote the manuscript. M.N., T.N., T.S., H.U. and K.S. reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ishiba, T. *et al.* Periostin suppression induces decorin secretion leading to reduced breast cancer cell motility and invasion. *Sci. Rep.* **4**, 7069; DOI:10.1038/srep07069 (2014).

EMBO Molecular Medicine

Epithelial-Mesenchymal Transition Spectrum Quantification and Its Efficacy in Deciphering Survival and Drug Responses of Cancer Patients

Tuan Zea Tan, Qing Hao Miow, Yoshio Miki, Tetsuo Noda, Seiichi Mori, Ruby Yun-Ju Huang and Jean Paul Thiery

Corresponding author: Jean Paul Thiery, Institute of Molecular and Cell Biology

Review timeline:

Submission date:	28 April 2014
Editorial Decision:	12 June 2014
Revision received:	30 July 2014
Editorial Decision:	04 August 2014
Revision received:	07 August 2014
Accepted:	08 August 2014

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

Editor: Roberto Buccione

1st Editorial Decision 12 June 2014

Thank you for the submission of your manuscript to EMBO Molecular Medicine. We have now received comments from the Reviewers whom we asked to evaluate your manuscript.

We apologise that it has taken more time than we would have liked to return a decision, but unfortunately the Reviewers delivered their evaluations with some delay. I trust that the inevitable frustration due to the delay will be somewhat tempered by the fact that the Reviewers are all quite supportive and, in my opinion, offer valuable suggestions to improve your manuscript.

Reviewer 1 specifically acknowledges the significant usefulness of the scoring tool and in this respect would like you to provide an algorithm to allow others to readily apply the approach to their favourite system. S/he also would like you to verify whether the signatures perform well with a reduced set of genes. Finally, this Reviewer suggests that a parallel should be drawn with previously reported cancer stemness signatures.

Reviewer 2 raises interesting points with respect to pancreatic cancer (with reference to the alleged negative correlation between EMT and survival) and also suggests further analysis comparing mesenchymal vs. epithelial cancer types. These points appear very interesting and while to specifically address them might not be required, I believe that the information gained would be potentially very valuable and would increase the impact of your work. Reviewer 2 notes that some discrepancies between certain clinical samples and theoretically corresponding cell lines were not discussed. S/he also challenges the contention that EMT reversal leads to less metastases, mentions other points of interest and cautions against overstating the clinical applications of your tool.

Reviewer 3 would like more information on how the EMT signatures other than the breast cancer one were derived and validated. S/he also suggests verifying EMT status after patient stratification based on different subtypes of a given cancer. This Reviewer, similarly to Reviewer 2, also mentions the growing evidence that EMT reversal actually increases metastatic colonisation. Finally, Reviewer 3 suggests that overall and disease-free survival as well should be considered in your evaluations.

In conclusion, while publication of the paper cannot be considered at this stage, we would be pleased to consider a suitably revised submission, provided that the Reviewers' concerns are addressed as outlined above.

I look forward to receiving you revised manuscript as soon as possible.

***** Reviewer's comments *****

Referee #1 (Comments on Novelty/Model System):

The presented study provides a potentially useful tool to evaluate and objectify EMT signatures in primary cancers as well as cancer-derived tumor cells. Most importantly, it would substantially increase the impact of the study if the authors provide an easy to handle on line tool allowing researchers to rapidly assess EMT signatures in their biological samples of interest or primary patient material. This would also increase the impact of the study and improve signatures to be used in evaluating therapy regimes.

Referee #1 (Remarks):

Tan et al. developed a method to quantify the EMT status of tumors and tumor-derived cell lines by developing a generic EMT gene expression signature. Based on transcriptomics of ovarian, breast, bladder, lung, colorectal and gastric cancers Tan et al. defined a gene expression signature that discriminates between epithelial (low, negative EMT score), mesenchymal (high, positive EMT score), or intermediate phenotypes, which display a certain degree of plasticity. Consistent with the hypothesis that this generic EMT signature monitors the cellular phenotype, functional studies demonstrated increased EMT scores upon e.g. knockdown of CDH1 or overexpression of TWIST in epithelial cancer cells. By 'EMT scoring' clinical samples and cell lines from various cancers, the authors identified cancer type-specific EMT states with e.g. predominant epithelial phenotypes in colorectal cancers. Most strikingly, however, the correlation of primary signatures with those observed in cancer-derived cells revealed a striking correlation and appeared barely biased by stromal content in primary material. In contrast to previous suggestions, poor patient overall (OS) and disease-free survival DFS were not exclusively associated with mesenchymal, but also with the epithelial cancer phenotypes. Moreover, by correlating the EMT score with chemotherapeutic responses of cancer cell lines Tan et al. suggest that the EMT-score indicates prevalence in the chemo-sensitivity of certain tumors instead of a general resistance or reduced sensitivity of MC-like tumors. This suggestion was further evaluated by analyzing Paclitaxel sensitivity of ovarian cancer patients. Supporting the hypothesis of phenotype-specific responses, Paclitaxel increased OS and DFS of patients suffering from mesenchymal ovarian carcinomas, but decreased the prognosis for patients suffering from epithelial ovarian carcinomas.
The manuscript presents a highly interesting tool, that could substantially contribute to objectifying the evaluation of EMT phenotypes in vitro as well as malignancies. However, in its present form the scoring remains difficult to apply, involves too many genes whereas other potential markers are not considered and lacks an option to include the estimation and correlation with stemness phenotypes.

Major aspects to be addressed:
1) The authors should at least provide a easy to use algorithm allowing other researchers to evaluate the EMT-status of their biological system of choice by gene expression signatures derived by microarrays.
2) The authors should evaluate if their signatures performs equally well with a reduced set of genes analyzed. This would potentially allow evaluating EMT signatures based on smaller qRT-PCR arrays. At least the authors should identify the 'best performing' genes.

3) Along these lines it remains elusive why the authors did not include miRNAs as additional markers, since these were shown to be substantially deregulated in various cancer according to the EMT-status, for instance the miR-200 family.
4) The authors should attempt to assess the expression of putative CSC-markers reported for various cancers and try to correlate this with their EMT signature. In addition the authors have to correlate their signature with previously reported stemness signatures (e.g. Ben-Porath et al., 2008.)

Minor aspects:

1) Wherefrom were the data derived used for analyses in Figure E1? Where are the references? The same applies for Figure 1B, C. The data processing and experimental procedures are not or barely described in the methods section or in the figure legend.
2) Figure E2B should be included in the main figure section. I would prefer one main figure together with Figure 3B. Accordingly, Figure 3A should be one main figure.

Referee #2 (Comments on Novelty/Model System):

The reviewers wish to state that they are not able to evaluate in detail the soundness of the bioinformatics/statistics in the manuscript as they are not experts in that area. They were, however able to evaluate the importance and context of the question addressed, novelty and the biological background.

Referee #2 (Remarks):

The authors provide a universal EMT-signature that can be applied to several cancer types. They show that this tool allows classification of tumor samples or cell lines due to their epithelial or mesenchymal genetic profile. In doing so, the authors very well appreciate the limitations of in vitro cell line models and distinguish between gene signatures of primary tumors and cell lines. In the end, the generated tumor-specific EMT-signature is correlated to clinical outcome. The authors address one of the most pressing questions in the field, i.e., is there a general EMT-signature and how does it affect clinical behaviour in different entities. They do so in an unbiased and careful manner, thereby putting out a tool (the generic EMT-signature) that should be of broad interest and utility to a lot of researchers from different fields of inquiry.

A few minor issues might be of interest to address.

For the creation of the generic, cancer-specific EMT-signature, the authors use gene sets of different tissues (bladder, breast, colorectal, gastric, lung and ovarian) (Figure 2). Given the implications of EMT in pancreatic cancer (PC) and the emerging evidence for a negative correlation between EMT and the survival rates of patients with PC (Ref. 1-3), inclusion of PC-specific EMT-signatures should be taken into consideration.
In Figure 3A, the generic EMT score is applied to several cancer entities, including data belonging to primary tumors and cell lines. Although many clinical samples and cell lines of the same cancer type overlap in their Epi/Mes classification, there are also cases of discrepancy (bladder, ovarian, lung and prostate carcinomas). Unfortunately, the authors do not discuss these differences. One possibility could be that the EMT-score for cell lines is less reliable due to smaller cohorts. Alternatively, it would mean that cell lines indeed transform during in vitro culturing and do not always reflect the phenotype of the correspondent primary tumor.
In the text belonging to Figure 3A, the authors draw the conclusion that the presented analysis might help to predict the origin of the tumor. One should be careful with this presumption, since at least in breast cancer it has been shown that basal-like tumors arise from luminal/epithelial, rather than from basal/mesenchymal populations (Ref. 4, and 5).
Next, the authors investigate whether the generic EMT-signature has clinical implications (Figure 4). They find that the EMT status does not entirely correlate with clinical response in patients suffering from breast or ovarian cancer. It is most surprising that the distribution of Epi and Mes phenotypes is very similar between patients that show a good clinical response and those suffering

from a progressive disease. Together with the drug sensitivity data, this analysis shows that EMT should not be the only factor taken into consideration when deciding over treatment regimens. Nevertheless, the authors were able to show that mesenchymal ovarian cancers respond better to chemotherapy than their epithelial counterparts. To consolidate the ability of the EMT score to predict sensitivity to chemotherapy it would be necessary to assess additional data from several cancer types and maybe even compare cancer entities that tend to be more mesenchymal (e.g. colon) to the more epithelial ones (e.g.: liver or renal).
In the discussion, the authors mention the EMT score as a method to assess effectiveness of EMT reversion therapies, aiming to inhibit the metastatic tumor potential. While the idea that the EMT score can detect phenotypic transitions is uncontestable, stating that the reversion of the EMT process leads to less metastasis is not entirely correct. It was recently shown that reversion of EMT even promotes metastatic colonization (Ref. 6).
Taken altogether, EMT scoring can be very useful when classifying tumor entities, but one should be aware of its limitations when it comes to the choice of therapy and the prediction of clinical outcome. The authors therefore provide a versatile tool for tumor classification based on their EMT status, but its power in clinical applications should not be overrated.
1. Nakajima S. et al. N-cadherin expression and epithelial-mesenchymal transition in pancreatic carcinoma. Clin Cancer Res. 2004 Jun 15;10(12 Pt 1):4125-33.
2. Hotz B. et al. Epithelial to mesenchymal transition: expression of the regulators snail, slug, and twist in pancreatic cancer. Clin Cancer Res. 2007 Aug 15;13(16):4769-76.
3. Arumugam T. et al. Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer. Cancer Res. 2009 Jul 15;69(14):5820-8.
4. Lim E. et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009 Aug;15(8):907-13.
5. Molyneux G. et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell. 2010 Sep 3;7(3):403-17.
6. Tsai JH et al. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. Cancer Cell. 2012 Dec 11;22(6):725-36.

Referee #3 (Comments on Novelty/Model System):

Future cancer subtype analysis and metastasis-free survival analysis are strongly suggested.

Referee #3 (Remarks):

In this manuscript the authors report a novel EMT scoring method to quantitatively estimate the degree of EMT in cell lines and tumors. The authors first developed cancer-specific EMT signatures using transcriptomics of a large collection of cancer cell lines and clinical samples, and demonstrated that they exhibit good correlation with published cancer-specific EMT signatures. From the established cancer-specific EMT signatures, they further derived a generic EMT signature and a method for universal EMT scoring. The authors also applied this EMT scoring to examine its efficacy in correlating EMT status with patient survival rates and responses to treatment.

This is an interesting study that is aims at developing methods to quantitatively assess the correlation between EMT status and cancer characteristics. The authors developed and applied a valid methodology, and the quality and explanation of data is adequate, making the conclusion robust. The paper is therefore of high scientific interest.

I suggest the following possible revisions to improve the quality of this manuscript:

1. Figure 1 and E1. The breast cancer-specific EMT signature is well explained and validated. However, it is not sure how some of the other cancer-specific EMT signatures are derived and verified, for example the ovarian and bladder cancers, as depicted in Figure 2A.

2. Figure 3 and E2. It is claimed by the authors that EMT status does not necessarily correlate with poorer survival. Did the authors try performing the analyses after stratifying patients based on different subtypes of the same cancer? For example, will performing Kaplan-Meier analysis by

subtypes of tumors (e.g. luminal, basal, ERBB2+, triple-negative subtypes of breast cancer) give a different result?

3. Figure 3 and E2. Given the critical roles of EMT in tumor cell dissemination and cancer metastasis, it is suggested that the authors also check metastasis-free survival, not only OS and DFS.

4. There is growing evidence suggesting that the reversal of EMT, the mesenchymal-epithelial transition (MET), may be necessary for efficient metastatic colonization. The authors should discuss this point and the potential complications in the interpretation of their results.

Minor concerns:

1. What are the values of the y axis in Figure 1C (top panel)?

1st Revision - authors' response 30 July 2014

Referee #1 (Comments on Novelty/Model System):

The presented study provides a potentially useful tool to evaluate and objectify EMT signatures in primary cancers as well as cancer-derived tumor cells. Most importantly, it would substantially increase the impact of the study if the authors provide an easy to handle on line tool allowing researchers to rapidly assess EMT signatures in their biological samples of interest or primary patient material. This would also increase the impact of the study and improve signatures to be used in evaluating therapy regimes.

Referee #1 (Remarks):

Tan et al. developed a method to quantify the EMT status of tumors and tumor-derived cell lines by developing a generic EMT gene expression signature. Based on transcriptomics of ovarian, breast, bladder, lung, colorectal and gastric cancers Tan et al. defined a gene expression signature that discriminates between epithelial (low, negative EMT score), mesenchymal (high, positive EMT score), or intermediate phenotypes, which display a certain degree of plasticity. Consistent with the hypothesis that this generic EMT signature monitors the cellular phenotype, functional studies demonstrated increased EMT scores upon e.g. knockdown of CDH1 or overexpression of TWIST in epithelial cancer cells. By 'EMT scoring' clinical samples and cell lines from various cancers, the authors identified cancer type-specific EMT states with e.g. predominant epithelial phenotypes in colorectal cancers. Most strikingly, however, the correlation of primary signatures with those observed in cancer-derived cells revealed a striking correlation and appeared barely biased by stromal content in primary material. In contrast to previous suggestions, poor patient overall (OS) and disease-free survival DFS were not exclusively associated with mesenchymal, but also with the epithelial cancer phenotypes. Moreover, by correlating the EMT score with chemotherapeutic responses of cancer cell lines Tan et al. suggest that the EMT-score indicates prevalence in the chemo-sensitivity of certain tumors instead of a general resistance or reduced sensitivity of MC-like tumors. This suggestion was further evaluated by analyzing Paclitaxel sensitivity of ovarian cancer

patients. Supporting the hypothesis of phenotype-specific responses, Paclitaxel increased OS and DFS of patients suffering from mesenchymal ovarian carcinomas, but decreased the prognosis for patients suffering from epithelial ovarian carcinomas.

Remark:

The manuscript presents a highly interesting tool, that could substantially contribute to objectifying the evaluation of EMT phenotypes in vitro as well as malignancies. However, in its present form the scoring remains difficult to apply, involves too many genes whereas other potential markers are not considered and lacks an option to include the estimation and correlation with stemness phenotypes.

We appreciate the reviewer's positive comments regarding our tool. We agree with the Referee#1 that the tool in its current form is relatively costly and difficult to apply, especially in the clinical setting. We have carefully considered the three limitations of our scoring system outlaid by Referee#1 and have provided a detailed response to each in the respective comments: i) involves too many genes (Comment 2); ii) Other potential markers are not considered (Comment 3); iii) Lacks an option to include the estimation and correlation with stemness (Comment 4).

Major aspects to be addressed:

Comment 1

The authors should at least provide an easy to use algorithm allowing other researchers to evaluate the EMT-status of their biological system of choice by gene expression signatures derived by microarrays.

Response: We agree with the Referee that it would be more useful if an online tool for EMT scoring is available. However, as we are in the process of patent filing (patent publication number: WO2013043132 A1; http://www.google.com/patents/WO2013043132A1?cl=en), we are not able to make the tool online at this moment. However, researchers can approach us through a temporary link http://www.csi.nus.edu.sg/bioinfo/index.php for computation of EMT score either using our signature or their own in a blinded manner without the need to login. Alternatively, researchers can request for the Matlab script upon signing of a Material Transfer Agreement (MTA) with our Institution. Details for such requests are given in Materials and Methods under 'Computation of EMT score'. The current script implementing the algorithm takes input formats from GSEA: GCT (for input gene expression data) and GMT (for gene signature). Details of the format can be obtained from the following website: http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats. We are currently preparing a follow-up paper in requirement of patent filing, in which we will make the online tool available once we are granted such clearance.

Comment 2

The authors should evaluate if their signatures perform equally well with a reduced set of genes analyzed. This would potentially allow evaluating EMT signatures based on smaller qRT-PCR arrays. At least the authors should identify the 'best performing' genes.

Response: We agree with Referee#1 that the tool will be more attractive in terms of cost and utility, especially in the clinical setting, if the number of genes could be reduced. Indeed, we are in the process of patent filing and are working on a follow-up paper using a reduced number of genes for diagnostic purposes (publication number: WO2013043132 A1). One possible way to identify the best-performing genes is based on the weights of differentially expressed genes between epithelial and mesenchymal states given in Table E1A. However, as pointed out by the reviewer, it is important that the reduced set will perform equally well. Thus, we have studied the effect of reducing the number of genes in the generic EMT signature in terms of correlating EMT scores between a reduced and the full EMT signature in the manuscript. We performed the analysis for both tumors and cell lines. We compared the overall concordance in tumor EMT phenotype estimated by the full and reduced EMT signatures, and re-validated the EMT scoring on the reduced set using the same set of functional intervention dataset as in Fig. E3 and Table E3 of the revised manuscript. The EMT score computed from the reduced set has been added to Table E3 and E4A. The analysis and result are given in the Expanded View section 'Effect of reducing genes in generic EMT signature' and in Fig. E5.

Comment 3

Along these lines it remains elusive why the authors did not include miRNAs as additional markers, since these were shown to be substantially deregulated in various cancer according to the EMT-status, for instance the miR-200 family.

Response: We agree with Referee#1 that including miRNAs in the signature might allow for the more precise quantification of EMT. However, as we seek to quantify EMT in various cancers, and because the development of our signature and scoring system was based on a relatively large number of cell lines and tumors, the amount of miRNA data is not as abundantly available as mRNA gene expression data. This prompted us instead to focus only on mRNA gene expression data. Furthermore, since our EMT scoring system is based on a generic method two-sample Kolmogorov-Smirnov test, the EMT signature is therefore customizable. The present generic EMT signature could be further refined or modified. In order to identify potential miRNA markers, we gathered 6 datasets where miRNA and mRNA expression data are available from GEO (bladder, GSE40355; pancreas, GSE32688; prostate, GSE21034; breast, ovarian, TCGA; multiple myeloma, GSE17498) (The Cancer Genome Atlas, 2012; The Cancer Genome Atlas Research, 2011;

Donahue *et al*, 2012; Hecker *et al*, 2013; Lionetti *et al*, 2009; Taylor *et al*, 2010), and correlated generic EMT score with miRNA expression. Looking at miRNA implicated in EMT (Hao *et al*, 2014; Zhang & Ma, 2012), we found that the miR-200 family members (miR-200a, miR-200b, miR-200c, miR-141, miR-429) negatively correlate with the generic EMT score (Fig. E2). However, we also noted that the miRNA involved in promoting or suppressing EMT might be cancer-dependent, as the correlation or anti-correlation were not consistent in all cancers or with previous reports (Hao *et al*, 2014; Zhang & Ma, 2012). This discrepancy may also stem from the platform-specific, cross-hybridization problem, as miRNA are short, closely related sequences (between family members), and hence measuring their expression is technically challenging (Mestdagh *et al*, 2014). Nevertheless, the expression of miR-200 and miR-34 (miR-34a, miR-34b, miR-34c) families were relatively consistent across bladder, breast, pancreas, prostate, and ovarian cancers; hence, scoring EMT using an miRNA-based EMT signature would complement and may allow for a more precise estimate of EMT when used in concert with an mRNA-based EMT signature. It is important to note that this assessment is preliminary, as only small cohorts were analysed. An in-depth investigation will be only possible when more data with both miRNA and mRNA gene expression data become available. We have included and discussed the result of the analysis in Fig. E2 and in the Expanded View section 'Generic EMT signature and miRNA' in the revised manuscript.

Comment 4

The authors should attempt to assess the expression of putative CSC-markers reported for various cancers and try to correlate this with their EMT signature. In addition the authors have to correlate their signature with previously reported stemness signatures (e.g. Ben-Porath et al., 2008.)

Response: Because EMT is associated with the acquisition of stemness (Frisch *et al*, 2013; Huang *et al*, 2012; Tam & Weinberg, 2013; Thiery *et al*, 2009), as suggested by the Referee, we have assessed the correlation of the generic EMT score and stemness using 21 published stem cell gene sets found in the Molecular Signature Database (Msigdb v4.0)(Subramanian *et al*, 2005), as well as those published stem cell markers (Medema, 2013). The enrichment score from ssGSEA (Verhaak *et al*, 2013) of the 21 stem cell related gene sets (*Beier Glioma stem cell*, *Ben-Porath embryonic stem cell v1*, *Ben-Porath embryonic stem cell v2*, *Ben-Porath NANOG targets*, *Ben-Porath NOS targets*, *Ben-Porath OCT4 targets*, *Ben-Porath SOX2 targets*, *Bhattacharya embryonic stem cell*, *BIOCARTA stem pathway*, *Boquest stem cell cultured vs fresh*, *Boquest stem cell*, *Conrad stem cell*, *Gal leukemic stem cell*, *Gentles leukemic stem cell*, *Hoebeke lymphoid stem cell*, *Jaatinen hematopoietic stem cell*, *Lee neural crest stem cell*, *Oswald hematopoietic stem cell in collagen gel*, *Pece mammary stem cell*, *Wong embryonic stem cell core*, *Yamashita liver cancer stem cell*) were computed and correlated with the EMT score for each cancer type.

The generic EMT score does not correlate universally with stemness in all cancer types. It is likely that different cancers employ different programs for expressing stem cell properties. Positive correlations between generic EMT score and stem cell signatures such as *Boquest stem cell*, *Pece mammary stem cell* or gene expression of stem cell markers such as *CD44* and *CXCR4* were observed in the majority of the cancer types. However, this analysis is preliminary due to certain limitations: the stem cell signatures were derived from different cell types, and no consideration was given to the different types of stem cells. For example, in breast cancer, there exist at least two types of stem cells (Liu *et al*, 2014). Nevertheless, we added this data into the Results section in the Expanded View, Fig. E11 and commented on it in the Discussion section of the main text.

Minor aspects:

1) Wherefrom were the data derived used for analyses in Figure E1? Where are the references? The same applies for Figure 1B, C. The data processing and experimental procedures are not or barely described in the methods section or in the figure legend.

Response: The data in Figures 1B, 1C and E1 (E3 in the revised manuscript) are from various datasets. The source of the downloaded dataset and the associated references were given in Tables E3 and E8. As pointed out by Referee#1, we now mention the source of the dataset and cite references in the figure legends of Figures 1B, 1C and E3. The data processing procedure and software used were given in 'Section Data preprocessing for Affymetrix microarray expression data'. Since the experimental procedures of these public datasets were described in their publications, and many datasets were involved, we did not duplicate these descriptions in the Materials and Methods. We instead provided the references in Tables E6 and E8.

2) Figure E2B should be included in the main figure section. I would prefer one main figure together with Figure 3B. Accordingly, Figure 3A should be one main figure.

Response: We have re-arranged the figures such that Figure E2B and Figure 3B form the new Figure 4 in the revised manuscript, and Figure 3A has been changed to the new Figure 3 in the revised manuscript, as suggested by Referee#1.

Referee #2 (Comments on Novelty/Model System):

The reviewers wish to state that they are not able to evaluate in detail the soundness of the bioinformatics/statistics in the manuscript as they are not experts in that area. They were, however able to evaluate the importance and context of the question addressed, novelty and the biological background.

Referee #2 (Remarks):

The authors provide a universal EMT-signature that can be applied to several cancer types. They show that this tool allows classification of tumour samples or cell lines due to their epithelial or mesenchymal genetic profile. In doing so, the authors very well appreciate the limitations of in vitro cell line models and distinguish between gene signatures of primary tumours and cell lines. In the end, the generated tumour-specific EMT-signature is correlated to clinical outcome. The authors address one of the most pressing questions in the field, i.e., is there a general EMT-signature and how does it affect clinical behaviour in different entities. They do so in an unbiased and careful manner, thereby putting out a tool (the generic EMT-signature) that should be of broad interest and utility to a lot of researchers from different fields of inquiry.

> *Response:* We thank Referee#2 for supporting the importance of the tool and for raising several issues critical to the improvement of our manuscript. We have provided detailed responses to each of the issues raised in the following sections.

A few minor issues might be of interest to address.

Comment 1:

For the creation of the generic, cancer-specific EMT-signature, the authors use gene sets of different tissues (bladder, breast, colorectal, gastric, lung and ovarian) (Figure 2). Given the implications of EMT in pancreatic cancer (PC) and the emerging evidence for a negative correlation between EMT and the survival rates of patients with PC (Ref. 1-3), inclusion of PC-specific EMT-signatures should be taken into consideration.

Ref. 1. Nakajima S. et al. N-cadherin expression and epithelial-mesenchymal transition in pancreatic carcinoma. Clin Cancer Res. 2004 Jun 15;10(12 Pt 1):4125-33.
Ref. 2. Hotz B. et al. Epithelial to mesenchymal transition: expression of the regulators snail, slug, and twist in pancreatic cancer. Clin Cancer Res. 2007 Aug 15;13(16):4769-76.
Ref. 3. Arumugam T. et al. Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer. Cancer Res. 2009 Jul 15;69(14):5820-8.

> *Response:* We thank Referee#2 for the references and for pointing out the implications of EMT in pancreatic cancer. We are in complete agreement with Referee#2 that EMT could play a role in pancreatic cancer. Indeed, in Fig. 3 of the revised manuscript, we observed an EMTed pancreatic cancer population in both tumors and cell lines. As we are seeking a

generic EMT signature that is common amongst the different cancer types and since Figure E3 in the revised manuscript shows that EMT scoring is capable of accurately predicting EMT status of pancreatic cancer cell lines, we believe the present EMT scoring is also applicable to pancreatic cancer. In addition, we showed in Table E1C that this generic EMT scoring is highly correlated with cancer-specific EMT. Hence, we opine that it is highly likely that incorporating a pancreatic-specific EMT signature would not change significantly the present generic EMT signature. Thus, we did not repeat the procedure to generate a pancreatic cancer-specific EMT signature or generic EMT signature, but rather, we performed additional analyses validating the applicability of this generic EMT signature in pancreatic cancer and added the data to the Results section, Expanded View and Fig. E4 of the revised manuscript. Furthermore, we have included two recent pancreatic datasets that have overall survival information, and have studied the correlation of EMT and survival in pancreatic cancer as well (Fig. 4A in revised manuscript). Even though it is not significant, there is a trend that EMT correlates with poorer overall survival, consistent with previously published results (Arumugam *et al*, 2009; Hotz *et al*, 2007; Nakajima *et al*, 2004).

Comment 2.1:

In Figure 3A, the generic EMT score is applied to several cancer entities, including data belonging to primary tumours and cell lines. Although many clinical samples and cell lines of the same cancer type overlap in their Epi/Mes classification, there are also cases of discrepancy (bladder, ovarian, lung and prostate carcinomas). Unfortunately, the authors do not discuss these differences. One possibility could be that the EMT-score for cell lines is less reliable due to smaller cohorts. Alternatively, it would mean that cell lines indeed transform during in vitro culturing and do not always reflect the phenotype of the correspondent primary tumour.

Response: We thank Referee#2 for pointing out this oversight. We agree with Referee#2 that the EMT score distribution of clinical samples and cell lines of the same cancer type were largely similar in most cases but that there were also discrepancies in terms of EMT score mean and range in some cases such as ovarian and prostate cancers. Although it is true that cell lines acquire additional transformations in culture, we believe the discrepancies stemmed from the differences in the number of samples, and that the data of cell lines and tumours were not paired. Hence, the different compositions for each cancer type such as histology, grade, and stage may have caused the difference in the distribution of the EMT scores. For example, in prostate cancer, the tumours were made up of primary and metastatic, recurrent and non-recurrent prostate carcinoma, whereas the 8 prostate cancer cell lines were derived mainly from metastatic sites (Table R1). We have provided this explanation of the discrepancies in the Results section in the revised manuscript.

Table R1: Prostate cell lines in Fig. 3

Cell Line	Cancer	Tissue Source	ATCC code/PMID
22RV1	prostate	prostate	CRL-2505
BPH-1	prostate	transurethral resection	7535634
DU 145	prostate	derived from metastatic site: brain	HTB-81
LNCaP clone FGC	prostate	derived from metastatic site: left supraclavicular lymph node	CRL-1740
MDA PCa 2b	prostate	derived from metastatic site: bone	CRL-2422
NCI-H660	prostate	derived from metastatic site: lymph node	CRL-5813
PC-3	prostate	derived from metastatic site: bone	CRL-1435
VCaP	prostate	derived from metastatic site: vertebral metastasis	CRL-2876

Comment 2.2:

In the text belonging to Figure 3A, the authors draw the conclusion that the presented analysis might help to predict the origin of the tumour. One should be careful with this presumption, since at least in breast cancer it has been shown that basal-like tumours arise from luminal/epithelial, rather than from basal/mesenchymal populations (Ref. 4, and 5).

Ref. 4. Lim E. et al. Aberrant luminal progenitors as the candidate target population for basal tumour development in BRCA1 mutation carriers. Nat Med. 2009 Aug;15(8):907-13.
Ref. 5. Molyneux G. et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell. 2010 Sep 3;7(3):403-17.

Response: Our intention was to say that the wide range of the EMT score observed within a cancer type might reflect the difference in cell of origin in the cancer. For example, in breast cancer, subtypes Luminal-A and -B originated from mature luminal cells (hence more Epi), whereas the Basal subtype originated from luminal progenitors (hence more Mes) (Lim *et al*, 2009). We did not mean to imply that EMT is predictive of the cell of origin. We have removed these sentences from the Results section to avoid this confusion in the revised manuscript.

Comment 3:

Next, the authors investigate whether the generic EMT-signature has clinical implications (Figure 4). They find that the EMT status does not entirely correlate with clinical response in patients suffering from breast or ovarian cancer. It is most surprising that the distribution of Epi and Mes phenotypes is very similar between patients that show a good clinical response and those suffering from a progressive disease. Together with the drug sensitivity data, this analysis shows that EMT should not be the only factor taken into consideration when deciding over treatment regimens. Nevertheless, the authors were able to show that mesenchymal ovarian cancers respond better to chemotherapy than their epithelial counterparts. To consolidate the ability of the EMT score to

predict sensitivity to chemotherapy it would be necessary to assess additional data from several cancer types and maybe even compare cancer entities that tend to be more mesenchymal (e.g. colon) to the more epithelial ones (e.g.: liver or renal).

Response: As per the suggestions of Referee#2, we have assessed additional data from 11 breast cancer cohorts (GSE48905, GSE33658, GSE23428, GSE22226, GSE18864, GSE28796, GSE16646, GSE22513, GSE4779, GSE18728, GSE50948)(Bauer *et al*, 2010; Carey *et al*, 2012; Esserman *et al*, 2012; Evans *et al*, 2012; Farmer *et al*, 2009; Knudsen *et al*, 2014; Korde *et al*, 2010; Lehmann *et al*, 2011; Massarweh *et al*, 2011; Prat *et al*, 2014; Silver *et al*, 2010), 3 colorectal cancer cohorts (GSE19862, GSE35452, GSE46862) (Gim J, 2014), a head & neck cancer cohort (GSE32877) (Tomkiewicz *et al*, 2012) and a melanoma cohort (GSE22968) (Beasley *et al*, 2011) where clinical response data was available. There is no survival data available in these cohorts. Interestingly, the data from the 11 breast cancer cohorts subjected to various treatments, including fulvestrant, anastrazole, carboplatin, doxorubicin, and other drugs (Fig. E8A), showed a similar distribution of Epi, intermediate and Mes breast cancers in each of the clinical response groups as we noted in Fig. 5A. Notably, the worst response group (progressive disease or residual disease) was made up of mostly Mes breast cancers. There is a trend of either an increasing proportion of Mes or a decreasing proportion of Epi toward chemoresistant groups. In the predominantly Epi colorectal and head & neck cancers and the predominantly Mes melanoma, because the distribution of EMT score does not allow us to segregate into Epi, intermediate and Mes groups, we instead investigated the EMT score profiles of responders and non-responders in these cancers (Fig. E8B). We found no significant difference between responders and non-responders in terms of EMT score, albeit there is a very slight trend that responders tend to have higher EMT score in predominantly Epi cancers (colorectal cancer), and a slight trend that responders tend to have a lower EMT score in predominantly Mes cancers (melanoma). We were not able to compare the results with that of the cell lines (Table E5) because the majority of these cohorts were treated with a combination of chemotherapeutics. Furthermore, as these data are from a relatively small cohort, further study is required to validate the current observation. These analyses have been added to the Results section, and the Expanded View of the revised manuscript.

On the other hand, we were not able to verify the findings of either the benefit of Mes tumour to paclitaxel in other cancer types because the data were not available. Instead, we found three datasets with survival and treatment information (glioma, GSE43388; ER+ breast cancer, GSE16391, and multiple myeloma, GSE9782) (Desmedt *et al*, 2009; Erdem-Eraslan *et al*, 2013; Mulligan *et al*, 2007). Interestingly, we observed a differential response in Epi and Mes tumours (Fig. E10). Even though, in general, patients with glioma receiving radiotherapy and chemotherapy have a better overall survival, the benefit is greater in

patients with a Mes glioma (p=0.0117). In contrast, patients with Epi multiple myeloma have a better disease-free survival rate when administered bortezomib instead of dexamethasone (p=0.0349). We observed no difference in ER+ tumours in patients administered with letrozole or tamoxifen in terms of EMT stratification. These data provide preliminary evidence of the potential benefit in stratifying patients by EMT status using generic EMT scoring. We concede that the cohorts were relatively small. However, we believe that our analysis would open the door to opportunities for incorporating the EMT scoring system into translational research protocols of clinical trials in the future when more data is available to assess the benefit of stratification by EMT.

Comment 4:

In the discussion, the authors mention the EMT score as a method to assess effectiveness of EMT reversion therapies, aiming to inhibit the metastatic tumor potential. While the idea that the EMT score can detect phenotypic transitions is uncontestable, stating that the reversion of the EMT process leads to less metastasis is not entirely correct. It was recently shown that reversion of EMT even promotes metastatic colonization (Ref. 6).

Ref. 6. Tsai JH et al. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. Cancer Cell. 2012 Dec 11;22(6):725-36.

Response: We thank the reviewer for raising this important point. Tsai et al. is a landmark paper and we fully agree with Tsai et al. as well as with Referees#2 and #3 that reversing EMT may promote metastatic colonization. This paradigm, to an extent, fits our previously proposed model where mesenchymal micro-metastases must re-acquire an epithelial phenotype to proliferate at the metastatic site (Thiery, 2002). The original text in the manuscript was not accurate ("the EMT reversion therapy as a means to reduce the metastatic potential") in that we are not postulating that EMT reversal therapy could *cure* metastasis, but instead we meant that EMT reversal therapy may *sensitize* cancer cells to specific drugs. Furthermore, our intent was not to achieve a full reversal of EMT—which, as pointed out by Referees#2, #3 and Tsai et al. could be detrimental—rather, we were referring to modifying the EMT status *along the spectrum* to improve patient responses to treatment. In our experience, the use of a single agent does not fully reverse EMT but it is sufficient to reduce invasion (unpublished), anoikis, spheroidogenesis and clonogenicity *in vitro* (Huang *et al*, 2013), and *in vivo* using orthotopic grafting of a human cancer line (Sim & Thiery, 2014; manuscript in preparation). The main challenge, however, is that we do not know exactly at which intermediate states and under what conditions or context cancer cells in the primary—or at distant sites—can exit dormancy and resume growth or develop chemo-resistance. This is a very complex problem requiring further investigations which are beyond the scope of the current study. We have amended the sentence to better reflect what we meant in the Discussion section of the revised manuscript.

Comment 5:

Taken altogether, EMT scoring can be very useful when classifying tumor entities, but one should be aware of its limitations when it comes to the choice of therapy and the prediction of clinical outcome. The authors therefore provide a versatile tool for tumor classification based on their EMT status, but its power in clinical applications should not be overrated.

> *Response:* We thank you the reviewer for supporting the utility of this tool, and we have re-written sentences in the manuscript regarding its clinical application such that they do not overstate the clinical utility of our tool in the revised manuscript.

Referee #3 (Comments on Novelty/Model System):

Future cancer subtype analysis and metastasis-free survival analysis are strongly suggested.

Referee #3 (Remarks):

In this manuscript the authors report a novel EMT scoring method to quantitatively estimate the degree of EMT in cell lines and tumors. The authors first developed cancer-specific EMT signatures using transcriptomics of a large collection of cancer cell lines and clinical samples, and demonstrated that they exhibit good correlation with published cancer-specific EMT signatures. From the established cancer-specific EMT signatures, they further derived a generic EMT signature and a method for universal EMT scoring. The authors also applied this EMT scoring to examine its efficacy in correlating EMT status with patient survival rates and responses to treatment.

This is an interesting study that is aims at developing methods to quantitatively assess the correlation between EMT status and cancer characteristics. The authors developed and applied a valid methodology, and the quality and explanation of data is adequate, making the conclusion robust. The paper is therefore of high scientific interest.

I suggest the following possible revisions to improve the quality of this manuscript:

Comment 1:

Figure 1 and E1. The breast cancer-specific EMT signature is well explained and validated. However, it is not sure how some of the other cancer-specific EMT signatures are derived and verified, for example the ovarian and bladder cancers, as depicted in Figure 2A.

> *Response:* Figure 1 depicted the procedure for deriving the cancer-specific EMT signature for the bladder, colorectal, gastric, and lung cancer-specific EMT signatures. Breast and