LETTERS

Genes that mediate breast cancer metastasis to the brain

Paula D. Bos[1], Xiang H.-F. Zhang[1], Cristina Nadal[1]†, Weiping Shu[1], Roger R. Gomis[1]†, Don X. Nguyen[1], Andy J. Minn[2], Marc J. van de Vijver[3], William L. Gerald[4], John A. Foekens[5] & Joan Massagué[1,6]

The molecular basis for breast cancer metastasis to the brain is largely unknown[1,2]. Brain relapse typically occurs years after the removal of a breast tumour[2–4], suggesting that disseminated cancer cells must acquire specialized functions to take over this organ. Here we show that breast cancer metastasis to the brain involves mediators of extravasation through non-fenestrated capillaries, complemented by specific enhancers of blood–brain barrier crossing and brain colonization. We isolated cells that preferentially infiltrate the brain from patients with advanced disease. Gene expression analysis of these cells and of clinical samples, coupled with functional analysis, identified the cyclooxygenase COX2 (also known as PTGS2), the epidermal growth factor receptor (EGFR) ligand HBEGF, and the α2,6-sialyltransferase ST6GALNAC5 as mediators of cancer cell passage through the blood–brain barrier. EGFR ligands and COX2 were previously linked to breast cancer infiltration of the lungs, but not the bones or liver[5,6], suggesting a sharing of these mediators in cerebral and pulmonary metastases. In contrast, ST6GALNAC5 specifically mediates brain metastasis. Normally restricted to the brain[7], the expression of ST6GALNAC5 in breast cancer cells enhances their adhesion to brain endothelial cells and their passage through the blood–brain barrier. This co-option of a brain sialyltransferase highlights the role of cell-surface glycosylation in organ-specific metastatic interactions.

Brain metastasis affects an estimated 10% of cancer patients with disseminated disease[2,8,9]. Even small lesions can cause neurological disability, and the median survival time of patients with brain metastasis is short. The two main sources of brain metastasis—adenocarcinomas of the lung or the breast—represent different models of the course of the disease. Metastasis from lung adenocarcinomas develops within months of diagnosis and affects several organs besides the brain[10]. This course suggests that aggressive pro-metastatic functions foster the colonization of several organs at once. In breast cancer, a long period of remission often precedes distant relapse[3,4], suggesting that breast cancer cells initially lack the full competence for outgrowth in distant organs but develop this under the selective pressure of different organ microenvironments. Breast cancer metastasis frequently becomes prevalent in one organ long before it does in others, and brain metastasis tends to be a late event[2]. The barriers to metastasis are distinct in different organs. Capillary endothelia are backed by a basement membrane in the lung[11] and also by tight junctions and astrocyte foot processes in the blood–brain barrier (BBB)[2,8], whereas the capillaries in the bone marrow and the liver are fenestrated[11,12]. The composition of the parenchyma also varies extensively between these organs. The protracted progression of disseminated cancer cells in different environments may give rise to metastatic speciation, as suggested by the coexistence of malignant cells with different organ tropisms in fluids from patients with advanced disease[5,13]. Analysis of such malignant cell populations has revealed genes that selectively mediate breast cancer metastasis to bones[13] or the lungs[5]. Here we adopted this approach to test the hypothesis that breast cancer infiltration of the brain requires general mediators of extravasation, complemented by specific enhancers of cell passage through the BBB.

We used oestrogen-receptor-negative (ER^-) pleural malignant cells from a Memorial Sloan-Kettering Cancer Center (MSKCC) breast cancer patient (CN34 sample), and also MDA-MB-231 cells (MDA231 for brevity)—an ER^- breast cancer pleural cell line previously used for the isolation of bone and lung metastatic cells[5,13] and brain metastatic cells[14]. CN34 and MDA231 cells were inoculated into the arterial circulation of immunodeficient female mice to isolate populations that target the brain (Fig. 1a–d). After tumour dissociation and expansion in culture, the resulting cell populations (brain metastatic derivative 1, BrM1) were subjected to a second round of *in vivo* selection, yielding BrM2 cell populations that showed a significant increase in brain metastatic activity (Fig. 1a, b). When grown as mammary tumours, CN34-BrM2 metastasized to brain in 42% (5 out of 12) of the mice, whereas parental CN34 mammary tumours yielded no brain metastases in ten mice. BrM2 cells showed no increase in bone or lung metastatic activity compared to the parental populations (Supplementary Table 1). MDA231 lung metastatic (LM2-4175) and bone metastatic (BoM-1833; refs 5, 13) derivatives were poorly metastatic to brain compared to BrM2 cells (Fig. 1b). The CN34-BrM2 and MDA231-BrM2 cell lines generated multifocal lesions in the cerebrum, the cerebellum and the brainstem (Fig. 1e, f and Supplementary Fig. 1a, b), and in the leptomeninges (Fig. 1g and Supplementary Fig. 1c, d). Larger nodules developed hemorrhagic cores and oedema (Fig. 1d). Astrogliosis occurred in the periphery of the tumours (Fig. 1h). All of these features are typical of brain metastasis in breast cancer patients[2,15]. Within 24 h of inoculation, BrM2 cells lodged in brain capillaries as single cells (Fig. 1i), suggesting that brain metastases resulted from an ability of these cells to breach the BBB.

Comparative genome-wide expression analysis demonstrated 243 genes that were overexpressed or underexpressed in the brain metastatic populations of both cell lines, or were upregulated in one cell system and overexpressed in the other, or were downregulated in one system and underexpressed in the other (Supplementary Table 2). To prioritize these candidate genes, we screened for those whose expression in breast tumours was associated with brain relapse. Univariate

[1]Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [2]Department of Radiation and Cellular Oncology, and Ludwig Center for Metastasis Research, The University of Chicago, Chicago, Illinois 60637, USA. [3]Department of Pathology, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. [4]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [5]Department of Medical Oncology, Erasmus MC Rotterdam, Josephine Nefkens Institute and Cancer Genomics Centre, Rotterdam, The Netherlands. [6]Howard Hughes Medical Institute, Chevy Chase, Maryland 20185, USA. †Present addresses: Institut de Malalties Hemato-Oncològiques, Hospital Clínic, 08036 Barcelona, Spain (C.N.); Oncology Programme, Institute for Research in Biomedicine, 08028 Barcelona, Spain (R.R.G.).

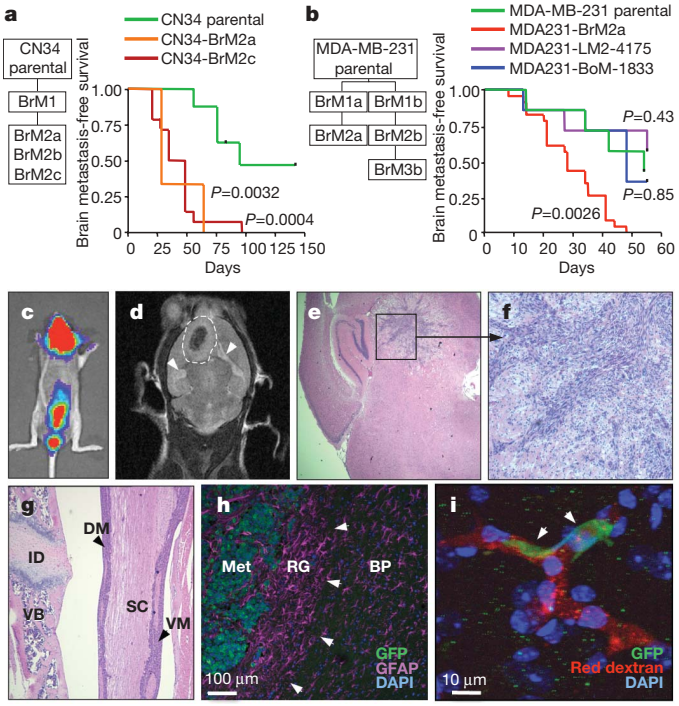

Figure 1 | Isolation and characterization of brain metastatic variants. **a**, **b**, Flowcharts of the *in-vivo*-selected brain metastatic derivatives, and Kaplan–Meier survival curves for brain metastasis-free survival of representative CN34 (parental $n = 8$, BrM2c $n = 14$, BrM2a $n = 3$) (**a**) and MDA231 (parental $n = 7$, LM $n = 7$, BoM $n = 7$, BrM2a $n = 23$) (**b**) cell line variants. A log-rank test was used to compare the survival curves of each cell line to the parental line. BoM, BrM and LM indicate bone, brain and lung metastatic derivative, respectively. **c**, Bioluminescence image of a mouse with brain and leptomeningeal metastasis by CN34-BrM2c cells. **d**, Magnetic resonance imaging (MRI) of a brain metastatic lesion (dashed line) showing a hemorrhagic core, and brain oedema (arrowheads). **e**, **f**, Representative haematoxylin and eosin (H&E)-stained sections of a mouse brain containing a CN34-BrM2c lesion (original magnification, ×2 (**e**) and ×10 (**f**)). **g**, H&E staining of a section showing MDA231-BrM2a cell colonization of the dorsal (DM) and ventral (VM) meninges. ID, intervertebral disc; SC, spinal cord; VB, vertebral body. Original magnification, ×5. **h**, MDA231-BrM2a brain metastatic lesion showing reactive glia (RG, arrowheads) around the metastatic lesion (Met). Tumour cells express green fluorescent protein (GFP), and glial cells are stained with the glial marker glial fibrillary protein (GFAP, purple). BP, brain parenchyma; DAPI, 4,6-diamidino-2-phenylindole. **i**, GPF^+ MDA231-BrM2a cells (arrowheads) arrested in brain capillaries (red, rhodamine dextran) 24 h after intracardiac injection into mice. Nuclei were stained with DAPI (blue).

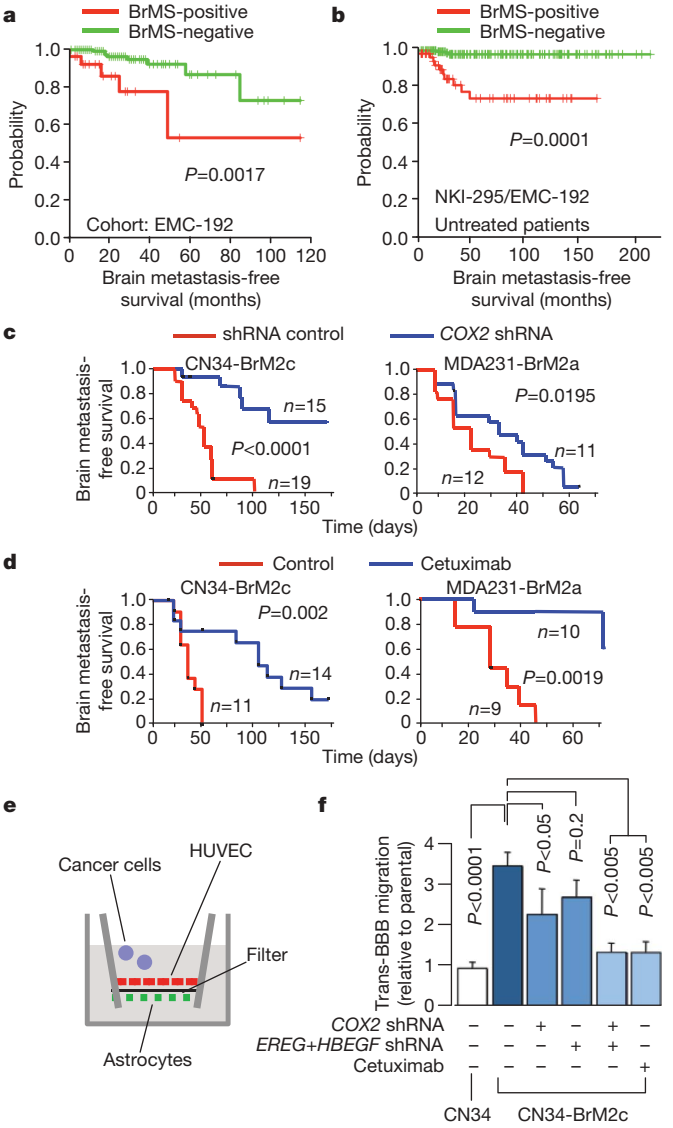

Figure 2 | COX2 and EGFR ligands as mediators of brain metastasis and BBB transmigration. **a**, **b**, Kaplan–Meier curves for brain metastasis-free survival on the basis of BrMS status in an independent cohort of 192 breast tumours (**a**), and in a combined cohort of 262 breast tumours from patients who received no adjuvant therapy (**b**). **c**, Kaplan–Meier curves for brain metastasis-free survival of mice injected with the indicated cell lines expressing short hairpin RNA (shRNA) vector control or shRNA targeting *COX2*. **d**, Kaplan–Meier curves for brain metastasis-free survival of mice injected with the indicated cell lines and treated with cetuximab or vehicle control. **e**, Schematic of the *in vitro* BBB model assay system. HUVEC, human umbilical vein endothelial cells. **f**, *In vitro* BBB transmigration activity of the indicated cell lines and conditions. The number of transmigrated cells relative to the parental cell lines is plotted. Error bars, s.e.m.; $n = 6$–20. *P* values were determined by log rank test (**a**–**d**) and one-tailed unpaired *t*-test (**f**).

analysis in a combined cohort of 368 clinically annotated breast tumours (MSK-82 and EMC-286 sets; Supplementary Table 3) showed 17 genes whose expression was correlated ($P < 0.05$) with brain relapse (Supplementary Table 4), and resembled the expression profile in the brain-metastatic-derived (BrM) cells (Supplementary Fig. 2a, b). A classifier trained with this brain metastasis gene set (BrMS) showed association with brain relapse in two independent breast tumour data sets (Fig. 2a and Supplementary Fig. 3a). The same procedures applied to randomly generated sets of 500 genes yielded no classifiers that performed in the data sets. The association of BrMS status with brain relapse remained significant within ER^- tumours (Supplementary Table 5 and Supplementary Fig. 3b), and was stronger in patients who received no adjuvant therapy ($P < 0.0001$; Fig. 2b and Supplementary Table 6). $BrMS^+$ tumours appeared in different molecular subtypes of breast cancer[16] (Supplementary Fig. 4a–c).

We do not interpret the results of the gene expression analysis as reflecting the only possible 17 genes associated with brain relapse. However, the expression of these 17 genes in breast tumours was not associated with relapse to bones, liver or lymph nodes (Supplementary Fig. 5a). Notably, six of these genes were shared with an 18-gene lung metastasis signature (LMS) that is associated with relapse to the lungs, but not to bones, liver or lymph nodes[17]. LMS^+ status was weakly associated with relapse to the brain, and $BrMS^+$ status with relapse to the lungs (Supplementary Fig. 5b–d). The shared genes include the prostaglandin-synthesizing enzyme cyclooxygenase-2 (*COX2*), which promotes extravasation in the lungs[6]; collagenase-1 (*MMP1*), which mediates invasion and extravasation[6,18]; angiopoietin-like 4 (*ANGPTL4*), which is induced by tumour-derived TGF-β and disrupts endothelial junctions[19]; latent TGF-β-binding protein (*LTBP1*), which controls TGF-β activation[20]; fascin-1 (*FSCN1*), which supports cancer

cell migration[21]; and the putative metastasis suppressor *RARRES3* (ref. 5). Furthermore, both gene sets include an EGFR ligand: heparin-binding EGF (*HBEGF*) in the BrMS, and epiregulin (*EREG*) in the LMS. *EREG* was highly expressed in CN34-BrM but not in MDA231-BrM cells.

These observations suggested a partial sharing of mediators of metastasis to the brain and lungs, a hypothesis that we tested by focusing on EGFR ligands and COX2. Prostaglandin production during inflammation increases BBB permeability[22]. HBEGF induces cancer cell motility and invasiveness[23]. The brain metastatic activity of BrM2 cells (Fig. 2c, d) was decreased by RNA interference (RNAi)-mediated knockdown of *COX2* expression[6] (Supplementary Fig. 6a, b), or by treatment of mice with cetuximab, which targets human EGFR[24]. To investigate cancer cell passage through the BBB, we used an *in vitro* model consisting of human primary endothelial cells and astrocytes (Fig. 2e and Supplementary Fig. 7a). This model generates a tight barrier that expresses brain endothelial markers and lacks permeability to albumin (Supplementary Fig. 7b–d)[25]. CN34-BrM2 and MDA231-BrM2 cells were three- to fourfold more active than their parental lines at migrating through this barrier. *COX2* knockdown inhibited this transmigration in both BrM2 lines, as did the addition of cetuximab to CN34-BrM2 cells, or the combination of *COX2*, *HBEGF* and *EREG* knockdowns (Fig. 2f and Supplementary Fig. 7e).

The ability of COX2 and EGFR ligands to prime breast cancer cells for extravasation into the brain may explain the association of lung and brain relapse in breast cancer[9,26]. However, given the course of breast cancer, we postulated that brain relapse also depends on selective mediators of infiltration through the unique barriers of the brain. As candidates we selected genes whose expression was increased more than threefold in all CN34-BrM2 and MDA231-BrM2 isolates, but not in bone metastatic[13] or lung metastatic MDA231 derivatives[5]. After excluding histone genes, we arrived at a set of 26 candidates (Supplementary Table 7). This set largely consists of cell–cell interaction components. The α2,6-sialyltransferase *ST6GALNAC5* stood out because its expression is normally restricted to the brain both in mice[7] and in humans (Supplementary Fig. 8). *ST6GALNAC5* messenger RNA levels were notably higher in brain metastatic derivatives than in parental cell lines in MDA231 samples (30-fold ± 1 higher; mean ± s.d.), CN34 cells (>100-fold), and in two other pleural-derived samples that were subjected to one cycle of selection for brain infiltration in mice (CN37-BrM1, 95-fold ± 23; CN41-BrM1, 72-fold ± 12).

Sambucus nigra agglutinin (SNA), a lectin that binds to α2,6-linked sialyl groups, specifically stained mammary tumours and brain lesions formed by BrM2 cells (Fig. 3a, b and Supplementary Fig. 9a, b). Fifty per cent (6 out of 12) of brain metastatic samples from breast cancer patients had areas with strong SNA staining, whereas 18% (2 out of 11) of lung metastatic samples had areas of low SNA staining, and the remaining 9 samples were negative (Fig. 3c, d). Furthermore, in a set of ER^- breast cancer metastasis to various sites, *ST6GALNAC5* mRNA levels roughly equalled that of the BrM2 cell lines in 23% (3 out of 12) of brain metastases, but not in metastases to other sites (Fig. 3e; $P = 0.04$, Fisher's exact test).

Sialyltransferases catalyse the addition of sialic acid to gangliosides and glycoproteins[27], and cell-surface sialylation has been implicated in cell–cell interactions[28]. CN34-BrM2 cells were more adhesive to monolayers of human primary brain endothelial cells than were the CN34 or *ST6GALNAC5*-knockdown CN34-BrM2 cells (Supplementary Fig. 10a). Notably, the knockdown of *ST6GALNAC5* decreased the BBB transmigration activity of CN34-BrM2 cells to ground level (Fig. 4a), and also decreased the brain metastatic activity of CN34-BrM2 cells (Fig. 4b). Brain metastasis was further decreased by combination with cetuximab treatment (Fig. 4b). *ST6GALNAC5* knockdown did not inhibit the growth of CN34-BrM2c cells in culture or as mammary tumours, the basal lung-seeding ability of these cells, or the aggressive lung-colonizing ability of the MDA231 derivative

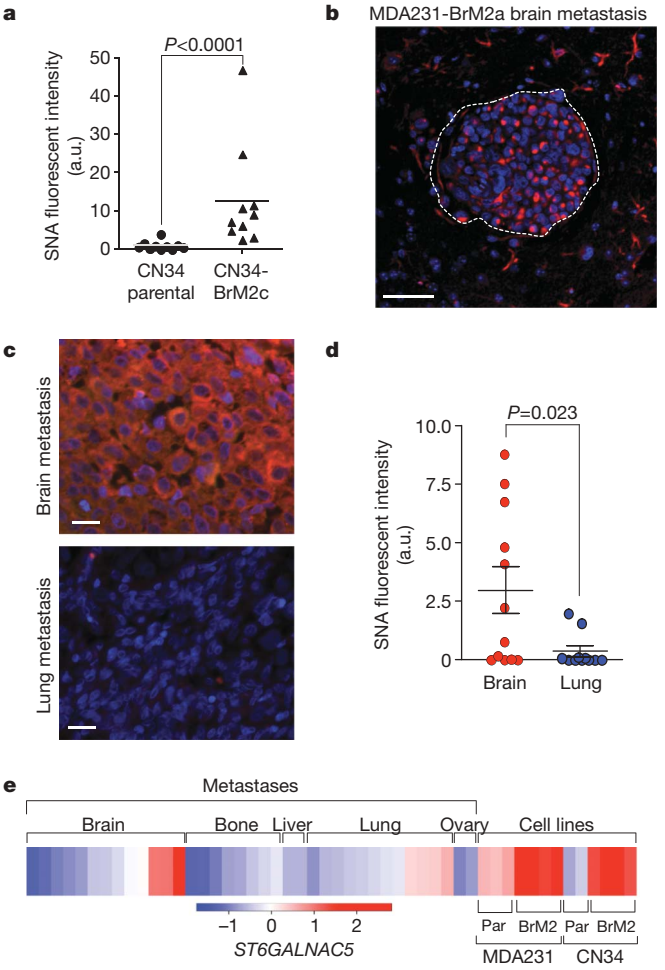

Figure 3 | *ST6GALNAC5* expression and activity in brain metastasis from breast cancer. **a**, Quantification of SNA staining in mammary fat pad tumours formed by parental CN34 or CN34-BrM2c cells in mice. a.u., arbitrary units. **b**, SNA staining of a mouse brain metastasis after intracardiac inoculation of MDA231-BrM2a cells. Scale bar, 50 μm **c**, SNA staining of representative human brain and lung metastases samples from the same breast cancer patient. Scale bars, 20 μm. **d**, Distribution of SNA staining intensity, quantified by Metamorph analysis, in 12 brain and 11 lung metastases resected from breast cancer patients. *P* values (**a**, **d**) were determined by Mann–Whitney one-tailed test. **e**, Heat map showing the relative *ST6GALNAC5* expression levels in a panel of 13 brain, 8 bone, 3 liver, 12 lung and 2 ovary human metastases from breast cancer patients. Included for comparison are the parental (par) and brain metastatic derivatives from MDA231 and CN34 cells. Data are on the basis of Affymetrix probe intensity.

line LM2-4175 (ref. 5) (Supplementary Fig. 10b–f). However, transduction of LM2-4175 cells with an *ST6GALNAC5* expression vector increased the ability of these cells to transmigrate across a BBB (Fig. 4c) and to infiltrate the brain (Supplementary Fig. 11a). Infiltrated brains showed a prevalence of micrometastases in mice inoculated with LM2-ST6 and of larger lesions in mice inoculated with BrM2 cells (Fig. 4d and Supplementary Fig. 11b). ST6GALNAC5 therefore mediates infiltration into the brain, and further mediators may be required for the expansion of the resulting foci into macrometastases.

Breast cancer cells can disseminate to the lungs from early stages of tumour development[29], indicating that cancer cells departing from breast tumours are competent for extravasation through lung microcapillary walls. Our results indicate that the expression of COX2 and HBEGF in primary tumours enhances cancer cells for extravasation through the non-fenestrated capillaries of the brain and lungs, whereas ST6GALNAC5 expression is co-selected with, and acts as a specific mediator of, cancer cell infiltration through the blood–brain

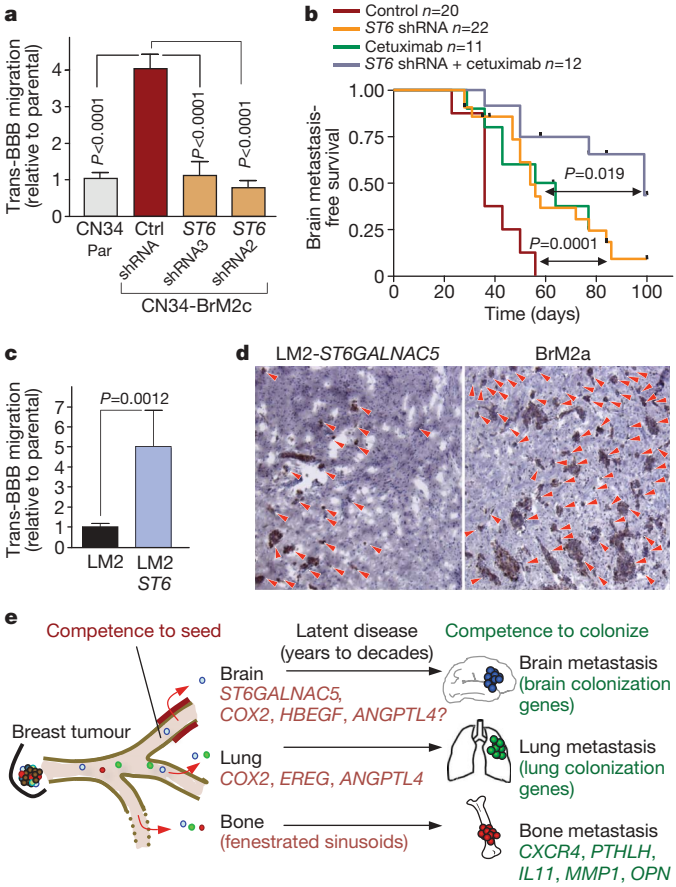

Figure 4 | The brain-specific sialyltransferase *ST6GALNAC5* as a BBB extravasation and brain metastasis gene. a, *In vitro* BBB transmigration activity of the indicated cell lines. Ctrl, control; par, parental; *ST6*, *ST6GALNAC5*. Error bars, s.e.m.; $n = 9$–27; *P* values were determined by Mann–Whitney one-tailed test. **b**, Kaplan–Meier curves for brain metastasis-free survival of mice injected with CN34-BrM2 cells expressing an shRNA targeting *ST6GALNAC5* or an empty vector control, and then treated with cetuximab or vehicle. *P* values were determined using log-rank test. **c**, *In vitro* BBB transmigration activity of LM2 cells transduced with an empty vector or with *ST6GALNAC5*. Error bars, s.e.m.; $n = 22$–27; *P* values were determined by Mann–Whitney one-tailed test. **d**, Anti-GFP immunostaining of representative lesions from mice injected intracardially with the indicated cell lines. Red arrowheads show individual tumour foci; original magnification, ×10. **e**, Schematic model of organ-specific metastatic extravasation of breast cancer cells. Extravasation into the bone marrow is a relatively permissive process owing to the fenestrated endothelium lining the sinusoid capillaries. Extravasation into the pulmonary or brain parenchyma requires specific functions for breaching the non-fenestrated capillary walls of these organs. Shared mediators of extravasation include, among others, *COX2* and EGFR ligands such as epiregulin and *HBEGF*. Passage through the BBB requires further mediators including, but not limited to, the brain-specific sialyltransferase *ST6GALNAC5*. Competence to colonize each organ requires additional mediators.

barrier (Fig. 4e). These findings draw attention to the role of cell-surface sialylation as a previously unrecognized participant in brain metastasis, and to the possibility of therapeutically disrupting these interactions. Our work also points to other candidate brain metastasis genes, including genes implicated in vascular permeability and leukocyte infiltration during brain inflammatory processes, and genes implicated in neurite extension and astrocyte cell processes. The role of these genes in brain metastasis and their interest as therapeutic targets is open to further analysis.

METHODS SUMMARY

CN34 tumour cells were isolated from the pleural effusion of a breast cancer patient treated at our institution, after written consent in accordance with Institutional Review Board (IRB) regulations. Brain metastatic populations from these cells and MDA-MB-231 cells were obtained by consecutive rounds of *in vivo* selection in 6–7-week-old beige nude and athymic mice, respectively. All animal work was done in accordance with the MSKCC Institutional Animal Care and Use Committee. Methods for RNA extraction, labelling and hybridization for DNA microarray analysis have been described previously[17]. Bioinformatics analyses with detailed descriptions can be found in the Methods. Knockdown and overexpression of candidate genes, and cetuximab inhibitor studies were performed as previously described[6]. The *in vitro* BBB model was set up as previously described[25], and modified to enable tumour cell counting. *Sambucus nigra* lectin staining was performed using standard histochemical techniques, and quantified using Metamorph software analysis. The Methods section provides further information, including malignant cell isolation from pleural fluids, tumour cell extraction and cell culture protocols, animal inoculation and bioluminescence imaging, generation of retroviral gene knockdown and overexpression vectors, transfections and infections, RNA and protein expression, *in vitro* BBB transmigration assay, endothelial cell adhesion assay, and metastatic tissue staining and quantification.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 February; accepted 26 March 2009.
Published online 6 May 2009.

1. Chiang, A. C. & Massague, J. Molecular basis of metastasis. *N. Engl. J. Med.* **359**, 2814–2823 (2008).
2. Weil, R. J. *et al.* Breast cancer metastasis to the central nervous system. *Am. J. Pathol.* **167**, 913–920 (2005).
3. Karrison, T. G., Ferguson, D. J. & Meier, P. Dormancy of mammary carcinoma after mastectomy. *J. Natl. Cancer Inst.* **91**, 80–85 (1999).
4. Schmidt-Kittler, O. *et al.* From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc. Natl Acad. Sci. USA* **100**, 7737–7742 (2003).
5. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
6. Gupta, G. P. *et al.* Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. *Nature* **446**, 765–770 (2007).
7. Okajima, T. *et al.* Molecular cloning of brain-specific GD1α synthase (ST6GalNAc V) containing CAG/glutamine repeats. *J. Biol. Chem.* **274**, 30557–30562 (1999).
8. El Kamar, F. G. & Posner, J. B. Brain metastases. *Semin. Neurol.* **24**, 347–362 (2004).
9. Lassman, A. B. & DeAngelis, L. M. Brain metastases. *Neurol. Clin.* **21**, 1–23 (2003).
10. Feld, R., Rubinstein, L. V. & Weisenberger, T. H. Sites of recurrence in resected stage I non-small-cell lung cancer: a guide for future studies. *J. Clin. Oncol.* **2**, 1352–1358 (1984).
11. Inoue, S. & Osmond, D. G. Basement membrane of mouse bone marrow sinusoids shows distinctive structure and proteoglycan composition: a high resolution ultrastructural study. *Anat. Rec.* **264**, 294–304 (2001).
12. Paku, S., Dome, B., Toth, R. & Timar, J. Organ-specificity of the extravasation process: an ultrastructural study. *Clin. Exp. Metastasis* **18**, 481–492 (2000).
13. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
14. Yoneda, T. *et al.* A bone-seeking clone exhibits different biological properties from the MDA-MB-231 parental human breast cancer cells and a brain-seeking clone *in vivo* and *in vitro*. *J. Bone Miner. Res.* **16**, 1486–1495 (2001).
15. Fitzgerald, D. P. *et al.* Reactive glia are recruited by highly proliferative brain metastases of breast cancer and promote tumor cell colonization. *Clin. Exp. Metastasis* **25**, 799–810 (2008).
16. Fan, C. *et al.* Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* **355**, 560–569 (2006).
17. Minn, A. J. *et al.* Lung metastasis genes couple breast tumor size and metastatic spread. *Proc. Natl Acad. Sci. USA* **104**, 6740–6745 (2007).
18. Egeblad, M. & Werb, Z. New functions for the matrix metalloproteinases in cancer progression. *Nature Rev. Cancer* **2**, 161–174 (2002).
19. Padua, D. *et al.* TGFβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**, 66–77 (2008).
20. Saharinen, J., Hyytiainen, M., Taipale, J. & Keski-Oja, J. Latent transforming growth factor-β binding proteins (LTBPs)–structural extracellular matrix proteins for targeting TGF-β action. *Cytokine Growth Factor Rev.* **10**, 99–117 (1999).
21. Adams, J. C. Roles of fascin in cell adhesion and motility. *Curr. Opin. Cell Biol.* **16**, 590–596 (2004).
22. de Vries, H. E. *et al.* The influence of cytokines on the integrity of the blood-brain barrier *in vitro*. *J. Neuroimmunol.* **64**, 37–43 (1996).
23. Miyamoto, S. *et al.* Heparin-binding epidermal growth factor-like growth factor as a novel targeting molecule for cancer therapy. *Cancer Sci.* **97**, 341–347 (2006).
24. Goldstein, N. I. *et al.* Biological efficacy of a chimeric antibody to the epidermal growth factor receptor in a human tumor xenograft model. *Clin. Cancer Res.* **1**, 1311–1318 (1995).

25. Eugenin, E. A. & Berman, J. W. Chemokine-dependent mechanisms of leukocyte trafficking across a model of the blood-brain barrier. *Methods* **29,** 351–361 (2003).
26. Slimane, K. *et al.* Risk factors for brain relapse in patients with metastatic breast cancer. *Ann. Oncol.* **15,** 1640–1644 (2004).
27. Harduin-Lepers, A. *et al.* The human sialyltransferase family. *Biochimie* **83,** 727–737 (2001).
28. Dall'Olio, F. & Chiricolo, M. Sialyltransferases in cancer. *Glycoconj. J.* **18,** 841–850 (2001).
29. Husemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13,** 58–68 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is dedicated to the memory of our colleague W. Gerald. We thank E. Eugenin, E. Brogi, M. Drobnjac, K. LaPerle, M. Smid, A. Viale and K. Manova-Todorova for advice and support. We thank L. DeAngelis, A. Lassman, E. Holland, J. Posner and members of the Massagué laboratory for discussions. This work was supported by grants from the National Institutes of Health (U54 CA126518), the Kleberg Foundation and the Hearst Foundation, and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO). J.M. is an investigator of the Howard Hughes Medical Institute.

Author Contributions P.D.B. and J.M. designed experiments, analysed data and wrote the manuscript. J.M. supervised the research. X.H.-F.Z. performed bioinformatics analyses. P.D.B. performed experiments. W.S. assisted with experiments. C.N. and R.R.G. isolated metastatic cells from clinical samples. D.X.N. helped with gliosis immunostaining and confocal microscopy. A.J.M. identified LMS clinical correlation with brain relapse. W.L.G., J.A.F. and M.J.V.d.V. obtained, classified and processed breast tumour samples. All authors discussed the results and commented on the manuscript.

Author Information The clinical microarray data on the brain metastatic cell lines have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under the GEO series accession number GSE12237. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M. (j-massague@ski.mskcc.org).

METHODS

Isolation of carcinoma cells from pleural effusions. Clinical specimens were obtained from three consenting patients (CN34, CN37 and CN41) with metastatic breast cancer treated at our institution, following IRB-approved protocols. Epithelial cells were obtained from pleural fluids as described before[30]. In brief, pleural fluid was collected in the presence of heparin (5 U ml^{-1}), and centrifuged at 150*g* for 10 min. Cell pellets were resuspended in PBS, red blood cells were lysed with ACK lysis buffer, and a fraction of the cells was subjected to negative selection to remove leukocytes (CD45$^+$ and CD15$^+$ populations). Cells were cultured for 24 h, and epithelial cells were sorted from this population using an anti-EpCAM antibody. The resulting cell population was transduced with a lentivirus expressing the triple-fusion reporter encoding herpes simplex virus thymidine kinase 1, GFP and firefly luciferase[31]. GFP-expressing cells were sorted and maintained at 5% CO_2 at 37 °C in M199 medium supplemented with 2.5% fetal bovine serum (FBS), 10 μg ml^{-1} insulin, 0.5 μg ml^{-1} hydrocortisone, 20 ng ml^{-1} EGF, 100 ng ml^{-1} cholera toxin, 1 μg ml^{-1} fungizone, and 100 U ml^{-1} penicillin/streptomycin, for approximately 1 week before mouse injection.

Isolation of brain metastatic cells. A cell suspension containing 10^5 CN34 breast cancer cells in a volume of 100 μl was injected in the left cardiac ventricle of anesthetized 6–7-week-old Cr:NIH-bg-nu-Xid mice. A cell suspension of 10^4 MDA-MB-231 breast cancer cells in a volume of 100 μl was injected in the left cardiac ventricle of anesthetized 6–7-week-old athymic mice. Tumour development was monitored by weekly bioluminescence imaging using the IVIS-200 imaging system from Xenogen as previously described[5]. Brain metastatic lesions were confirmed by MRI and histological analysis after necropsy. Brain lesions were localized by *ex vivo* bioluminescence imaging, and resected under sterile conditions. Half of the tissue was fixed with 4% paraformaldehyde (PFA), and processed for histological analysis. The other half was minced and placed in culture medium containing a 1:1 mixture of DMEM/Ham's F12 supplemented with 0.125% collagenase III and 0.1% hyaluronidase. Samples were incubated at room temperature for 4–5 h, with gentle rocking. After collagenase treatment, cells were briefly centrifuged, resuspended in 0.25% trypsin, and incubated for a further 15 min in a 37 °C water bath. Cells were resuspended in culture medium and allowed to grow to confluence on a 10-cm dish. GFP$^+$ cells were sorted for further propagation in culture or inoculation in mice. All animal work was done following a protocol approved by the MSKCC Institutional Animal Care and Use Committee.

Histological analysis and microscopy. Brain metastatic lesions were fixed with 4% PFA overnight, washed twice with PBS, dehydrated in 50% ethanol, and subsequent 70% ethanol, and embedded in paraffin for H&E staining. For all other purposes, animals were perfused with 10 ml PBS, and pre-fixed with 5 ml of 4% PFA. Lesions were extracted and post-fixed with 4% PFA for a further 2 h, incubated in a solution of 30% sucrose in PBS for 1–2 days, and processed for OCT compound embedding and montage. Assessment of reactive glia was performed by staining with the astrocyte marker GFAP (DAKO), followed by detection with a fluorescently labelled secondary antibody. Detection of tumour foci in the brain parenchyma was achieved by immunohistochemistry using an anti-GFP antibody (Invitrogen). For this purpose, the brain hemispheres were separated longitudinally and mounted in OCT. Four sections from each hemisphere were obtained, separated by 300 μm. In total, eight brain regions were stained with anti-GFP and microscopically explored using ×5 magnification. Microscopic analysis was performed using a Zeiss Axioplan2 microscope, and foci quantification was achieved using Metamorph software analysis. For detection of tumour cells in the brain microvasculature, 10^6 brain metastatic cells were injected into the left cardiac ventricle of anesthetized mice. Enhancement of the green fluorescence was obtained by labelling the tumour cells with 5 μM CFMDA cell-tracker dye (Invitrogen) for 45 min before injection. To visualize the brain vasculature, mice were injected with 2 mg g^{-1} of body weight rhodamine-labelled 70-kDa dextran (Invitrogen) via retro-orbital inoculation 1 h before culling. Animals were perfused and culled 24 h after tumour cell inoculation, and brain was processed for OCT compound embedding. Thirty-micrometre sections were examined on an Upright Leica TCS SP2 confocal microscope, and ×63 images were collected.

RNA isolation and gene-expression profiling. RNA was extracted from exponentially growing cells using the RNeasy mini kit (Qiagen). Labelling and hybridization of the samples to HG-U133A gene expression chip (Affymetrix) were performed by the MSKCC Genomics Core Facility using standard methodology, as previously described[17].

Data analysis was performed using the GeneSpring 7.2 software. The raw data was filtered by intensity values equal to or greater than 150. Class comparison between parental and brain metastatic populations (three biological replicates of the MDA231 parental versus four brain metastatic derivatives, and two biological replicates of CN34 parental versus four brain metastatic derivatives) was performed to identify gene expression changes of 2.5-fold associated with the brain metastatic phenotype ($P < 0.05$). This step yielded 271 genes (310 probe sets) that were differentially expressed between the parental and the brain-metastatic CN34 cell lines, and 179 genes (210 probe sets) between the parental and the brain-metastatic MDA231 cell lines. From these lists, we selected genes with a concordant expression change of at least 1.5-fold in the other cell system, genes that were upregulated in one system and consistently expressed at a high level in the other, and genes that were downregulated in one system and consistently underexpressed in the other. As a result, we obtained a final list of 280 probe sets, corresponding to 243 genes that are associated with brain metastatic phenotype in CN34 and MDA231 cells (Supplementary Table 2).

BrMS derivation and clinical sample analysis. Microarray data from four cohorts of breast tumours were used for analysis. The MSK-82 cohort was more locally advanced compared to either the NKI-295 or the EMC-286 series (91% T2-T4 and 66% node positive in MSK-82, compared to 47% T2-T4 and 49% node positive in NKI-295, and 51% T1, 46% T2 and 0% node positive in EMC-286). EMC-192 is a heterogeneous cohort that includes 144 tumours from patients that relapsed and received first-line chemotherapy for metastatic disease in an adjuvant or recurrence setting, and 48 from patients that were node-negative and did not receive adjuvant systemic therapy. Sites of recurrence and treatment information for each of these data sets are shown in Supplementary Table 3.

The MSK-82 and EMC-286 cohorts were analysed on Affymetrix HG-U133A platform, and the EMC-192 cohort on HG-U133 plus 2.0. The NKI-295 set was analysed on Agilent microarrays. To achieve statistical power given the limited incidence of brain metastasis in these cohorts, we merged the MSK-82 and EMC-286 cohorts. All data sets were first transformed to log-2 scales and median-centred. *Z* transformation was then performed to normalize gene expression across all samples in each cohort, using the MAS5.0 normalization approach[32].

The 243 gene set (280 probe set) associated with brain metastasis was used to fit a Cox hazard ratio regression model to gauge the association of each gene with brain or lung metastasis-free survival in the combined MSK-82/EMC-286 cohort. This was achieved using the survival package in the R statistical software. Wald test was used to calculate the *P* values. We designated 'brain metastasis gene set' (BrMS) the 17 genes with $P < 0.05$ from the 243 gene set. To examine whether these 17 genes were obtained owing to artefactual effects caused by merging the MSK-82 and EMC-286 cohorts, we performed principal component analysis and tested whether the two data sets differentially distributed along the top two principal components (Supplementary Fig. 12). This analysis revealed no segregation between the two data sets ($P = 0.11$ and 0.28 for the top two principal components by *t*-test) (Supplementary Fig. 9), arguing that the statistical significance and classification power of the 17 genes cannot be attributed to the merge of the two data sets. The identification of BrMS$^+$ tumours was achieved by unsupervised hierarchical clustering of tumours in the MSK-82 and EMC-286 as a training cohort. The resulting cluster-tree was cut at different distance cutoffs to yield different numbers (2 to 10) of sub-clusters. In each case, the cluster that most resembled the gene expression pattern of BrM cells was compared with the other clusters for enrichment of brain relapse events, using Fisher's exact test. The best cutoff was determined when such cluster not only maintained the resemblance of gene expression pattern to BrM cells, but also best-segregated brain relapse events. This cluster was defined as BrMS$^+$. Heat maps were generated using the gplots package of R statistical program.

For each patient, metastasis-free survival is defined as the time interval between the surgery and the diagnosis of metastasis. When there are several metastatic sites, because the chronological order of different metastatic sites on the same patient was not annotated, it has been assumed that all metastases were diagnosed at the same time, and each patient was scored for each of the corresponding sites in our analyses.

BrMS$^+$ and BrMS$^-$ tumours were used to train a support vector machine (package e1071, R statistical program). We used a linear kernel and expression values of the 17-gene BrMS as features. The trained classifier was then applied to the NKI-295 and EMC-192 cohorts to predict BrMS$^+$ tumours. We performed Kaplan–Meier analysis and log-rank tests on the survival rates of the predicted BrMS$^+$ and BrMS$^-$ tumours in the NKI-295 and EMC-192 data sets, using the survival package of R. The specificity and sensitivity of the BrMS and LMS classifiers as predictors of brain metastatic recurrence is shown in Supplementary Table 6.

As a control, we randomly generated ten sets of 500 genes each, and determined the correlation of the genes in each set with brain metastasis-free survival in the 368-tumour cohort. We then selected univariately significant genes, trained classifiers using EMC/MSK-368 and tested the classifiers on the independent EMC192 and NKI295 data sets as we did with the BrMS. Such randomly derived classifiers yielded *P* values for brain relapse association ranging from 0.1

to 0.8 (data not shown). Thus, the performance of the BrMS as a correlate of brain relapse was not due to chance.

Knockdown and overexpression cell lines. Knockdown of *COX2* and *EREG* with a validated hairpin was achieved as previously described[6]. Knockdown of *HBEGF* was achieved with pRetroSuper vector targeting the sequence 5′- GGTA TGCTGTCATGGTCCT-3′, and knockdown of *ST6GALNAC5* was achieved by targeting the sequences 5′-CATAAGCAACTCAACAATA-3′ (shRNA2), and 5′-GAGCACATCTCCACTGACT-3′ (shRNA3). Overexpression of *ST6GALNAC5* was achieved by cloning the open reading frame (ORF) of this gene into the pBabe-Puro retroviral vector from complementary DNA obtained from normal human brain using the primers 5′-GGAATTCATGTACCCATACGATGTTC CAGATTACGCTAAGACCCTGATGCGCCA-3′ and 5′- ACGCGTCGACTTAGA ACACAGGTTTATTCT-3′.

The efficiency of the knockdown and overexpression was confirmed by quantitative PCR with reverse transcription (qRT–PCR) TaqMan gene expression assays (Applied Biosystems), or western immunoblotting analysis (anti-COX2 antibody, Cayman Chemical). β2-microglobulin and actin were used as endogenous controls for qRT–PCR and western blot, respectively. The viral particles for infection of the brain-metastatic derivatives were obtained by transfection of the GPG29 amphotropic packaging cell line, and collection of supernatants 48 and 72 h after transfection. Supernatants were filtered and centrifuged at 43,000*g* to concentrate the viral particles, and were used to infect sub-confluent cultures in the presence of 5 μg ml^{-1} polybrene overnight. Puromycin (2 μg ml^{-1}) was used to select for stable cell lines. Only cell lines with a transduction rate over 80–90% were used for further studies.

Cetuximab treatment. Biweekly intraperitoneal injection of 1 mg cetuximab antibody (ImClone) was performed as previously described[6]. Animals were given one or two doses of cetuximab before intracardiac inoculation of the tumour cells, and were maintained on drug treatment until the end of the experiment.

***In vitro* blood–brain barrier assay.** Primary human umbilical vein endothelial cells (HUVEC, ScienCell) were co-cultured with human primary astrocytes (ScienCell), on opposite sides of a polylysine-treated, gelatin-coated tissue culture transwell insert for 3 days as previously described[25]. In brief, 3 μm pore PET tissue culture inserts (Fisher) were treated with polylysine (1 μg ml^{-1}, Millipore) overnight, washed four times, and coated with 0.2% gelatin (Sigma) for a minimum of 30 min. Inserts were placed upside-down in a 15 cm plate, and 10^5 primary human astrocytes were plated on the membrane surface. Astrocytes were fed every 15 min for 5 h, and the inserts were then flipped and placed in 24-well plates. Fifty-thousand endothelial cells were plated on the upper chamber of the inserts, and cultures were placed in the incubator, without further perturbation. Thee days later, the tightness of the barriers was tested by permeability to serum albumin. Evans blue-conjugated albumin (0.45% in phenol red-free medium) was added to the upper chamber and incubated for 30 min at 37 °C. Medium from the bottom chamber was collected, and absorbance was measured at 620 nm. Controls included astrocytes alone, endothelial cells alone, astrocytes plated on both sides of the insert, and insert alone. Immunofluorescence staining was performed with antibodies against the endothelial cell marker von Willebrand factor (Sigma), the tight junction marker zonula occludens 1 (Zymed), and the astrocyte marker GFAP (Dako). The expression of the brain endothelial markers glucose transporter-1 (*GLUT1*, also known as *SLC2A1*) and γ-glutamyl transpeptidase (*GGT1*) was evaluated by qPCR using TaqMan gene expression assays (Applied Biosystems) after RNA extraction from the top cell layer of the BBB inserts.

For BBB transmigration assays, cancer cells were labelled with 5 μM CFMDA cell tracker green (Invitrogen) for 45 min, and recovered overnight before assaying. Fifty-thousand cells were seeded on the upper chamber and incubated for 14–18 h. Inserts were washed with PBS and fixed with 4% PFA for 20 min. The membranes were removed from the plastic insert and mounted on a microscope slide. Pictures from 5–8 inserts per experiment were taken, and the number of transmigrated cells was counted. For cetuximab experiments, tumour cells were pre-treated with 10 μg ml^{-1} cetuximab in the medium for 5 h, and the same concentration was added to the tumour cell suspension seeded on the top of the endothelial layer.

Lectin staining. SNA staining was performed on perfused, paraffin-embedded xenograft tumour tissue as previously described[33]. In brief, after standard deparafinization, sections were washed with PBS, and endogenous peroxidase was quenched by incubation in 0.3% H_2O_2 in methanol for 30 min at room temperature. Sections were washed three times with PBS and blocked in 10% donkey serum for 30 min at room temperature. Labelling with biotin-conjugated SNA was carried out at a concentration of 100 μg ml^{-1} for 45 min, followed by three washes with PBS. An Alexa-568 conjugated-tyramide amplification kit (Invitrogen) was used following manufacturer's procedures to detect the biotinylated lectin. Sections were mounted with Prolong Gold mounting medium (Invitrogen), and images were taken using a Zeiss Axioplan2 microscope. The same protocol was followed for SNA staining of human breast cancer metastatic tissues, except that SNA was used at 10 μg ml^{-1}. Metamorph analysis was used to quantify the intensity of the lectin staining, and the resulting values were plotted in a scatter plot graph for comparison.

Metastatic sample gene-expression analysis. Human breast carcinoma metastasis specimens were obtained from the files of the Department of Pathology in compliance with protocols approved by the MSKCC IRB. Samples were snap-frozen in liquid nitrogen and stored at −80 °C. Each sample was examined histologically using H&E-stained cryostat sections. Regions were manually dissected from the frozen block to provide consistent tumour cell content of greater than 70% in tissues used for analysis. All studies were conducted under MSKCC IRB-approved protocols. RNA was extracted from frozen tissues by homogenization in Trizol reagent (GIBCO/BRL) and evaluated for integrity. cDNA was synthesized from total RNA using a T7-promoter-tagged-dT primer. RNA target was synthesized by *in vitro* transcription and labelled with biotinylated nucleotides (Enzo Biochem). Labelled target was assessed by hybridization to Test3 arrays (Affymetrix). All gene expression analysis was carried out using HG-U133A GeneChip. Gene expression was quantified using MAS 5.0 or GCOS (Affymetrix).

Oestrogen receptor status of the metastatic samples was confirmed by examination of the probe '205225_at' that represents the oestrogen receptor gene *ESR1*. Only metastases with raw intensity values <1,000 were kept for further analyses. For *ST6GALNAC5* expression, we first set the median expression value of all probes on each chip to be zero, and carried out a *Z*-transformation across different samples. The *ST6GALNAC5* expression level was then interrogated by the intensity of the probe '220979_s_at', and a heat map was made accordingly with the previously described tools.

Oncomine gene expression data analysis. Relative levels of *ST6GALNAC5* mRNA expression in human tissues were obtained by Oncomine Cancer Microarray database analysis (http://www.oncomine.org)[34] of a published gene expression data set[35]. The data was log-2-transformed, with the media set to zero and s.d. set to one.

Cell adhesion assay. Primary human brain microvascular endothelial cells (hBMVECs, ScienCell) were grown to confluency in 12-well plates. Before seeding the tumour cells, hBMVEC monolayers were washed twice with 0.5% BSA in PBS. Tumour cells were briefly trypsinized, resuspended in medium containing 0.5% BSA, and counted. Five-hundred-thousand cells were plated in each well, and allowed to adhere to the monolayer for 30 min. Plates were washed three times for 5 min each, shaking. Cells were lysed with 100 μl Passive lysis buffer (Promega) for 1 h, shaking. Firefly luciferase activity was determined using an Orion microplate luminometer (Berthold Detection Systems). Assays were performed in quadruplicate.

Other tissue culture procedures. Primary human endothelial cells and astrocytes were cultured in M199 medium supplemented with 50 mg ml^{-1} ascorbic acid, 25 mg ml^{-1} heparin, 3 mg ml^{-1} endothelial cell growth supplement (Sigma), 5 μg ml^{-1} bovine brain extract (Clonetics), 20% FBS, 5% human serum (Biocell), 1 μg ml^{-1} fungizone, and 100 U ml^{-1} penicillin/streptomycin. GPG29 cells were cultured in DMEM supplemented with 20 ng ml^{-1} doxycycline, 2 μg ml^{-1} puromycin, 0.3 mg ml^{-1} G418, and 10% FBS. 293T/17 packaging cell lines used for lentiviral production, and MDA-MB-231 parental cell lines and derivatives were cultured in DMEM supplemented with 10% FBS, 1 μg ml^{-1} fungizone, and 100 U ml^{-1} penicillin/streptomycin. All transfections were performed using Lipofectamine2000 (Invitrogen). GPG29 cells were maintained in DMEM supplemented with 10% FBS and 1 mM sodium pyruvate after transfection. *In vitro* proliferation quantification was achieved by plating 10^4 cells in each of 24-well plates, and lysing them in triplicates every other day with 100 μl with Passive lysis buffer (Promega) for 1 h, shaking. Firefly luciferase activity was determined using an Orion microplate luminometer (Berthold Detection Systems).

Other animal procedures. Orthotopic tumour growth was measured by injecting 1×10^6 viable single cells in a 1:1 mixture of PBS and growth-factor-reduced Matrigel (BD Biosciences) into mammary gland 4 in a total volume of 50 μl as previously described[5]. Primary tumour growth rates were analysed by measuring tumour length (*L*) and width (*W*), and calculating tumour volume on the basis of the formula $\pi LW^2/6$. For experimental lung metastasis assays, 2×10^5 cells were resuspended in 0.1 ml PBS and injected into the lateral tail vein. Lung metastatic progression was monitored and quantified using non-invasive bioluminescence as previously described[5].

30. Gomis, R. R. *et al.* C/EBPβ at the core of the TGF-β cytostatic response and its evasion in metastatic breast cancer cells. *Cancer Cell* **10,** 203–214 (2006).
31. Ponomarev, V. *et al.* A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur. J. Nucl. Med. Mol. Imaging* **31,** 740–751 (2004).
32. Cheadle, C., Cho-Chung, Y. S., Becker, K. G. & Vawter, M. P. Application of z-score transformation to Affymetrix data. *Appl. Bioinformatics* **2,** 209–217 (2003).

33. Kaneko, Y., Yamamoto, H., Colley, K. J. & Moskal, J. R. Expression of Galβ1,4GlcNAc α2,6-sialyltransferase and α2,6-linked sialoglycoconjugates in normal human and rat tissues. *J. Histochem. Cytochem.* **43,** 945–954 (1995).
34. Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6,** 1–6 (2004).
35. Shyamsundar, R. *et al.* A DNA microarray survey of gene expression in normal human tissues. *Genome Biol.* **6,** R22 (2005).