Analysis of tumour- and stroma-supplied proteolytic networks reveals a brain-metastasis-promoting role for cathepsin S

Lisa Sevenich¹, Robert L. Bowman¹,⁹, Steven D. Mason¹,⁹, Daniela F. Quail¹, Franck Rapaport², Benelita T. Elie¹, Edi Brogi³, Priscilla K. Brastianos⁴,⁵, William C. Hahn⁴, Leslie J. Holsinger⁶, Joan Massagué¹,⁷,⁸, Christina S. Leslie³ and Johanna A. Joyce¹,⁷,⁸,¹⁰

Metastasis remains the most common cause of death in most cancers, with limited therapies for combating disseminated disease. While the primary tumour microenvironment is an important regulator of cancer progression, it is less well understood how different tissue environments influence metastasis. We analysed tumour–stroma interactions that modulate organ tropism of brain, bone and lung metastasis in xenograft models. We identified a number of potential modulators of site-specific metastasis, including cathepsin S as a regulator of breast-to-brain metastasis. High cathepsin S expression at the primary site correlated with decreased brain metastasis-free survival in breast cancer patients. Both macrophages and tumour cells produce cathepsin S, and only the combined depletion significantly reduced brain metastasis in vivo. Cathepsin S specifically mediates blood–brain barrier transmigration through proteolytic processing of the junctional adhesion molecule, JAM-B. Pharmacological inhibition of cathepsin S significantly reduced experimental brain metastasis, supporting its consideration as a therapeutic target for this disease.

Cancer cells in a primary tumour are adept at exploiting their local tissue microenvironment. In contrast, when metastatic cells leave these favourable surroundings, they must possess or acquire traits to enable survival and colonization of potentially hostile tissue environments. The obstacles that metastasizing tumour cells encounter vary between organs, and are influenced by non-cancerous stromal cells in the local microenvironment. For example, the blood–brain barrier (BBB), composed of endothelial cells, astrocytes and pericytes, presents a more formidable structure for tumour cells to penetrate, compared with the fenestrated capillaries in bone marrow. Tumour cells that extravasate in different tissue microenvironments then encounter distinct cell types that can positively or negatively regulate metastatic outgrowth. Indeed, dissemination can occur to multiple organs, yet metastatic tumours typically grow in only one or a few sites, indicating critical roles for the microenvironment in this process.

Multiple steps of the metastatic cascade require proteolytic activity. Most studies so far have focused on how proteases regulate angiogenesis and invasion in the primary tumour. Proteases are organized into highly regulated networks in which their activity is controlled by endogenous inhibitors and interacting partners. The mechanisms by which proteases regulate tumour progression are highly specific, including cleavage of cell-adhesion molecules and processing of growth factors, chemokines and kinases, and are not limited to protein degradation and matrix remodelling. We therefore investigated the expression of proteolytic effectors in different metastatic sites to determine whether protease function is important in the rate-limiting steps of extravasation, seeding and outgrowth.

Although previous studies profiling constituent cell types of different tumour microenvironments have been informative in identifying stromal gene signatures, often with prognostic value, they involved manipulation of the tumour to isolate individual cell types, and in most cases stromal cells alone were analysed, without comparative expression data for the tumour cells. Moreover, the focus has

¹Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ²Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³Pathology Department, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ⁵Division of Hematology/Oncology, Massachusetts General Hospital, Boston, Massachusetts 02214, USA. ⁶Virobay Inc., 1360 Willow Road, Menlo Park, California 94025, USA. ⁷Brain Tumor Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁸Metastasis Research Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁹These authors contributed equally to this work.

Received 5 October 2013; accepted 20 June 2014; published online 3 August 2014; DOI: 10.1038/ncb3011

© 2014 Macmillan Publishers Limited. All rights reserved.
largely been on tumour–stroma interactions in the primary tumour, without consideration of the metastatic process. Here, we specifically examine the interplay between cancer cells and the microenvironment in intact xenograft tumours at distinct stages of metastasis using a dual species-specific microarray platform. We focused on three organ sites to which breast cancer commonly metastasizes: the brain, bone, and lung. We find that tumour-cell-derived proteases and their inhibitors predominantly undergo stage-specific changes in expression during metastatic seeding and outgrowth in different organs, whereas stromal-derived genes are primarily regulated in a tissue-specific manner. From these analyses, we show that cathepsin S expression is regulated in both a cell-type-specific and stage-dependent manner, and enhances breast-to-brain metastasis.

RESULTS

Differential expression of proteases and protease inhibitors in different metastatic microenvironments

To investigate tumour–stroma interactions in different environments we used a mouse model for organ-specific experimental metastasis (Fig. 1a). In this model, three different metastatic variants of the human breast cancer cell line MDA-MB-231 (refs 20–22) were injected either intracardially or intravenously into immunocompromised mice, resulting in the development of brain, bone, or lung metastases. Whereas previous studies focused on profiling tumour-cell-specific expression in each of these cell line variants20–22, we have been able to additionally capture the stromal contribution by removing intact whole tumours at distinct stages of metastatic seeding and outgrowth in different organs, followed by expression analyses (Fig. 1a and Supplementary Fig. 1).

We used a platform that enabled simultaneous analysis of tumour and stromal gene expression: the ‘HuMu Protin’ custom array (Hu: human; Mu: murine, Prot: protease and In: inhibitor), which surveys proteases, their endogenous inhibitors and interacting partners. The uniqueness of this microarray is based on the species-specificity of the probe sets, with no cross-reactivity between human and mouse genes. This platform thus allowed us to distinguish between messenger RNA expression changes in the tumour (human) and stromal (mouse) gene space in response to metastatic seeding and outgrowth (early- and late-stage metastases respectively, Supplementary Fig. 1), with the goal of identifying tumour–stroma interactions that modulate organ-specific metastasis. Each of the metastatic cell variants was transduced with a triple-fusion TK–GFP–Luc imaging vector, enabling non-invasive bioluminescence imaging (BLI) as a read-out of metastatic burden, as previously described. Early- and late-stage metastases in each organ site were collected on the basis of BLI output, as described in the Methods, corresponding to micrometastatic and macrometastatic disease respectively.

Principal component analysis was used to evaluate global trends in proteolytic network expression across tissue and stage for both tumour cells and stromal cells (Fig. 1b). Analysis of tumour-cell-specific gene expression revealed pronounced changes between early- and late-stage metastases across all three organs. Meanwhile, stromal genes were differentially expressed between early- and late-stage metastases in a tissue-dependent manner. The brain and bone stroma showed the most robust changes in gene expression as metastases progressed. Across all tissues, there were few changes in gene expression between the normal tissue (that is non-tumour burdened), and the stroma of early metastases (brain, bone = 0 genes, lung = 3 genes, Supplementary Table 1a). This could reflect the relatively low disease burden at the early stages, resulting in a minimal impact on the organ as a whole. Alternatively, expression changes in proteolytic genes in the stroma may not be as important in the earliest stages of metastatic extravasation and seeding, possibly owing to a predominant role for tumour-supplied proteases or non-protease factors in the stroma.

Differential gene expression analyses revealed that many genes changed in tumour cells in the brain, bone and lung (242, 241 and 245 genes respectively) between early- and late-stage metastases (Fig. 1c and Supplementary Table 1b–d). By comparison, there were fewer stage-specific differentially expressed genes identified in the brain and bone stroma (40 and 44 genes respectively), and only one differentially expressed gene, haptoglobin, in the lung stroma (Fig. 1d and Supplementary Table 1e–g). A substantial proportion of tumour-derived genes were shared across all three metastatic sites (176 genes, Supplementary Fig. 2a), whereas few stromal genes were shared across >1 metastatic site (10 genes, Supplementary Fig. 2b). Rather, there were multiple tissue-specific proteases and inhibitors in the brain, bone and lung stroma (Supplementary Fig. 2c). Quantitative real-time PCR (qPCR) confirmed tissue-specific enrichment of representative candidates for each organ in normal tissue, and early- and late-stage metastases (Supplementary Fig. 2d). Representative proteases and protease inhibitors that exhibited stage-specific gene expression changes between early and late metastases were validated using immunostaining (Fig. 2a–c) and qPCR (Supplementary Fig. 2a). We also validated several genes identified in lung metastasis xenografts (Supplementary Table 1a) in the immunocompetent MMTV-PyMT model of breast-to-lung metastasis (Supplementary Fig. 3a–g).

We next examined whether expression changes in stromal cells in organ-specific metastases are a general response to tumour cell colonization of the respective tissue, or whether they are specific to the metastatic cell variant used. In the models used here, bone metastases occasionally develop in animals inoculated with the brain-metastatic (Br-M) variant, and conversely brain metastases can be observed in mice inoculated with bone-metastatic (Bo-M) cells. This allowed us to compare stromal and tumour gene expression in the ‘matched’ (Br-M to brain, Bo-M to bone) and ‘mismatched’ (Br-M to bone, Bo-M to brain) samples. Interestingly, for the genes tested, we found that stromal expression changes depend on tumour–stroma interactions specific to the metastatic tumour cell variant (Supplementary Fig. 3a,b). In contrast, tumour gene expression in different metastatic variants responds to the same microenvironment in a similar manner, suggesting an important effect of the stroma on the tumour gene expression program (Supplementary Fig. 3c,d).

Cathepsin S is negatively associated with metastasis-free survival in patients with brain metastasis

Whereas previous analyses have identified genes associated with site-specific metastases, few studies have uncovered genes that are concordantly or discordantly expressed in tumour cells and stroma. We therefore took advantage of the species specificity of the HuMu arrays to separately profile stroma- and tumour-derived genes in cross-species analyses. We identified genes for which both human and mouse homologues were significantly altered in each metastatic
Figure 1 HuMu ProtIn array enables simultaneous acquisition of gene expression changes in tumour and stromal cells. (a) Schematic of the experimental design employed to analyse tumour-stroma interactions in different metastatic microenvironments using the dual species-specific HuMu ProtIn array in xenografted animals. Br-M: brain metastatic; Bo-M: bone metastatic; and Lu-M: lung metastatic variants of the MDA-MB-231 cell line. (b) Principal component analysis of the HuMu array data: the first and second component are plotted on the x and y axes respectively. These two components together represent the largest sources of variation in the data set. The first and second components represent 89.98% and 8.44% respectively of the variance in the tumour gene space, and 90.83% and 4.06% in the stromal gene space. This analysis revealed variation in tumour gene expression that was predominantly associated with differences between early- and late-stage metastasis. Meanwhile, variation in the stroma was associated with both stage and tissue. Dotted ellipses were drawn manually to indicate related data points within stage or organ. (c,d) Heat maps of tumour- (c) and stroma- (d) derived genes that were differentially expressed between early and late metastases across different organ sites. The lung stroma did not show extensive differences between early and late stages (Supplementary Table 1g).

site (Fig. 3a and Supplementary Fig. 4a,b). Cathepsin S showed an intriguing expression pattern: while tumour-derived cathepsin S was high in early brain metastases and decreased in late-stage metastases, stromal cathepsin S exhibited the inverse pattern, with higher expression in late-stage brain metastases versus early-stage. qPCR using species-specific probes for cathepsin S confirmed these data in an independent sample set (Fig. 3b). To distinguish between the cellular sources of cathepsin S, we will refer to tumour/human CTSS in capitals, and stromal/mouse Ctss in lower-case.

We investigated whether there were any associations between CTSS expression at the primary site and organ-specific metastasis-free survival (MFS) using a data set of locally advanced primary breast cancer with clinical annotation. Patients were separated into three equal tertiles of low, medium and high CTSS expression as described in the Methods. Kaplan–Meier analysis was used to assess MFS for brain, bone and lung. Interestingly, the high CTSS expression group was associated with decreased MFS only for the brain, and not bone or lung (Fig. 3c). This was further evident in a complementary Cox proportional hazards model analysis (Supplementary Fig. 4c).

We similarly determined whether other tumour genes that were differentially expressed in the experimental model (Fig. 1c) were associated with differences in patient survival (Supplementary Table 2). In addition to CTSS, 26 other genes were significantly associated with brain MFS. Of these, 23 genes were negatively associated with brain MFS (Supplementary Fig. 4c and Table 2a). Thirty genes were associated with bone MFS, of which 6 genes were negatively associated (Supplementary Fig. 4d and Table 2b). Fifty-nine genes were associated with lung MFS, of which 45 genes were negatively associated.
Figure 2 Independent validation of differentially expressed genes in experimental brain, bone and lung metastases. (a–e) Representative images of control (non-tumour-burdened) tissue and early- and late-stage site-specific metastases (classified by BLI intensity as in Supplementary Fig. 1; n = 3 samples for each stage and tissue) showing immunofluorescence staining of tumour- and stromal-derived proteases and protease inhibitors exhibiting stage-dependent expression changes in the HuMu ProtIn array. (a) Brain sections were stained with antibodies against the protease CTSZ (red) and the protease inhibitor TIMP2 (red) as representative candidates that were differentially expressed in tumour cells. (b) Bone sections were stained with antibodies against the protease ADAM17 (red) and the protease inhibitor SERPINB10 (red) to represent differentially expressed candidates in tumour cells in bone metastases. (c) Lung sections were stained with antibodies against the protease MMP24 (red) and the protease inhibitor SERPINE2 (red) to confirm stage-differential expression in tumour cells in lung metastases. (d) Staining for the stromal-derived protease Bace1 (red) and the protease inhibitor Timp1 (red) confirmed stage-specific expression changes in GFAP+ astrocytes (white). (e) Staining for the protease Ctse (red) and the protease inhibitor Csta (red) confirmed stage-specific stromal changes in bone metastasis. CD68+ macrophages (white) were identified as the predominant source for Csta in bone metastases. All sections were stained with GFP (green) to visualize tumour cells and DAPI as a nuclear counter stain. Scale bars, 50 μm.

(Supplementary Fig. 4e and Table 2c). Although tumour cells underwent largely congruent changes in gene expression from early- to late-stage metastases across the three organs (Supplementary Fig. 2a), only 20 of these genes were significantly associated with MFS at multiple sites, whereas most genes were associated with tissue-specific MFS (Supplementary Fig. 4f and Table 2a–c).

CST7 in brain metastasis, and CTSS and SERPINA3 in bone metastasis, were the only genes that showed similar stage-dependent and cell-type-specific expression changes to CTSS in brain metastasis (Fig. 3a and Supplementary Fig. 4a). Given that we did not observe an association of CTSS expression with patient bone MFS, and neither CST7 nor SERPINA3 expression associated with brain and bone MFS respectively (Supplementary Fig. 4a,b), we chose to further investigate the potential role of cathepsin S specifically in brain metastasis, a function not previously ascribed to this protease, or any cathepsin family member.
Figure 3 Cathepsin S shows highly regulated stage- and cell-type-specific expression changes in experimental brain metastases, and that cathepsin S expression in primary breast tumours is inversely correlated with brain MFS in patients. (a) Cross-species scatter plot shows log-fold expression changes in the tumour and stromal gene space in early- versus late-stage brain metastases. Differentially expressed genes in either the stromal (mouse) or tumour (human) gene space are shown in pink or black respectively. Genes that are differentially expressed in both the stromal and tumour gene space are shown in purple. Grey dots represent homologues with insufficient fold change as detailed in the Methods. Horizontal and vertical lines denote minimum and maximum values; horizontal line: median). P values were obtained using two-tailed unpaired t-test for b and a log-rank test for c. **P < 0.01 and ***P < 0.001.
The patient expression data above were derived from whole-tumour samples, thus precluding cell-type-specific expression analyses. We therefore stained a set of patient samples of brain metastases, with matched primary breast tumours in a subset of cases (Supplementary Table 3). Across all samples (breast cancer and brain metastases), we found the main cell types contributing to the tumour mass were cytokeratin (CK)⁺ tumour cells and CD68⁺ macrophages, with a minor fraction representing CK⁻CD68⁻ cells (Fig. 3d,e and Supplementary Fig. 5a–d). CTSS levels were highest in CD68⁺ macrophages, with expression also in CK⁺ tumour cells, albeit at lower levels than in macrophages, in both primary tumours and matched brain metastases (Fig. 3d,f and Supplementary Fig. 5a,b,e). CTSS expression in tumour cells was observed in all molecular subtypes of breast cancer analysed (Fig. 3d,f and Supplementary Fig. 5a,b,e and Table 3).

**Combined depletion of cathepsin S in tumour and stromal cells reduces experimental brain metastasis**

We investigated the stromal cell source of Ctss in the experimental brain metastasis model. Seedling and outgrowth of brain metastasis induced a stromal response characterized by accumulation of astrocytes and macrophages/microglia in metastatic lesions (Supplementary Fig. 1d). Detection of cathepsin S using an antibody that recognizes both mouse and human homologues, in combination with cell-type-specific markers, identified macrophages as the predominant stromal cell type expressing Ctss in brain metastases and normal brain (Fig. 4a). We observed a gradual increase of Ctss expression in Iba1⁺ macrophages from normal brain to early- and late-stage metastases. CTSS expression was also detectable in tumour cells, although at lower levels than in macrophages, mirroring the patient analyses. At late stages, CTSS expression was undetectable in most tumour cells. We found a similar expression pattern in an immunocompetent brain metastasis model (Supplementary Fig. 3g). These data confirm the stage- and cell-type-dependent expression changes at the protein level as predicted by the HuMu array.

Given the reciprocal, cell-type-specific expression pattern of cathepsin S, we sought to investigate whether tumour and stromal sources play important, perhaps complementary roles in the seeding and outgrowth of experimental brain metastases. To address this, we performed short hairpin (sh)-RNA-mediated CTSS knockdown (KD) in Br-M cells, achieving a 90% reduction of CTSS expression at both the mRNA and protein level, and a corresponding reduction in secreted CTSS protein (Supplementary Fig. 6a–c). CTSS KD did not alter tumour cell proliferation in culture (Supplementary Fig. 6d). After backcrossing Cts knock out (KO) mice25 into the athymic/nude background, we generated four experimental groups (shown in Fig. 4b), to analyse the effects of targeting tumour or stromal cathepsin S alone, or in combination, compared with the control group. Interestingly, only the combined removal of tumour and stromal cathepsin S significantly reduced brain metastasis incidence (Fig. 4b, Control (Ctrl); Cts wild-type (WT) versus CTSS KD; Cts KO), whereas targeting either source separately had no effect. A separate cohort of mice for all four experimental groups was aged until day 35 post-tumour cell injection, which was selected as the time point by which all mice in the control group had developed brain metastases (Fig. 4b). Brain imaging at this endpoint revealed a 64% decrease in BLI signal only in the CTSS KD; Cts KO group (Fig. 4c,d). Together, these results indicate that while there is a stage-dependency to cell-type-specific cathepsin S expression, contributions from both cellular sources are apparently required to regulate brain metastasis.

**Cathepsin S promotes transmigration of the BBB by metastatic cells**

To gain insights into the mechanisms underlying impaired metastatic seeding and/or outgrowth specifically in the CTSS KD; Cts KO group, we next analysed multiple tumorigenic processes at day 35. We found that the size and proliferation rate of brain metastases in CTSS KD; Cts KO mice were significantly lower than any of the other groups (Fig. 4c–e), whereas there were no differences in apoptosis (Supplementary Fig. 6e). The small lesions that did develop in the CTSS KD; Cts KO mice were closely apposed to the vasculature, with most tumour cells being only 1 cell diameter from the vessel, and a pronounced reduction in growth away from blood vessels (Fig. 5a,b). Similarly, analysis of the GFP⁺ tumour cell area relative to the CD34⁺ blood vessel area confirmed this reduction (Supplementary Fig. 6f). Critically, Cts deletion did not alter blood vessel density or permeability in the normal brain of non-tumour-bearing animals (Supplementary Fig. 6g,h). These results are suggestive of either a potential defect in seeding of brain metastatic cells in the earliest stages, or a subsequent impairment in colonization, or perhaps deficiencies in both processes.

To investigate these possibilities, we assessed metastatic seeding across the four experimental groups. We examined the earliest steps of brain metastatic cell homing and survival26, specifically the first 48 h. We found that 24 h after CTSS KD tumour cell injection, BLI signal was reduced in both WT and Cts KO mice, with a further decrease in BLI signal at 48 h only in the CTSS KD; Cts KO group (Fig. 6a,b). Similarly, analysis of the proportion of viable tumour cells still within the blood vessel lumen (intravascular), in the process of extravasating, or fully extravascular, revealed significant differences in the CTSS KD; WT group at 24 h, and in the CTSS KD; Cts KO group at both 24 h and 48 h (Fig. 6c).

Given the initial reduction in tumour cell extravasation in CTSS KD; Cts WT mice (Fig. 6a,c), although the incidence of detectable brain metastasis was ultimately not affected (Fig. 4b), we assessed subsequent metastatic colony outgrowth. While there was an initial trend towards delayed growth in the CTSS KD; Cts WT cohort, brain metastases ultimately grew to the same extent as controls (Supplementary Fig. 7a). In contrast, tumour growth kinetics in the CTSS KD; Cts KO group did not recover over the same period (Supplementary Fig. 7a). This suggests that tumour- and stromal-derived cathepsin S show some functional redundancy during seeding and outgrowth, and the impact of each cellular source is most likely regulated by differential expression levels at distinct stages. In addition, tumour-cell-derived CTSS may be important for the initial steps of BBB transmigration and extravasation into the brain, whereas stromal-supplied Cts is subsequently involved in supporting tumour cell survival to successfully form brain micrometastases, and only their combined depletion impairs both metastatic seeding and outgrowth. A similar finding was reported in a colorectal carcinoma model, where depletion of both tumour and stromal cathepsin S was required to slow tumour growth27.
Figure 4 Macrophages are the predominant source of stromal-derived cathepsin S and only combined depletion of tumour- and stromal-derived cathepsin S reduces experimental brain metastasis. (a) Representative images of normal brain, early- and late-stage brain metastasis (classified by BLI intensity) co-stained for Ctss/CTSS (red) and GFP (tumour cells; green) or Iba1 (macrophages/microglia; white). Tumour-cell-derived CTSS is indicated by the arrowhead and macrophage-derived Ctss is indicated by the arrow. Images are representative of 5 independent specimens for each stage. (b) Kaplan-Meier curve shows the percentage of brain metastasis-free animals in the four experimental groups indicated in the table. Ctrl; Ctss WT (n = 21 mice), CTSS KD; Ctss WT (n = 7 mice), Ctrl; Ctss KO (n = 13 mice), and CTSS KD; Ctss KO (n = 11 mice). (d) Representative ex vivo BLI images of the 4 experimental groups as shown in c. (e) Quantification of tumour cell proliferation (percentage of Ki67+GFP+ cells) on day 35 after tumour cell inoculation. Ctrl; Ctss WT (n = 8 mice), CTSS KD; Ctss WT (n = 8 mice), Ctrl; Ctss KO (n = 6 mice), and CTSS KD; Ctss KO (n = 10 mice). Scale bars, 50 μm. Circles represent individual mice and horizontal lines represent the mean ± s.e.m. P values were obtained with Mantel-Cox log-rank test for MFS and with two-tailed unpaired t-test for numerical data. *P < 0.05, **P < 0.01 and ***P < 0.001.

Cathepsin S promotes BBB transmigration through junctional protein cleavage

The BBB is a selective barrier between the systemic circulation and the brain, which is formed by specialized endothelial cells, pericytes and astrocytes. Although the BBB restricts the entry of most macromolecules, it is not an impenetrable barrier to transmigration of metastasizing cancer cells. We therefore examined the potential role of tumour-cell-supplied CTSS in BBB penetration, by using an in vitro BBB assay. We performed genetic or pharmacological depletion of CTSS in Br-M cells through shRNA-mediated KD or a cathepsin S-specific inhibitor VBY-999 respectively, which did not affect Br-M cell viability (Supplementary Fig. 8a). CTSS depletion did not affect the ability of Br-M cells to cross a BBB formed by human umbilical endothelial vein cells (HUVECs) and astrocytes (Supplementary Fig. 8b). In contrast, when human brain microvascular endothelial cells (HBMECs) were used instead of...
HUVECs, this presented a stronger barrier to Br-M transmigration, which was impaired by 55–65% through genetic or pharmacological depletion of CTSS respectively (Fig. 6d). HBMEC monolayers (without astrocytes) also significantly decreased CTSS KD Br-M cell transmigration (Supplementary Fig. 8b), whereas transmigration across HUVECs or astrocytes alone was not altered by CTSS depletion (Supplementary Fig. 8b).

Tight and adherens junctions between adjacent cells are critical for maintaining BBB integrity, and are composed of different proteins including junctional adhesion molecules (JAMs), occludin, claudins and cadherins. Therefore, we investigated whether any of these proteins represented potential CTSS substrates. We performed biochemical cleavage assays using recombinant CTSS and recombinant proteins for each of the potential substrates, performed biochemical cleavage assays using recombinant CTSS and recombinant proteins for each of the potential substrates.

As the effects of CTSS depletion or inhibition on Br-M transmigration were observed only when HBMECs were used in the BBB assay, and given the organ-specificity of Jam-B expression (Fig. 7c), we reasoned that Jam-B was the most relevant substrate in this context. We aimed to identify the putative cleavage location for CTSS in Jam-B. We compared fragment sizes of each cleavage product that was detectable with Jam-A, -B or -C specific antibodies to fragments that contain the immunoglobulin (Ig)G1 domain, which is linked to recombinant Jam proteins (Supplementary Fig. 8d). The molecular weight of Jam cleavage products and pH dependence of Jam processing by CTSS suggests that all 3 family members share a similar but not fully conserved CTSS cleavage site that is linked to recombinant Jam proteins (Supplementary Fig. 8d). The molecular weight of Jam cleavage products and pH dependence of Jam processing by CTSS suggests that all 3 family members share a similar but not fully conserved CTSS cleavage site that is linked to recombinant Jam proteins (Supplementary Fig. 8d).
in biochemical studies. Cleavage in this region probably leads to shedding of the JAM extracellular domain, thereby disrupting cell–cell adhesion. We performed cell-based assays to determine whether tumour-cell-secreted CTSS mediates removal of JAM-B from the HBMEC surface (Fig. 7d). Indeed, incubation of HBMECs with Br-M cell-conditioned media led to a CTSS-mediated accumulation of JAM-B in HBMEC-conditioned media after 2–4 h. The effect was decreased by addition of VBY-999 (Fig. 7e and Supplementary Fig. 8f). These results are consistent with impaired BBB transmigration in vitro and in vivo when CTSS is targeted, as CTSS-mediated cleavage of the JAM-B extracellular domain would be expected to disrupt tight junction integrity, thereby allowing tumour cells to breach the BBB.

Cathepsin S inhibition reduces experimental brain metastasis formation

We next examined whether pharmacological inhibition of cathepsin S reduced metastatic seeding and colonization in a preclinical prevention trial (Fig. 8a). Mice were treated with VBY-999 for 2 days to inhibit cathepsin S activity before Br-M cell inoculation, and were then continuously treated with VBY-999 until the trial endpoint of 35 days post-tumour cell inoculation. Pharmacokinetic analysis showed that VBY-999 levels in the plasma were significantly above the required concentration for target inhibition at the time of Br-M cell inoculation and confirmed that VBY-999 efficiently crosses the BBB with stable concentrations in the brain.

**Figure 6** Cathepsin S mediates BBB transmigration of brain metastatic cells. (a) Quantification of BLI intensity at the indicated time points relative to BLI signal immediately after tumour cell inoculation. Ctrl; Ctss WT (n = 10 mice), CTSS KD; Ctss WT (n = 9 mice), Ctrl; Ctss KO (n = 8 mice), and CTSS KD; Ctss KO (n = 8 mice) for the 24 h time point, and n = 5 for each group for the 48 h time point. (b) Representative BLI images in the 4 experimental groups immediately (0 h) and 48 h after tumour cell injection in vivo (top panels) and ex vivo (lower panel). Images are representative of the mice analysed in a. (c) Tumour cells were categorized on the basis of their localization relative to the vasculature defined as intravascular, extravasating and extravascular, and the percentage of viable tumour cells in each category was quantified at the indicated time points. Ctrl; Ctss WT (n = 4 mice), CTSS KD; Ctss WT (n = 3 mice), Ctrl; Ctss KO (n = 3 mice), and CTSS KD; Ctss KO (n = 4 mice) for the 24 h time point, and n = 4 mice for each group for the 48 h time point. (d) Quantification of the number of transmigrated Br-M Ctrl and CTSS KD cells in the presence or absence of the cathepsin-S-specific inhibitor VBY-999 through an in vitro BBB formed with HBMECs in co-culture with astrocytes. Two hundred fields of view were analysed per sample. Br-M Ctrl Vehicle (n = 25 samples), Br-M Ctrl VBY-999 (n = 22 samples), Br-M KD Vehicle (n = 15 samples), and Br-M CTSS KD VBY-999 (n = 15 samples) were analysed in total from three independent experiments. Circles represent individual mice, horizontal lines represent the mean ± s.e.m. for numerical data shown in a. Graphs represent mean ± s.e.m. in d. Categorical data are plotted as stacked bars. P values were obtained with two-tailed unpaired t-test for numerical data and with an ordinal Chi-square test for categorical data. NS, not significant; *P < 0.05; **P < 0.01; and ***P < 0.001.
Figure 7 Cathepsin S cleaves tight junction proteins that regulate BBB integrity. (a) Western blot analysis of CTSS-mediated cleavage of recombinant tight junction proteins (JAM-A, -B and -C; occludin (OCLN); and claudins (CLDN)-3 and -5), and adherens junction proteins (cadherin 5 (CDH5) and CD31) for the indicated time points at pH 4.5 and pH 6.0 in the presence or absence of the cathepsin-S-specific inhibitor VBY-999. VBY-999 was used at 10 μM, a concentration that efficiently inhibits cathepsin S. (b) mRNA expression of the tight junction and adherens junction molecules in HUVECs and HBMECs (n=9 samples for each cell line). All assays were run in triplicate and gene expression was normalized to B2M. Expression is depicted relative to expression in HBMECs. (c) Representative images of control brain, bone and lung sections stained for the tight junction proteins Jam-B, Ocln or Cldn 3 (white), with Cd31 (red) to visualize blood vessels. DAPI was used as a nuclear counterstain. Images are representative of 3 independent specimens. (d) Schematic of the cell-based cleavage assay. (e) Western blot analysis showing increased JAM-B in HBMEC-conditioned media (CM) after incubation with Br-M-cell-conditioned media for the indicated time points. Addition of the cathepsin-S-specific inhibitor VBY-999 (10 μM) resulted in reduced accumulation of JAM-B in HBMEC-conditioned media at the indicated time points. Incubation with PBS pH 6.0, 0.05 mM dithiothreitol served as a control for baseline JAM-B shedding from HBMEC. Scale bars, 20 μm. Graphs represent mean ± s.e.m. P values were obtained using two-tailed unpaired t-test. NS, not significant; ***P < 0.001. Each western blot shows the representative result of three independent experiments. Uncropped images of blots are shown in Supplementary Fig. 9.
Figure 8 Pharmacological inhibition of cathepsin S reduces brain metastasis formation in a preclinical trial. (a) Schematic of the prevention trial experimental design. (b) Quantification of VBY-999 concentrations in plasma and brain tissue at the indicated time points after treatment started ($n=3$ mice for each group). (c) Quantification of BLI intensity in the head region at the indicated time points after Br-M cell inoculation. $n=20$ mice for vehicle group (5% dextrose in water) and $n=21$ mice for VBY-999 treatment group (100 mg kg$^{-1}$ d$^{-1}$). The BLI signal in the VBY-999 versus control group is reduced by 77, 70 and 65%, respectively at each of the three successive time points indicated. (d) Representative BLI images at the trial endpoint, day 35 after Br-M cell inoculation. Images are representative of the number of mice as in c. (e) Quantification of BLI intensity at day 35 after Bo-M tumour cell inoculation in the bone and spine region. Vehicle ($n=12$ mice) and VBY-999 ($n=13$ mice). (f) Representative BLI and X-ray images at day 35 after Bo-M cell inoculation. Arrows indicate osteolytic lesions. Images are representative of the number of mice as in e. Bars represent mean ± s.e.m. for b, circles represent individual mice, horizontal lines represent the mean ± s.e.m. for c.e. $P$ values were obtained using two-tailed unpaired t-test. NS, not significant; $^*P<0.05$.

Throughout the trial (Fig. 8b). Interestingly, BLI signal was reduced by 65–77% at different time points during the trial and at the 35 day endpoint (Fig. 8c,d).

Initiation of VBY-999 treatment in fully established, end-stage brain metastases did not alter tumour burden (Supplementary Fig. 7b,c), indicating that targeting this enzyme is most critical in seeding and early outgrowth. We investigated the organ specificity of cathepsin S inhibition by assessing bone metastasis in prevention trials, using two different approaches. First, as bone and spine metastases can develop in the Br-M model, we assessed whether these lesions were affected by VBY-999. There was no significant difference between the treatment groups, which was supported by the finding that genetic depletion of cathepsin S did not affect bone metastasis formation (Supplementary Fig. 7d,e). Similarly, a VBY-999 prevention trial specifically in the bone metastasis model showed no change in BLI output or development of osteolytic metastases (Fig. 8e,f and Supplementary Fig. 7f). These data are consistent with our finding that CTSS expression levels in patients correlated only with brain MFS, and not bone MFS. In sum, cathepsin S inhibition is efficient in substantially and specifically reducing brain metastasis if cathepsin S activity is blocked throughout the disease course.

**DISCUSSION**

Here we identify a brain-metastasis-promoting role for cathepsin S. Using gene expression screens of distinct metastatic microenvironments in animal models, and analyses of patient samples, we uncovered a specific association between high cathepsin S levels...
and the development of brain metastases. Interestingly, cathepsin S was regulated in both a cell-type-specific and stage-dependent manner, with high expression in tumour cells in early-stage metastases, which is subsequently downregulated concomitant with increased macrophage-specific expression in late-stage lesions. This pattern of expression shares intriguing similarities with lung metastatic niche development, whereby tenascin C (TNC) is initially provided by the tumour cells to support extravasation into the lung. TNC production later switches to infiltrating stromal cells as lung metastases expand\textsuperscript{38}. Cathepsin S also shows cell-type-specific expression in normal tissues, and is predominantly produced by immune and stromal cells\textsuperscript{6,39,40}. Induction of cathepsin S expression in tumour cells of epithelial origin may indicate a type of ‘leukocytic mimicry’ whereby metastatic tumour cells could implement immune cell-like expression programs that enhance mobilization and cell motility. Indeed, similarities between the behaviour of metastatic tumour cells and leukocytes have been noted\textsuperscript{41}, which could be particularly advantageous during BBB transmigration\textsuperscript{2}.

Considering that extravasation into the brain parenchyma is a rate-limiting step in brain metastasis\textsuperscript{10,42,43}, we further investigated the contribution of cathepsin S to this specific stage. We identified JAM-B cleavage by cathepsin S as an important mechanism by which BBB transmigration of Br-M cells is enhanced. Consistently, genetic or pharmacological targeting of cathepsin S significantly impaired experimental brain metastasis. We found that both tumour- and stroma-derived cathepsin S are important for metastatic seeding and outgrowth, and infer that this relates to a requirement for high levels of the enzyme overall, rather than distinct functions in these two cell types. Thus, although the levels of tumour CTSS are highest at the step of metastatic cell extravasation, there is nonetheless stromal Ctsx already expressed at this stage, albeit at much lower levels than in later stages. Similarly, while stromal cells become the predominant source of cathepsin S during colonization, there is still some tumour CTS present. Therefore, both sources need to be depleted, otherwise there is some degree of functional compensation from the cell type that is not targeted. Given that our analysis of breast cancer patients revealed a significant association between high cathepsin S levels in the primary site and reduced brain MFS, it may be instructive to specifically consider this patient population for any future therapeutic intervention to reduce brain metastasis by cathepsin S inhibition.

The proteolytic gene expression screens performed herein resulted in the identification of numerous differentially expressed genes that are regulated in a stage- or tissue-specific manner in different metastatic microenvironments. By querying whether expression of these genes in primary breast cancer patients was associated with MFS in brain, bone or lung, we were able to apply an additional filter that allowed subsequent restriction of the gene lists to only those that showed a significant correlation with survival, which represent interesting targets for future investigation. Moreover, determining expression levels of these genes at the primary site may be useful as prognostic markers to evaluate the risk for patients to develop tissue-specific metastasis. As such, the identification of proteases in different metastatic microenvironments, many of which can be pharmacologically inhibited\textsuperscript{40–49}, reveals an abundance of potential targets for metastatic disease.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Note:** Supplemental Information is available in the online version of the paper.

**ACKNOWLEDGEMENTS**

We thank K. Simpson and X. Chen for excellent technical support, and members of the J.A.J. laboratory for insightful discussion. We thank L. Akkari, O. Olson and D. Yan for reading the manuscript. We are grateful to P. Bos for advice on the BBB assays and experimental brain metastasis model. We thank the MSKCC Core Facilities of Genomics, Molecular Cytology, Small Animal Imaging and Monoclonal Antibody Production for technical assistance. We thank H. Chapman (UCSF) for providing cathepsin S KO mice. This research was supported by the following: US National Cancer Institute program grants of the Integrative Cancer Biology Program (CA148967; J.A.J., C.S.L.) and Tumour Microenvironment Network (CA126518; J.A.J., J.M.), the Health Research Science Board of New York, and the Alasan and Sandra Gerry Metastasis Research Initiative (J.A.J.), Deutsche Forschungsgemeinschaft (SE2324/1-1; L.S.). US National Cancer Institute F31 fellowship CA167863 and Gerstner Sloan Kettering graduate program (R.L.B.), US National Cancer Institute F32 fellowship CA130329 (S.D.M.), and Canadian Institutes of Health Research (D.F.Q.).

**AUTHOR CONTRIBUTIONS**

L.S., S.D.M. and J.A.J. designed experiments. L.S., S.D.M., D.F.Q. and B.T.E. performed experiments and analysed data. R.L.B., E.R. and C.S.L. performed computational analyses, E.B.K., W.C.H., L.I.H. and J.M. provided patient samples or reagents. J.A.J., L.S. and R.L.B. wrote the manuscript. All authors edited or commented on the manuscript. J.A.J. conceived and supervised the study.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Published online at www.nature.com/dofinder/10.1038/ncb3011 Reprints and permissions information is available online at www.nature.com/reprints

1. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. Nat. Med. 19, 1423–1437 (2013).
2. Nguyen, D. X., Bos, P. D. & Massague, J. Metastasis. From dissemination to organ-specific colonization. Nat. Rev. Cancer 9, 274–284 (2009).
3. Goss, P. E. & Chambers, A. F. Does tumor dormancy offer a therapeutic target? Nat. Rev. Cancer 10, 871–877 (2010).
4. Affara, N. I., Andreu, P. & Coussens, L. M. Delineating protease functions during cancer development. Methods Mol. Biol. 539, 1–32 (2009).
5. Mason, S. D. & Joyce, J. A. Proteolytic networks in cancer. Trends Cell Biol. 21, 228–237 (2011).
6. Gocheva, V. et al. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. Genes Dev. 20, 543–556 (2006).
7. Murphy, G. The ADAMs: signaling scissors in the tumour microenvironment. Nat. Rev. Cancer 8, 929–941 (2008).
8. Butler, G. S. & Overall, C. M. Proteomic identification of multitasking proteins in unexpected locations complicates drug targeting. Nat. Rev. Drug Discov. 8, 935–948 (2009).
9. Lopez-Otin, C. & Hunter, T. The regulatory crosstalk between kinases and proteases in cancer. Nat. Rev. Cancer 10, 278–292 (2010).
10. Kessenbrock, K., Plaks, V. & Werb, Z. Matrix metalloproteinases: regulators of the tumor microenvironment. Cell 141, 52–67 (2010).
11. Allinen, M. et al. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell 6, 17–32 (2004).
12. Finak, G. et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat. Med. 14, 518–527 (2008).
13. Enco, N., Truitt, M., Olson, P., Arron, S. T. & Hanahan, D. Cancer-associated fibroblasts are activated in incipient neoplasia to orchestrate tumor-promoting inflammation in an NF-κB-dependent manner. Cancer Cell 17, 135–147 (2010).
14. Katz, A. M. et al. Astrocyte-specific expression patterns associated with the PDGF-induced glioma microenvironment. PLoS ONE 7, e32453 (2012).
15. Seaman, S. et al. Genes that distinguish physiological and pathological angiogenesis. Cancer Cell 11, 539–554 (2007).
16. Ojala, L. S., Whittaker, C. A., Condeelis, J. S. & Pollard, J. W. Gene expression analysis of macrophages that facilitate tumor invasion supports a role for Wnt-signaling in mediating their activity in primary mammary tumors. J. Immunol. 184, 702–712 (2010).
17. Paget, S. The distribution of secondary growths in cancer of the breast. 1889. Cancer Metastasis Rev. 8, 98–101 (1989).
18. Disibio, G. & French, S. W. Metastatic patterns of cancers: results from a large autopsy study. Arch. Pathol. Lab. Med. 132, 931–939 (2008).
19. Lee, Y. T. Breast carcinoma: pattern of metastasis at autopsy. J. Surg. Oncol. 23, 175–180 (1983).
20. Bos, P. D. et al. Genes that mediate breast cancer metastasis to the brain. Nature 459, 1005–1009 (2009).
21. Kang, Y. et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell 3, 537–549 (2003).
22. Minn, A. J. et al. Genes that mediate breast cancer metastasis to lung. Nature 436, 518–524 (2005).
23. Schwartz, D. R. et al. Hu/Mu Protin oligonucleotide microarray: dual-species array for profiling protease and protease inhibitor gene expression in tumors and their microenvironment. Mol. Cancer Res. 5, 443–454 (2007).
24. 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).
25. Shi, G. P. Cathepsin S required for normal MHC class II peptide loading and germinal center development. Immunity 10, 197–206 (1999).
26. Lorger, M. & Felding-Habermann, B. Capturing changes in the brain microenvironment during initial steps of breast cancer brain metastasis. Am. J. Pathol. 176, 2958–2971 (2010).
27. Small, D. M. et al. Cathepsin S from both tumor and tumor-associated cells promote cancer growth and neovascularization. Int. J. Cancer 133, 2102–2112 (2013).
28. Abbott, N. J., Ronnback, L. & Hansson, E. Astrocyte-endothelial interactions at the blood–brain barrier: an overview of structure, regulation, and regulator substances. CNS Neurosci. Ther. 13, 785–799 (2007).
29. Eichler, A. F. et al. Cross-species hybridization of microarrays for studying tumor transcriptome of brain metastasis. Proc. Natl. Acad. Sci. USA 108, 17456–17461 (2011).
30. Kirschke, H., Wiederanders, B., Bromme, D. & Rinne, A. Cathepsin S from bovine spleen. Purification, distribution, intracellular localization and action on proteins. Biochem. J. 264, 467–473 (1989).
31. Liu, W. Y., Wang, Z. B., Zhang, L. C., Wei, X. & Li, L. Tight junction in blood–brain barrier: an overview of structure, regulation, and regulator substances. Trends Pharmacol. Sci. 29, 22–28 (2008).
32. Gallagher, F. A. et al. Magnetic resonance imaging of pH in vivo using hyperpolarized 13C-labelled bicarbonate. Nature 453, 940–943 (2008).
33. Choe, Y. et al. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J. Biol. Chem. 281, 12824–12832 (2006).
34. Nolan, D. J. et al. Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. Dev. Cell 26, 204–219 (2013).
35. Oliveira, M. et al. Improvement of cathepsin S detection using a designed FRET peptide based on putative natural substrates. Peptides 31, 562–567 (2010).
36. Chapman, H. A. Endosomal proteases in antigen presentation. Curr. Opin. Immunol. 18, 78–84 (2006).
37. Gupta, S., Singh, R. K., Dastidar, S. & Ray, A. Cysteine cathepsin S as an immunomodulatory target: present and future trends. Expert Opin. Ther. Targets 12, 291–299 (2008).
38. Small, D. M. et al. Cathepsin S from both tumor and tumor-associated cells promote cancer growth and neovascularization. Int. J. Cancer 133, 2102–2112 (2013).
39. Abbott, N. J., Ronnback, L. & Hansson, E. Astrocyte-endothelial interactions at the blood–brain barrier: an overview of structure, regulation, and regulator substances. Trends Pharmacol. Sci. 29, 22–28 (2008).
40. Binosi, M. L., Nagler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. Proteomics 10, 5363–5373 (2011).
41. Binns, D. et al. Cathepsin S from bovine spleen. Purification, distribution, intracellular localization and action on proteins. Biochem. J. 264, 467–473 (1989).
42. Lee, Y. T. Breast carcinoma: pattern of metastasis at autopsy. J. Surg. Oncol. 23, 175–180 (1983).
43. Eichler, A. F. et al. Cross-species hybridization of microarrays for studying tumor transcriptome of brain metastasis. Proc. Natl. Acad. Sci. USA 108, 17456–17461 (2011).
44. Turk, B. Targeting proteases: successes, failures and future prospects. Nat. Rev. Cancer 11, 352–363 (2011).
45. Palermo, C. & Joyce, J. A. Cysteine cathepsin proteases as pharmacological targets in cancer. Trends Pharmacol. Sci. 29, 22–28 (2008).
46. Gupta, S., Singh, R. K., Dastidar, S. & Ray, A. Cysteine cathepsin S as an immunomodulatory target: present and future trends. Expert Opin. Ther. Targets 12, 291–299 (2008).
47. Madsen, C. D. & Sahai, E. Cancer dissemination—lessons from leukocytes. Dev. Cell 19, 13–26 (2010).
48. Small, D. M. et al. Cathepsin S from both tumor and tumor-associated cells promote cancer growth and neovascularization. Int. J. Cancer 133, 2102–2112 (2013).
49. Abbott, N. J., Ronnback, L. & Hansson, E. Astrocyte-endothelial interactions at the blood–brain barrier: an overview of structure, regulation, and regulator substances. Trends Pharmacol. Sci. 29, 22–28 (2008).
50. Binosi, M. L., Nagler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. Proteomics 10, 5363–5373 (2011).
51. Binns, D. et al. Cathepsin S from bovine spleen. Purification, distribution, intracellular localization and action on proteins. Biochem. J. 264, 467–473 (1989).
52. Lee, Y. T. Breast carcinoma: pattern of metastasis at autopsy. J. Surg. Oncol. 23, 175–180 (1983).
53. Eichler, A. F. et al. Cross-species hybridization of microarrays for studying tumor transcriptome of brain metastasis. Proc. Natl. Acad. Sci. USA 108, 17456–17461 (2011).
54. Turk, B. Targeting proteases: successes, failures and future prospects. Nat. Rev. Cancer 11, 352–363 (2011).
55. Palermo, C. & Joyce, J. A. Cysteine cathepsin proteases as pharmacological targets in cancer. Trends Pharmacol. Sci. 29, 22–28 (2008).
56. Gupta, S., Singh, R. K., Dastidar, S. & Ray, A. Cysteine cathepsin S as an immunomodulatory target: present and future trends. Expert Opin. Ther. Targets 12, 291–299 (2008).
METHODS

Mice. All animal studies were approved by the Institutional Animal Care and Use Committee of Memorial Sloan-Kettering Cancer Center (MSKCC). Athymic/nude mice were purchased from NCI Frederick or bred within the MSKCC animal facility. The cathespin S KO mouse line (Ctss KO) was generated as described previously22 and backcrossed for 6 generations to the athymic/nude background. NOD/SCID mice were purchased from Charles River Laboratories. MMTV-PyMT (ref. 50) immunocompetent transgenic mice (FVB/n) were bred within the MSKCC animal facility.

Cell lines. Brain- (Br-M), bone- (Bo-M) and lung- (Lu-M) metastatic variants were isolated from the human breast cancer cell line MDA-MB-231 (denoted 'parental') were provided by J.M. and labelled with a triple-imaging vector (TK–GFP–Luc; TGL; ref. 24) to allow for non-invasive in vivo imaging of tumour growth over time. The MDA-MB-231 variants were cultured in DMEM + 10% FBS. Mouse Br-M variants were derived from the TSI cell line23, which was previously isolated from a MMTV-PyMT mammary tumour, through sequential rounds of in vitro and in vivo selection for homing to the brain. These are denoted PyMT-BrM cells. Human umbilical vein endothelial cells (HUVECs) were purchased from the ATCC. Human brain microvascular endothelial cells (HBMECs) and human astrocytes (HA) were purchased from ScienCell. HUVECs and HBMECs were cultured on gelatin-coated cell culture dishes, and HA on poly-L-lysine-coated cell culture dishes in endothelial cell media (ECM, ScienCell) + 10% FBS supplemented with endothelial cell growth factors (ECGF).

Generation of experimental brain, bone and lung metastases. For brain and bone metastasis generation in xenografted mice, 1 × 10^6 brain-metastatic cells (Br-M) or 1 × 10^6 bone-metastatic cells (Bo-M) were inoculated into the lateral ventricle of 6–8-week-old female athymic/nude mice. For lung metastasis generation, 1 × 10^5 lung metastatic cells (Lu-M) were injected into the lateral tail vein of 6–8-week-old female NOD/SCID mice. For brain metastasis generation in immunocompetent mice, 1 × 10^5 PyMT-BrM cells were inoculated into the left cerebral ventricle of 6–8-week-old female FVB/n mice. Early- and late-stage metastases were defined by bioluminescence intensity (BLI) at the time of tissue collection for samples used in the microarray analysis and the independent sample set used for validation. Brain metastases that had a BLI output between 4.3 × 10^2 and 4.2 × 10^3 photons s^-1 were classified as early-stage metastases and were collected between 2–4 weeks after tumour cell inoculation. Late-stage brain metastases had a BLI output between 1.6 × 10^5 and 6.4 × 10^5 photons s^-1 and were collected between 5–8 weeks after tumour cell inoculation. Histological and morphometric analyses of these different stages showed that early-stage brain metastases are comprised of clusters of ~50–200 cells, and can be considered similar to ‘micrometastases’. Late-stage and late-stage metastases consist of clusters of ~5,000–15,000 cells, corresponding to ‘macrometastases’. Representative images of the different stages are shown in Supplementary Fig. 1d. Early-stage bone metastases were defined by a BLI intensity that ranged between 6.3 × 10^1 and 1.1 × 10^5 photons s^-1 and were collected 3 weeks after tumour cell inoculation. Late-stage bone metastases showed a minimal BLI intensity of 8 × 10^4 photons s^-1 and a maximal BLI intensity of 2.5 × 10^6 photons s^-1 and were collected 5 weeks after tumour cell inoculation.

Histological and morphometric analyses of bone metastases showed that early-stage lung metastases are comprised of clusters of ~50–200 cells, and late-stage metastases comprised of ~2,000–10,000 cells. Representative images of the different stages are shown in Supplementary Fig. 1e. The generation of ‘mismatched’ samples (Br-M in bone or Bo-M in brain) followed the same criteria for early- and late-stage metastasis. For the experiments shown in Supplementary Fig. 3a–e, stromal-derived candidates were chosen to represent genes that show expression changes in an organ-specific manner (see Supplementary Fig. 2a and Table 1e–g). Tumour-derived genes were selected to represent genes with expression changes confined to only one organ (see Supplementary Fig. 2a and Table 1b–d). Early-stage lung metastases were collected 48 h after tumour cell inoculation. The BLI intensity at this time point ranged between 2.1 × 10^4 and 1.7 × 10^5 photons s^-1. Late-stage lung metastases were collected 3 weeks after tumour cell inoculation with an average BLI intensity between 8.1 × 10^3 and 3 × 10^5 photons s^-1. Histological and morphometric analyses of lung metastases showed that early-stage lesions are comprised of cells diffusely present throughout the lung (~2,000–4,000 cells per sectional plane, per entire lung), and late-stage metastatic clusters consist of ~1,000–5,000 cells. Representative images of the different stages are shown in Supplementary Fig. 1f. Late-stage lung metastases from the spontaneous MMTV-PyMT breast-to-lung metastasis model were collected from 14-week-old female PyMT mice.

For tissue isolation, mice were anesthetized with 10 mg ml^-1 ketamine; 1 mg ml^-1 xylazine and retro-orbitally injected with 15 mg ml^-1 luciferin. Mice were then intracardially perfused with PBS. Tumour-burdened tissue was identified by the presence of BLI signal for brain and bone metastases. For lung metastases, part of the left lung lobe was collected. Snap-frozen samples were collected for RNA and protein isolation and tissues were fixed in 4% paraformaldehyde (PFA) for histology.

Microarray analysis. For microarray analysis, all samples were prepared and processed by the Genomics Core Facility at MSKCC. RNA was isolated using Trizol (Invitrogen) and the quality was assessed on an Agilent Bioanalyzer. Total RNA was reverse transcribed and labelled with the Genechip® 3 IVT Express Kit (Affymetrix). The resulting cRNA was hybridized to HuMu ProIn chips (Affymetrix). All bioinformatics analyses were completed in R using the Bioconductor suite22. The 'afk' package27 was used for robust multi-array average normalization followed by quantile normalization. Mouse and human sample and probes were normalized separately. With the exception of the cross-species scatter plots, all subsequent bioinformatics analyses regarded the tumour and stroma separately.

The 'limma' package28 was used to identify differentially expressed genes, defined by a fold-change cutoff of ±2 and a false discovery rate of 10%. Tissue-specific genes were identified by the intersection of pairwise comparisons: for example, lung–stroma-specific genes were identified by the intersection of genes significantly enriched in lung versus bone and genes significantly enriched in lung versus brain. Stage-specific genes were identified in a tissue-specific manner comparing early- and late-stage metastases.

Principal component analysis was completed using the covariance matrix in the 'princomp' function in R. Homologues for mouse and human genes were identified using the HomoloGene Database through the NCBI (http://www.ncbi.nlm.nih.gov/ homologene). Homologue pairs were plotted with mouse/stroma tissue-specific, early versus late, fold change on the x axis, and human/tumour tissue-specific, early versus late, fold change on the y axis.

External data sets and survival analysis. For gene expression analysis of mouse endothelial cell lines, raw data from GSE47087 (ref. 34) were imported into R and normalized as above. For patient analysis, normalized gene expression data were downloaded from the GEO (GSE12276; ref. 20). Each gene was mean centred and scaled by standard deviation. Patients were split into tertiles (lower 33%, middle 33%, upper 33%) of CTSS gene expression for Kaplan–Meier survival analysis. The scaled, continuous CTSS gene expression was used for hazard ratio calculation. Similar analyses were completed for genes in Supplementary Fig. 4c–e. Survival analysis was completed using the 'survival' package in R (ref. 35). Hazard ratios were determined utilising the 'coxph' function from the 'survival' package. Nominal P values are reported for hazard ratio significance in Supplementary Table 2 with a significance cutoff of 0.05 used to identify genes significantly associated with metastasis-free survival. P values were generated using the log-rank statistic for Kaplan–Meier analysis and Wald's test for the hazard ratio analysis.

Clinical samples. Primary breast tumours and brain metastases were provided by E.B., P.K.B. and W.C.H. This research was approved by MSKCC Institutional Review Board (IRB) protocol 06-107 and Dana-Farber/Harvard Cancer Center (DF/HCC) protocol 10-454. All participants provided written consent. Information about the clinical samples can be found in Supplementary Table 3.

Generation of CTSS KD lines. Five shRNA sequences targeting CTSS were obtained from the RNAi Codex and RNAi Consortium. shRNA sequences were inserted into the targeting hairpin sequence for the pRetroSuper vector. Correct insertion into the vector was verified by digestion and sequencing of the vector. Plasmids with the correct shRNA targeting sequence were transfected into H292 viral packaging cells. Viral particles were concentrated from the H292 cell supernatant, added to the target cells in the presence of Polybren and cells were selected with puromycin. The following shRNA sequences were used for their KD efficiency: CTSS-1 shRNA, 5′-CTAAATTAACCCAGCTA-3′; CTSS-2 shRNA, 5′-CAGAGAATATCACATATA-3′; CTSS-3 shRNA, 5′-GATAAAGTTTGGTCAAG TAA-3′; CTSS-4 shRNA, 5′-CTTAAAAACCTAAAGCTT-3′; and CTSS-5 shRNA, 5′-CCTTTGCCATACGATTTGGA-3′. CTSS-shRNA was used for subsequent experiments to target CTSS with 90% KD efficiency. A non-targeting shRNA (5′-GCCGATAAAAATATACCTGTT-3′) was used as a control.

Targeting tumour- and stroma-derived cathepsin S in vivo. Br-M cells (1 × 10^6 cells; Br-M CTSS KD or Br-M Ctrl) were inoculated into the left ventricle of 6–8-week-old female athymic/nude or C57BL/6 athymic/nude mice. Metastasis formation was monitored once weekly by BR using a Xenogen IVIS-200 Optical In Vivo Imaging System to determine metastasis incidence in the four experimental groups shown in the table in Fig. 4b. Lesions that reached the cutoff of 1 × 10^4 photons s^-1 in a defined region in the head area were scored positive for brain metastasis in the incidence curve presented in Fig. 4b. In addition, numerical values of the increase in BLI intensity presented the kinetics of tumour progression (Supplementary Fig. 7a). An independent cohort of mice was injected with Br-M cells as described above.
and euthanized at day 35 after tumour cell inoculation for subsequent analysis of proliferation, apoptosis, angiogenesis and metastatic outgrowth.

For in vivo extravasation experiments, athymic/nude or Cts5 KO athymic/nude mice were inoculated with 5 × 10^3 Br-M or C57 mice. After tumour cell inoculation and the BLI intensity was plotted relative to the BLI intensity immediately after tumour cell inoculation (0 h time point). Analyses of BLI intensity were performed blinded to the group allocation. No statistical method was used to predetermine sample size.

Identification of cathepsin S inhibitor VBY-999. VBY-999 was provided by Virobay and is part of an extensive structure-based drug discovery program. VBY-999 is a covalent reversible inhibitor with an electrophilic nitrile warhead. The detailed chemical synthesis and structure of compounds in the structural series including VBY-999 can be found in issued US Patent 7,547,701. Recombinant purified human and mouse cathepsin S were used to assess potency of VBY-999 and determine inhibition constants. Activity on the peptide substrate Z-Leu-Arg-MCA was determined in vitro by measuring hydrolysis of the substrate with spectrophotometric quantification of AMC. VBY-999 was preincubated with cathepsin S for 15 min at room temperature (25°C) after which the substrate was added to initiate the 30 min reaction. Assay incubation buffer included 25 mM CH_3COONa, pH 4.5, 2.5 mM dithiothreitol (DTT) and 0.05 M NaCl. Appropriate reaction conditions and peptide substrates for other cysteine and serine proteases were utilized to screen for selectivity of VBY-999 for cathepsin S. VBY-999 has an inhibition constant K_{i} = 300 μM on the purified human cathepsin S enzyme, and >3,000-fold selectivity versus the related cathepsins K, L, B and F. Potency on the closely related cathepsins K, I, and F was K_{i} > 3 μM, with potency on cathepsin B K_{i} = 700 nM. VBY-999 inhibited mouse cathepsin S enzyme activity. VBY-999 has an inhibition constant K_{i} = 690 nM on mouse cathepsin S. No measurable inhibition was detected for any other cysteine, serine or aspartyl proteases tested.

VBY-999 inhibitor preclinical trial. For administration to mice, VBY-999 was formulated in a nanoparticle-based suspension formulation and further diluted in 5% dextrose in water (DSW) at a concentration of 10 mg/ml. Subcutaneous dosing of VBY-999 provided a dosing formulation and route that allowed high and sustained plasma concentrations of the drug to be achieved, which was confirmed using a bioanalytical LC-MS/MS method after 2 and 7 days of treatment (Fig. 8B). This results in full inhibition of the enzyme target for the duration of the trial, following once-daily dosing. To determine whether VBY-999 had sufficient penetration of the CNS to be available for cathepsin S inhibition in the brain and at the blood–brain barrier (BBB), and to confirm that concentrations in the brain remain stable throughout the duration of the trial, VBY-999 concentration was determined by LC-MS/MS at day 2, day 7 and day 37 after treatment initiation (Fig. 8B). These data indicate that VBY-999 levels in the plasma and brain were significantly above the required concentration for target inhibition at the time of tumour cell inoculation and remain stable throughout the 37-day treatment schedule at a level sufficiently greater than the enzyme inhibition constant, and are thus expected to effectively inhibit cathepsin S activity. For the prevention trials, mice were randomly assigned into vehicle and VBY-999 treatment groups and treatment was started two days before tumour cell inoculation (d = −2). Mice were dosed with 10 mg/kg VBY-999 intraperitoneal (IP) or vehicle (D5W) by subcutaneous injection once daily for 7 days starting two days before tumour cell inoculation (d = −2). At day 0, athymic/nude mice were inoculated with 1 × 10^6 Br-M Ctrl cells or 1 × 10^6 Br-M M, with 0.03% Tween for 20 min. Donkey anti-mouse HRP-labelled secondary antibody (Jackson Immunoresearch, West Grove, PA) was used at a dilution of 1:200 in amplification buffer for 8 min. Following the signal amplification reaction, donkey anti-goat Alexa568 (Molecular Probes) was used at a dilution of 1:200 in 0.25% PNB for 1 h at room temperature. The automated deparanization/rehydration, citrate buffer-based antigen retrieval and blocking of unspecific protein binding and endogenous peroxidase was followed by incubation with mouse anti-human CD68 (Dako) primary antibody and goat anti-human CTSS (R&D Systems) or mouse anti-human CK (Dako) and goat anti-human CTSS (R&D Systems) overnight at 4°C. Sections were incubated with donkey-anti mouse HRP-labelled secondary antibody (Jackson Immunoresearch, 1:200) in 0.25% PNB buffer in PBS for 1.5 h followed by incubation with Alexa488-labelled tyramide (Invitrogen) at a 1:200 dilution in amplification buffer for 8 min. Following the signal amplification reaction, donkey anti-goat Alexa568 (Molecular Probes) was used at a dilution of 1:500 in 0.25% PNB for 1 h at room temperature. Frozen sections that were used for immunofluorescence, frozen sections were thawed and dried at dry heat and rehydrated. VBY-999 was dissolved in 0.5% PNB (phosphate-NaCl buffer) in PBS for at least 1 h at room temperature, followed by incubation with goat anti-human CTSS primary antibody dilution 1:100 in 0.25% PNB overnight at 4°C. Donkey anti-goat Alexa568 secondary antibody (Molecular Probes) was used at a dilution of 1:500 in 0.25% PNB for 1 h at room temperature.

Tissue preparation and immunostaining. Tissue for frozen histology was fixed in 4% PFA overnight, and brain and lung samples were subsequently transferred into 30% sucrose until the tissue was fully equilibrated. Bone samples were decalcified in 0.5 M EDTA, pH 8.0 for 14 days. All tissues were then embedded in OCT (Tissue-Tek) and 10 μm or 20 μm cryostat tissue sections were used for all subsequent analyses. For immunofluorescence, frozen sections were thawed and dried at room temperature and rehydrated. For the standard staining protocol, tissue sections were blocked in 0.5% PNB in PBS for at least 1 h at room temperature, followed by 3% H_2O_2 in 0.1 M PBS to block endogenous peroxidase activity. Tissue information and dilutions are listed in Supplementary Table 5. Fluorophore-conjugated secondary antibodies (Molecular Probes) were used at a dilution of 1:500 in 0.25% PNB for 1 h at room temperature.
Microscopy and image analysis. Tissue sections and cells on coverslips were visualized under a Carl Zeiss AxioImager Z1 microscope equipped with an ApoTome.2 and a TissueGnostics stage to allow for automated image acquisition. The analyses of proliferation and apoptosis were performed using TissueQuest analysis software (TissueGnostics) as previously described\(^1\). All parameters of metastatic outgrowth and angiogenesis were quantified using MetaMorph software (Molecular Devices). Briefly, vasculature was visualized by Texas red lectin (Vector Laboratories) injections or by staining of the endothelial cell marker CD34. Tumour cells were detected by their expression of the GFP reporter. The area covered by CD34 and GFP staining was quantified. To determine the number of tumour cells that are present within an area of 1 to >4 average tumour cell diameters, the blood vessel area was dilated by 1–4 average tumour cell diameters with an increment of 1 tumour cell diameter and the number of tumour cells in each area was determined. Tumour cells that were localized outside an area of 4 average tumour cell diameter were defined as >4 tumour cell diameter away from CD34 blood vessels as illustrated in Fig. 5b. Vessel density was quantified as the area covered by Texas red lectin relative to the area covered by DAPI.

To histologically quantify the percentage of intravascular, extravasating or extravasated tumour cells (Fig. 6c), brain sections were stained for TUNEL to exclude non-viable tumour cells from the analysis. Brain sections were automatically acquired using TissueQuest software (TissueGnostics), which used a z-stack (5 images above and below the focal plane, 0.3 μm steps, ×20 objective) to generate a maximal intensity projection image of each acquired brain area. Tumour cells were detected by Cell Tracker green (Invitrogen) and vasculature was visualized by Texas red lectin (Vector Laboratories) injections. Tumour cells were counted manually and their localization relative to the vasculature was determined.

For analysis of human samples, 5–10 fields of view were acquired using a ×20 objective (total magnification ×200) and a Zeiss Apotome to ensure that cells were in the same optical section. The number of CK⁺ tumour cells and CD68⁺ macrophages, and their relative CTSS intensities (CTSS index), was evaluated using CellProfiler 2.0 software. A CellProfiler module was generated that allowed for the detection of tumour cells and macrophages based on their DAPI and CK signal, or DAPI and CD68 signal, respectively. The CTSS signal intensity was measured in the whole cell population (DAPI⁺) and associated with a specific cell type (macrophages or tumour cells), and the proportion of CTSS signal associated with CK⁺ tumour cells or CD68⁺ macrophages was calculated relative to the overall CTSS signal intensity in all DAPI⁺ cells. Analyses of histological assessment were performed blinded to the group allocation.

Measurement of vessel permeability. Athymic/nude mice (6–8 week-old) were injected with Evan’s blue dye (30 mg/kg⁻¹) into the tail vein. Thirty minutes after injection, mice were anaesthetized and perfused with acidic fixative (1% paraformaldehyde (Sigma)) to extract Evan’s blue at 60°C overnight. Absorbance was measured at 610 and 740 nm on a spectraMax 340pc plate reader (Molecular Devices).

In vitro and cell-based cleavage assays. Recombinant active CTSS was obtained from R&D Systems. CTSS was activated at 50 ng/μl in 50 mM sodium acetate, 5 mM DTT, 0.25 M NaCl (pH 4.5) for 1.5 h at 37°C. For the in vitro cleavage reaction, activated CTSS was incubated with recombinant proteins in the presence or absence of the cathepsin S inhibitor VBY-999 (10 μM) for 0, 10 or 20 min in 50 mM sodium acetate, 5 mM DTT, 0.25 M NaCl at pH 4.5 and pH 6.0. Details about the recombinant proteins used in the in vitro cleavage assay can be found in Supplementary Table 6. The cleavage reaction was stopped by adding SDS sample buffer and reducing agent (Invitrogen) and the samples were boiled at 95°C for 5 min. Aliquots were subjected to western blot analysis as described above. Information about the antibodies can be found in Supplementary Table 5. All experiments were repeated independently at least three times.

For cell-based cleavage assays, HBMECs were grown to 100% confluence in a 10 cm plate. Conditioned medium from Br-M cells was collected as described above. Two hundred microlitres of concentrated Br-M-conditioned media (collected from two 10 cm plates of confluent Br-M cells) was diluted in 6.5 ml PBS pH 6.0 + 0.05 mM DTT for each 10 cm plate of HBMECs. The cleavage reaction was performed in the presence or absence of VBY-999 (10 μM) for 0, 1 h, 2 h and 4 h. PBS pH 6.0 + 0.05 mM DTT was used as a control. The supernatant from HBMEC cell layers was collected after the indicated time points, concentrated and subjected to western blot analysis as described above.

Proliferation assays. Cell growth rate was determined using an MTT cell proliferation kit (Roche). Briefly, cells were plated in triplicate in 96-well plates (2.5 × 10⁴ for Br-M Ctrl and Br-M CTSS KD cells) in the presence or absence of 0.1–100 μM VBY-999. Ten microlitres of MTT labelling reagent was added to each well and then incubated for 4 h at 37°C, followed by the addition of 100 μM MTT solubilization reagent overnight. Absorbance was measured at 595 nm and 750 nm on a spectraMax 340pc plate reader (Molecular Devices).

Data presentation and statistical analysis. Data are presented as means with standard error (s.e.m.) or as statistical scatter plots using GraphPad Prism Pro5. Numeric data were analysed using unpaired two-tailed Student’s t-test unless otherwise noted. Kaplan–Meier survival curves, heat maps and scatter plots were generated in R v 2.15.2 using the R ggplot2 or ggplot2 packages. P values were generated using the Log-Rank statistic for Kaplan–Meier analysis and Wild’s test for the hazard ratio. P < 0.05 was considered statistically significant. All code used to analyse the data and generate the plots is available at: https://bitbucket.org/bowmanr/joyceclab-humu-brain-met-cstc.

Accession numbers. Data sets generated in this study are available online at NCBI GEO, Accession Number GSE47930. The accession numbers for the previously published data sets that were reanalysed in this study are available online at NCBI GEO, and Accession Numbers GSE12276 (ref. 20) and GSE47067 (ref. 34).

50. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Mol. Cell. Biol. 12, 954–961 (1992).
51. Shree, T. et al. Macrophages and cathepsin proteases blunt chemotherapeutic response in breast cancer. Genes Dev. 25, 2465–2479 (2011).
52. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80 (2004).
53. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307–315 (2004).
54. Smyth, G. K. in Bioinformatics and Computational Biology Solutions Using R and Bioconductor (eds Gentleman, R., Carey, V., Dudoit, S., Irizarry, R. A. & Huber, W.) 397–420 (Springer, 2005).
55. Therneau, T. A Package for Survival Analysis in S. R package version 2.37-7, (2014); http://cran.r-project.org/package=survival.
56. Gochova, V. et al. IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. Genes Dev. 24, 241–255 (2010).
Supplementary Figure 1 Characterisation of the stromal cell types in early- and late-stage brain, bone and lung metastasis. (a-c) Quantification and representative images of in vivo BLI intensity are shown for (a) brain metastases ($n=7$ and 9 mice), (b) bone metastases ($n=6$ and 7 mice), and (c) lung metastases ($n=4$ mice per group) for early and late stages. Circles represent individual mice and horizontal lines represent the mean ± s.e.m. (d-f) Representative images of control tissue and early- and late-stage brain, bone, and lung metastases stained for GFP (tumour cells) and the endothelial marker CD34 or the macrophage markers CD68 (lung and bone) or Iba1 (brain). DAPI staining was used to visualise cell nuclei. Brain metastasis sections were also stained with the astrocyte-specific marker GFAP. Images are representative of three independent samples per stage. Scale bar indicates 50 µm.
Supplementary Figure 2. Tissue- and stage-specific gene expression changes in tumour and stroma. (a) Venn diagram of the tumour-derived genes that are significantly different between early and late metastases across different metastatic sites (Fig. 1c). Of the 308 genes significantly different between early and late stages in either brain, bone or lung metastases, 176 genes change by stage in all three sites. (b) Venn diagram depicting the overlap of the 75 stroma-derived genes that are significantly different between early- and late-stage metastases across brain, bone and lung (Fig. 1d). Unlike the tumour-derived genes depicted in (a), there were no stromal-derived genes that were significantly different between early- and late-stage metastases in all three tissues investigated. (c) Heatmap depicting tissue-specific gene expression in the brain, bone and lung stroma. No tissue-specific genes were identified for tumour-derived genes. Proteases are denoted in purple, endogenous inhibitors in red, and their interacting partners in black. (d) qPCR confirmed the tissue-enriched expression pattern of Htra1 for brain, Mmp13 for bone, and Mmp11 for lung in control tissue, early- and late-stage metastases. n=6, 4, 4 samples for control tissues, n=5, 6, 4 samples for early stages, and n=6, 6, 4 samples for late stages in brain, bone and lung respectively. Graphs represent mean + s.e.m. P values were obtained using two-tailed unpaired t-test: *P<0.05, **P<0.01, and ***P<0.001.
Supplementary Figure 3 Independent validation of differentially expressed genes in experimental brain, bone and lung metastases. qPCR was used to confirm stage-differential expression changes, and to determine if the expression changes in stromal cells in brain or bone metastasis are a general response to tumour cell colonisation of the respective organ, or if the expression changes depend on the tumour cell variant. Hatched bars represent ‘mismatched’ samples (Br-M in bone and Bo-M in brain). Filled bars indicate the ‘matched’ samples. (a-b) qPCR for Ctsb, Cst7, Pcsk1n, Serpin1 (brain) in (a) and Adams4, Adams12, Casp2, Ctsε (bone) in (b) confirmed stage-differential expression changes in a tissue-dependent manner. ‘Mismatched’ samples did not show significant changes. (c, d) qPCR for ADAM21, CTSZ, FAU, TIMP2 (brain) in (c) and ADAM17, CASP3, DPP8 (bone) in (d) confirmed stage-differential expression changes in a tissue-dependent manner. ‘Mismatched’ samples (Br-M to bone and Bo-M to brain) revealed that tumour cells underwent similar significant expression changes between early and late stages as identified for matched samples in the respective tissue. (e) qPCR for SERPINE2 confirmed stage-differential expression changes in Lu-M tumour cells in lung metastasis, while those changes were not present in Br-M and Bo-M cells in brain or bone metastasis, respectively. (f) qPCR for Serpina3n and Timp1 confirmed stage-differential expression changes (control vs. late) in the stroma of lungs from xenografted animals as well as lungs from the immunocompetent, syngeneic MMTV-PyMT breast cancer model. (g) Immunofluorescence staining for Serpina3n and Timp1 in lungs of MMTV-PyMT breast cancer model (upper panels) and Ctsε (red) in co-staining with GFP (tumour cells; green) or Iba1 (macrophages; white) in the syngeneic PyMT-BrM model (lower panels). Images are representative of three independent samples per stage. Scale bar indicates 50 µm. For qPCR validation: n = 11, 15, 15, 13, 7, 15 samples in (a), n = 11, 15, 15, 13, 7, 15 samples in (b), n = 8, 15, 7, 9, 5, 5 samples in (c), n = 8, 6, 15, 12, 3, 5 samples in (d), n = 5 and 9 samples for control lung and late-stage metastases (xenograft model), and n = 6 and 9 for control lung and late-stage metastases (syngeneic model). mRNA expression is depicted relative to early-stage metastases in (a-e) and relative to control tissue in (f). All assays were performed in triplicate and gene expression was normalised to Ubc for all stromal genes, B2M for tumour cell-derived genes. P values were obtained using two-tailed unpaired t-test: NS = not significant, *P < 0.05, **P < 0.01, and ***P < 0.001.
Supplementary Figure 4  Identification of genes associated with metastasis-free survival (MFS) and differentially expressed between the tumour and stroma.  

(a–b) As shown in Figure 3a for brain metastasis, cross-species scatter plots depict expression changes in the tumour and stromal gene space during the transition from early- to late-stage metastasis for (a) bone and (b) lung metastasis. Differentially expressed genes in the stroma (mouse) or in the tumour (human) gene space are shown in pink or black respectively. Genes that are differentially expressed in both the stroma and the tumour gene space are shown in purple. 

(c–e) Differentially expressed genes shown in Figure 1c were analysed for association with MFS for either (c) brain, (d) bone, or (e) lung metastasis, depending on the tissue in which the gene was differentially expressed. Scaled gene expression values were used in a Cox proportional hazards model as described in the methods. Hazard ratios (HR) and 95% confidence intervals (CI) are shown for each organ site. HR <1.0 is associated with better patient prognosis, whereas HR >1.0 is associated with poor patient prognosis. For CTSS, the HR for brain MFS was 1.4 (95% CI: 1.05-1.89; \( P = 0.0209 \)). 

(f) Genes depicted in (c–e) are shown in the Venn diagram, where few genes were found to be significantly associated with MFS in multiple tissues. A single gene, SLPI, was significantly associated with MFS in all three tissues, and high expression of levels of SLPI correlated with poor patient prognosis. Hazard ratio significance was determined using Wald’s test with a nominal \( P \) value cutoff of < 0.05.
Supplementary Figure 5 Tumour cells and macrophages are the major constituent cell types of patient brain metastases and express high levels of CTSS. (a-b) Representative images of (a) primary breast cancer and (b) brain metastases patient samples stained for CTSS (red) in combination with the macrophage marker CD68 (white) or a pan-cytokeratin (CK) antibody to visualise tumour cells (green). DAPI staining was used to visualise cell nuclei (blue). All patient samples listed in Supplementary Table 3 were stained and analysed and representative images are shown. The patient samples shown here represent different subtypes of breast cancer based on ER/PR/Her2 status. Scale bar indicates 100 µm in the upper panel rows and 20 µm in the lower panel rows. (c) Quantification of proportions of CK+ tumour cells and CD68+ macrophages in brain metastases samples (n=7 samples from patients for which there was no matched primary breast tumour tissue available). (d) Combined quantification of proportions of CK+ tumour cells, CD68+ macrophages, and the remaining CK-CD68- cell population in primary breast cancer (BC) (n=6 samples) and brain metastasis samples (BrM) (n=13; either matched to the primary, or unmatched samples). (e) Quantification of the CTSS index as a measure of relative CTSS protein levels in tumour cells and macrophages (n=7 samples). Data in (c) and (e) are presented as box plots. Boxes: values between the 25th and 75th percentile; whiskers: minimum and maximum values; horizontal line: median. Data in (d) are presented as stacked bars + s.e.m.
Supplementary Figure 6  Cathepsin S deficiency does not affect viability of Br-M cells or vessel formation in mice. (a) Quantification of CTSS mRNA expression in Br-M Ctrl and Br-M CTSS KD tumour cell lines (n=7 samples per group). Expression is depicted relative to Br-M Ctrl cells. All assays were run in triplicate and gene expression was normalised to B2M. (b) Western blot analysis of CTSS expression levels in cell lysates and conditioned media (CM) from Br-M Ctrl and Br-M CTSS KD cells. Western blots shows representative results from 3 independent experiments. qPCR and western blotting revealed a 90% knockdown efficiency for CTSS at both the mRNA and protein levels. (c) Representative images of CTSS immunofluorescence staining of Br-M Ctrl and Br-M CTSS KD cell lines. DAPI staining was used to visualise cell nuclei (blue). Scale bar indicates 20 µm. (n=3 replicates). (d) CTSS KD does not affect cell viability in culture as determined by MTT assays (n=4 independent experiments). (e) Quantification of tumour cell apoptosis (percentage of TUNEL^GFP^ cells) on day 35 post-tumour cell inoculation. Ctrl; Ctss WT (n=3 mice), CTSS KD; Ctss WT (n=4 mice), Ctrl; Ctss KO (n=6 mice), and CTSS KD; Ctss KO (n=3 mice). (f) Brain metastases size and vessel density were defined as the area covered by GFP (tumour cells) or CD34 (endothelial cells) respectively, and the GFP-covered area was quantified relative to CD34-covered area using Metamorph image analysis. Ctrl; Ctss WT (n=4 mice), CTSS KD; Ctss WT (n=4 mice), Ctrl; Ctss KO (n=6 mice), and CTSS KD; Ctss KO (n=4 mice) (g) Quantification of vessel density in the brain, defined as the ratio of Texas Red Lectin^ area to total DAPI area (n=4 for WT mice, n=6 for Ctss KO mice), and (h) assessment of vessel permeability by intravenous injection of Evan’s blue dye (n=8 mice for each group). Each western blot shows the representative result of three independent experiments. Uncropped images of blots are shown in Supplementary Fig. 9.
Supplementary Figure 7  Cathepsin S deficiency alters the kinetics of brain metastasis formation, though not bone metastasis, but does not affect tumour growth in a regression trial. (a) Quantification of BLI intensity in vivo at the indicated time points after tumour cell inoculation. Ctrl; Ctss WT (n=16 mice), CTSS KD; Ctss WT (n=15 mice), Ctrl; Ctss KO (n=11 mice), and CTSS KD; Ctss KO (n=10 mice). (b) Schematic of the regression trial experimental design. (c) Quantification of the BLI intensity from d0-d35 after Br-M tumour cell inoculation (n=7 mice per group) in the regression trial setting. (d) Quantification of the BLI intensity in the bone and spine region of vehicle-treated and VBY-999-treated animals from the prevention trial, at d35 following Br-M tumour cell inoculation (n=23 mice for vehicle group and n=21 mice for VBY-999 group). (e) Quantification of the BLI intensity in the bone region at d35 following Br-M tumour cell inoculation, in the four experimental groups. Ctrl; Ctss WT (n=10 mice), CTSS KD; Ctss WT (n=8 mice), Ctrl; Ctss KO (n=10 mice), and CTSS KD; Ctss KO (n=11 mice). (f) Quantification of the BLI intensity in the skull region (which may arise from skull and/or brain lesions) of vehicle-treated and VBY-999-treated animals from the prevention trial, at d35 following Bo-M tumour cell inoculation (n=12 mice for vehicle group and n=13 mice for VBY-999 group). Graphs represent mean ± s.e.m or circles represent individual mice and horizontal lines represent the mean ± s.e.m. P values were obtained using two-tailed unpaired t-test: NS = not significant, * P<0.05.
Supplementary Figure 8 Cathespin S deficiency impairs transmigration in an in vitro BBB assay, and sequence analysis identifies a putative cleavage site for cathepsin S. (a) Pharmacological inhibition of cathepsin S with increasing concentrations of the cathepsin S-specific inhibitor VBY-999 (0 µM (Vehicle) to 100 µM) did not affect Br-M cell viability as determined by MTT assays (n=3 independent experiments). (b) Quantification of the number of transmigrated Br-M Ctrl and CTSS KD cells. n=15 samples per condition with 200 fields of view (FOVs) analysed per sample in three independent experiments. (c) Expression of tight junction and adherens junction proteins was confirmed in an independent data set (GSE47067) of FACS sorted endothelial cells. Only Jam-B and Ocln are significantly enriched in brain endothelial cells compared to either lung or bone endothelial cells. n=3, 3, 6 samples analysed per organ. (d) Western blot analysis of recombinant JAM-A, -B and -C cleavage by CTSS, in the presence or absence of VBY-999 (10 µM), using an antibody that detects the IgG1 domain in the recombinant JAM fusion proteins. (e) Alignment of the amino acid sequence of JAM-A, -B, and -C. Motifs that are conserved in all 3 family members are highlighted in dark purple, and motifs that are conserved in 2 of the 3 JAM family members are depicted in light purple. The putative cleavage location for cathepsin S is indicated by the red box. (f) Quantification of JAM-B cell based cleavage experiments (n=3 independent experiments). Graphs represent mean ± s.e.m. in (a) and (b), and box plots in (c). Boxes: values between the 25th and 75th percentile; whiskers: minimum and maximum values; horizontal line: median. Circles represent individual samples and horizontal lines represent the mean ± s.e.m. in (f). P values were obtained using two-tailed unpaired t-test: NS = not significant, *P<0.05, **P<0.01, and ***P<0.001. Each western blot shows the representative result of three independent experiments. Uncropped images of blots are shown in Supplementary Fig. 9.
Supplementary Figure 9 Uncropped western blots. Uncropped images of immunoblots displayed in the main and supplementary figures. Dashed boxes indicate regions that were cropped.
Supplementary Table Legends

**Supplementary Table 1.** Tumour- and stroma-derived changes in gene expression between early and late stages of brain, bone and lung metastases. Gene symbol, gene name, P value and fold change of expression differences in the experimental metastasis models are indicated in each column. Fold change is depicted such that a negative (-) value is associated with downregulation in late-stage metastases, while a positive value is associated with upregulation in late-stage metastases. For Supplementary Table 1a, positive values are associated with upregulation in early-stage metastases compared to normal lung, while negative values are associated with downregulation in early-stage lung metastases compared to normal lung. To identify subtle changes in gene expression between normal lung and early stage lung metastases, a fold change cutoff of +1.5 was used in conjunction with a nominal P-value cut off of 0.05. For every other analysis, a 2.0 fold change cutoff was used. P values were calculated as described in the methods using a two-tailed Student’s t-test.

**Supplementary Table 2.** Tumour-derived stage-specific genes that are associated with metastasis-free survival (MFS) in patient datasets. (a-c) Hazard ratio, 95% confidence interval (CI), and P values for genes identified from analysis of patient dataset GSE12276, which are also listed in Supplementary Fig. 4. The “metastatic site association” column denotes whether a gene shows association with patient MFS in multiple tissues, as summarised in Supplementary Figure 4f. Tumour-derived differentially expressed genes (DEG) whose expression changes by stage in the experimental model (Fig. 1c, Supplementary Table 1) were assessed for MFS using a cox proportional hazards model, as described in the Methods. Hazard ratios with 95% CIs that do not cross 1.0 are considered significant. Nominal P values were determined using Wald’s test. (d) Summary of the number of DEGs associated with MFS at each organ site. These tests aimed to address whether selecting genes for differential expression at the metastatic site enriched for genes associated with site-specific MFS, and demonstrate that the set of genes differentially expressed in the bone is enriched for genes associated with bone MFS, while the brain and lung DEGs are not significantly enriched for genes associated with brain or lung MFS. (e) Summary of DEGs in the bone, and their association with brain and lung MFS. We sought to determine if the significant enrichment of genes associated with bone MFS in the bone DEG set was specific to the bone, or whether there was also an association with brain and lung MFS. These hypergeometric tests demonstrate that the set of genes differentially expressed in the bone is enriched for genes associated with bone MFS only, and not genes associated with brain or lung MFS. Differences in the numbers between Supplementary Table 1 and here are due to incomplete overlap between coverage of genes on the HuMu ProtIn array, and genes in the GSE12276 patient dataset as indicated. P values were generated using a hypergeometric test.

**Supplementary Table 3.** Summary of patient information of primary breast cancer and brain metastases samples. For patients 1-6, matched pairs of primary breast cancer and brain metastasis samples were available. For patients 7-13 only brain metastasis samples were available. MFS: metastasis-free survival. ER: estrogen receptor, PR: progesterone receptor, HER2: human epidermal growth factor receptor 2, Pos: positive, Neg: negative, N/A: not assessed, MFS: metastasis-free survival.

**Supplementary Table 4.** List of Taqman probes used for qPCR.

**Supplementary Table 5.** List of antibodies used for immunofluorescence (IF) and Western blotting (WB).

**Supplementary Table 6.** List of recombinant proteins used for in vitro cleavage assays.