HHS Public Access

Author manuscript

*Cancer Discov*. Author manuscript; available in PMC 2020 July 01.

Published in final edited form as:

*Cancer Discov*. 2020 January ; 10(1): 86–103. doi:10.1158/2159-8290.CD-19-0384.

# Circulating tumor cells exhibit metastatic tropism and reveal brain metastasis drivers

**Remi Klotz**[1,2], **Amal Thomas**[#3], **Teng Teng**[#1,2], **Sung Min Han**[1,2], **Oihana Iriondo**[1,2], **Lin Li**[1,2], **Sara Restrepo-Vassalli**[4], **Alan Wang**[1,2], **Negeen Izadian**[1,2,5], **Matthew MacKay**[1,2], **Byoung-San Moon**[1,6], **Kevin J. Liu**[1,2], **Sathish Kumar Ganesan**[1,2], **Grace Lee**[1,2], **Diane S. Kang**[1,2], **Charlotte S. Walmsley**[7], **Christopher Pinto**[7], **Michael F. Press**[2,8], **Wange Lu**[1,6], **Janice Lu**[2], **Dejan Juric**[7], **Aditya Bardia**[7], **James Hicks**[4], **Bodour Salhia**[2,9], **Frank Attenello**[10], **Andrew D. Smith**[3], **Min Yu**[1,2]

[1]Department of Stem Cell Biology and Regenerative Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA, 90033, USA.

[2]USC Norris Comprehensive Cancer Center, Keck School of Medicine of the University of Southern California, Los Angeles, California, 90033, USA.

[3]Department of Molecular and Computational Biology, USC David and Dana Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, California, 90089, USA.

[4]Bridge Institute, USC David and Dana Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, California, 90089, USA.

[5]MS Biotechnology program, California State University Channel Islands, Camarillo, California, 93012, USA.

Corresponding author: Min Yu, Harlyne J. Norris Research Tower, 1450 Biggy Street, NRT 3507, Los Angeles, CA 90033. Phone: (323) 442-7943. minyu@med.usc.edu.

Current address for Byoung-San Moon is Therapeutics and Biotechnology Division, Bio Platform Technology Research Center, Korea Research Institute of Chemical Technology, Daejeon, Republic of Korea.

AUTHORS' CONTRIBUTIONS

**Conception and design:** R. Klotz, M. Yu

**Development of methodology:** R. Klotz, A. Thomas, A.D. Smith, M. Yu

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** R. Klotz, T. Teng, S.M. Han, O. Iriondo, L. Li, S. Restrepo-Vassalli, A. Wang, N. Izadian, B.S. Moon, K.J. Liu, S. Kumar Ganesan, G. Lee, D.S. Kang, C.S. Walmsley, C. Pinto, W. Lu, J. Lu, D. Juric, A. Bardia, B. Salhia, F. Attenello, M. Yu

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** R. Klotz, A. Thomas, S. Restrepo-Vassalli, M. MacKay, M.F. Press, J. Hicks, A.D. Smith, M, Yu

**Writing, review, and/or revision of the manuscript**: All authors

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** R, Klotz, A. Thomas, N. Izadian, G. Lee, M. Yu

**Study supervision:** M. Yu

Disclosure of potential conflicts of interest: M. Yu. is the founder and director of CanTraCer Biosciences Inc. A. Bardia has a role in consultant/advisory board for Novartis, Pfizer, Genentech, Merck, Immunomedics, Sanofi, Diiachi and Biothernostics. J. Lu has a role in advisory board for Pfizer, Novartis, Radius and Daiichi. M.F. Press has a role in consultant/advisory board for Biocartis SA, Cepheid, Puma Biotechnology, Science Branding Communications and Zymeworks Inc. The other authors declare no competing interests.

Data and software availability

RNA-seq, ATAC-seq sequencing data are available in GEO, with accession number GSE112856. The codes used to process, generate figures in this study can be found at https://github.com/amalthomas111/CTCproject

[6]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Keck School of Medicine of the University of Southern California, Los Angeles, CA, 90033, USA.

[7]Massachusetts General Hospital Cancer Center, Boston, Massachusetts, 02114, USA.

[8]Department of Pathology, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA.

[9]Department of Translational Genomics, Keck School of Medicine of the University of Southern California, Los Angeles, California, 90033, USA.

[10]Neurological Surgery, Keck School of Medicine of the University of California, Los Angeles, California, 90033, USA.

[#] These authors contributed equally to this work.

## Abstract

Hematogenous metastasis is initiated by a subset of circulating tumor cells (CTCs) shed from primary or metastatic tumors into the blood circulation. Thus, CTCs provide a unique patient biopsy resource to decipher the cellular subpopulations that initiate metastasis and their molecular properties. However, one crucial question is whether CTCs derived and expanded *ex vivo* from patients recapitulate human metastatic disease in an animal model. Here, we show that CTC lines established from breast cancer patients are capable of generating metastases in mice with a pattern recapitulating most major organs from corresponding patients. Genome-wide sequencing analyses of metastatic variants identified semaphorin 4D (SEMA4D) as a regulator of tumor cell transmigration through the blood-brain-barrier and MYC as a crucial regulator for the adaptation of disseminated tumor cells to the activated brain microenvironment. These data provide the direct experimental evidence of the promising role of CTCs as a prognostic factor for site-specific metastasis.

### Keywords

Circulating tumor cells; Brain metastasis; Tumor microenvironment; Semaphorin 4D; Blood-brain-barrier

## INTRODUCTION

Brain metastasis has devastating prognosis and accounts for significant morbidity and mortality in cancer patients (1). Therapeutic options are limited as most systemic therapeutic molecules, such as hydrophobic chemotherapeutic drugs, cannot penetrate the blood brain barrier (BBB), and surgical resection is only offered in a limited number of clinical scenarios (2). The brain microenvironment is unique due to the tight control imposed by the BBB to prevent breaches by most immune and tumor cells. Thus, compromising the BBB plays a crucial role in the metastatic colonization of tumor cells to the brain (3). Although important insights have been achieved using various cell line modeling and imaging, our molecular knowledge of the pathophysiology of brain metastasis is still limited, which impedes the development of predictive and therapeutic approaches (1,4). To the best of our knowledge, no studies have identified the underlying molecular mechanisms that allow patient-derived

precursors of brain metastasis to transit to the brain. These precursors belong to a subset of circulating tumor cells (CTCs) that enter the bloodstream and are expected to be uniquely capable of extravasation through the BBB (5). Mounting evidence has shown that CTCs reflect disease progression and treatment responses and therefore have considerable promise as a “liquid biopsy” for monitoring active tumor biology (6). However, these brain metastasis-initiating CTCs have yet to be fully characterized mostly because of the difficulties to study the biological properties of this population due to the scarcity of CTCs recovered from each patient’s blood sample (7). To overcome this limitation, we have investigated the recently-established first cohort of long-term maintained patient-derived CTC lines via *ex vivo* cultures of CTCs isolated from patients with metastatic luminal breast cancers (8). This unique cell resource provided novel insights into molecular features that allow a subset of CTCs to adapt and grow in the brain.

## RESULTS

### Patient-derived CTC lines recapitulate human metastases in mice

A crucial question is whether CTCs isolated from cancer patient blood can generate metastases with similar tropism(s) in mice, thus reflecting their metastasis-initiating properties (9). In order to assess the metastatic potential of these luminal type breast cancer patient-derived CTC lines, we have utilized an experimental mouse model for metastasis by injecting GFP-LUC labeled CTCs directly into the left ventricle of the heart in female immunodeficient NSG mice (Fig. 1A). The ability of 4 CTC lines (BRx07, BRx42, BRx50 and BRx68) to invade and colonize an organ was monitored by bioluminescent imaging for at least 5 months (Fig. 1B). BRx07 and BRx68 have a high metastatic potential (more than 80% of mice had metastases after 3 months) with the generation of simultaneous metastases in the bones, lungs and ovaries. However, the BRx07 and BRx68 mice remained brain-metastases free for up to 8 months. In contrast, BRx50 and BRx42 demonstrated a metastatic preference for the brain despite their low metastatic potential (Fig. 1C; Supplementary Fig. S1A). Of note, the metastatic signal evolved quite differently over time for each organ (Fig. 1C), reflecting a possible dynamic interaction of tumor cells with the microenvironment during colonization of secondary organs. Interestingly, the metastatic tropism of each CTC lines in mice partially reflected the secondary lesions as found in the corresponding patients, as shown by clinical data (Fig. 1D; Supplementary Table S1). The fact that the metastatic recapitulation is not one to one exact match of patient metastases could be due to the differences between species, or to the possibility that CTCs may shed from the most active metastases—therefore the metastases generated in mice may indicate the origin of the CTCs and/or the potential capacity of those CTCs in subsequent organs. Remarkably, of the 4 breast cancer patients from whom CTC lines were generated, BRx42 patient developed a brain metastasis one year after CTC isolation (Fig. 1D; Supplementary Table S1). CTC line BRx42 generated from this patient showed the highest risk of brain metastases in mice (Fig. 1B and 1C; Supplementary Fig. S1A). Next, we explored the possibility of enriching tumor cell subpopulations with enhanced metastatic activity to the bone, lung, and brain by an *in vivo* selection for specific metastatic tropisms. After CTC intracardiac inoculation in mice (generation 1), tumors were dissociated, and the resulting metastatic tumor cells (referred to as BrM1, BoM1 or LuM1 for generation 1 brain, bone or

lung metastasis respectively) were subjected to a new round of *in vivo* selection (Fig. 1A). Cell morphology and ER expression were conserved in metastatic variants after 8–12 months of *ex vivo* culture (Supplementary Fig. S1B). In contrast to their respective parental CTC lines, these metastatic variants exhibited a reduced cell proliferation and viability after culture (Supplementary Fig. S1C and S1D). In the BRx50 line, two rounds of *in vivo* selection yielded BRx50BrM2, which exhibited a significant increase in brain metastatic activity (Fig. 1E; Supplementary Fig. S1E; Supplementary Table S2). BRx50BrM2 generated brain metastases in 50% (6 out of 12) of mice (generation 3), whereas parental BRx50 metastasized to the brain in 5% (1 out of 20) of mice (generation 1). However, the brain metastatic activity of BRx50BrM2/3 decreased after a prolonged time in culture. Similar to the parental BRx50, BRx50BrM2 did not metastasize to the lung, but did show an increase in bone and ovary metastases (Supplementary Fig. S1F; Supplementary Table S2). Similarly, we showed that bone tropism can be enriched in CTC lines BRx68. Although the lung metastasis signals did not show statistical significance, BRx07LuM2 generated lung metastases in 100% of the mice (increased from 70% in the first generation), and showed a decrease in bone metastases compared to the parental BRx07 (Fig. 1E; Supplementary Fig. S1F; Supplementary Table S2). Interestingly, BRx68BoM1 were significantly enriched for bone and brain tropism. Whereas parental BRx68 did not form brain metastases in 20 mice, BRx68BoM1 yielded brain metastases in 2 out of 16 mice (Supplementary Table S2). The co-enrichment for bone and brain tropism observed in CTC lines BRx50 and BRx68 suggests a possible partial sharing of metastatic drivers to those 2 organs.

## Global analysis of gene expression and chromatin accessibility in CTC-derived metastases and tumor microenvironment

We next explored global genetic and gene expression changes that mediate organ-specific CTC metastasis. We used RNA-seq and ATAC-seq to assay gene expression and chromatin accessibility in parental CTC and CTC-derived metastatic cells. Tumor cells were detected and sorted by tumor cell-specific GFP expression. Principal component analysis (PCA) revealed a pronounced clustering of samples according to the patient (Fig. 2A and 2B)—in agreement with previous reports of extensive transcriptional diversity between tumors (10). Compared with publicly available RNA-seq data from breast cancer cell lines (11), CTC lines and CTC-derived metastatic cells clustered distinctly, with closest distance to luminal-type cell lines (Supplementary Fig. S2A). Pairwise differential gene expression analyses revealed unique genes differentially expressed in metastatic tumor cells isolated from brain, bone, lung and ovary (Fig. 2C). Ovary metastatic tumor cells demonstrated the greatest number of differentially expressed genes: 2205 genes were uniquely upregulated compared to tumor cells from other sites. Brain and bone metastatic tumor cells shared the largest set of differentially expressed genes: 786 genes are upregulated in both sites (398 genes shared only between brain and bone, as well as 332 genes and 56 genes shared together with lung and ovary). This observation supports our above-mentioned result, suggesting a partial sharing of drivers of brain and bone metastases. Pathway analysis of genes uniquely upregulated in brain metastasis (Supplementary Table S3) indicated peroxisome, oxidative phosphorylation, Huntington’s disease and Parkinson’s disease pathways as the top enriched pathways in brain metastatic cells (FDR =0.013). The complete result of pathway analysis is shown in Supplementary Fig. S2B. To investigate the tumor microenvironment changes in

different metastases, we used RNA-seq to analyze expression changes in stromal cells after tumor formation in the brain, lung and bone microenvironments relative to control stromal cells from tumor free mice. PCA of the gene expression in stromal cells showed a clear tissue-specific clustering with a high variance between metastatic bone samples (Fig. 2D). Stromal gene expression change was the most pronounced between normal and metastatic tissue in the brain (208 genes), followed by bone (63 genes) and lung (2 genes)(Fig. 2E; Supplementary Table S4–S6). Pathway analysis of the upregulated genes in the brain metastatic microenvironment showed the significantly enriched scores in multiple signatures related to immune responses (Fig. 2F), and 54% of the differentially upregulated genes overlapped with a gene signature of activated primary microglia (ref. (12); Fig. 2G; Supplementary Table S7).

## Identification of SEMA4D associated with brain metastasis

Copy number variation (CNV) analyses revealed an amplification of chromosome 9q (chr9q13–34) in BRx50 brain metastatic variants (Fig. 3A; Supplementary Fig. S3A and S3B), which is absent in bulk DNA analysis of parental BRx50. We postulated that the chromosome 9q amplification in BRx50BrM is a clonally selected event for the initiation of brain metastasis. To investigate for the existence of pre-existing amplification in a small subset of CTCs, we sorted 61 single cells from the parental BRx50 for CNV analysis and found that 4 out of 61 CTCs carried the identical chromosome 9q amplification (Fig. 3A). We identified genes residing on chromosome 9q, for which the expression is altered in tumor cells with a preferential tropism for the brain (Fig. 3B; Supplementary Table S8). To further refine those genes, we examined whether there was any association between gene expression in the primary tumor and brain metastasis relapse using a dataset of 204 advanced primary breast tumors with clinical annotation (13). Kaplan-Meier analysis of metastasis free survival (MFS) for bone, lung and brain identified the *SEMA4D* gene as a candidate gene for brain metastasis (Table 1). High expression of the *SEMA4D* in the primary site was associated with a significant decrease of MFS in the brain, but not in the lung and bone (Fig. 3C). Beside *SEMA4D*, the *CTSL* gene was associated with both brain and lung relapse. Given that *SEMA4D* expression is associated with brain relapse only, we decided to further investigate the potential role of SEMA4D in brain metastasis formation. Chromatin accessibility around the *SEMA4D* transcription start site (TSS) was higher in CTC-derived brain metastases compared to other metastatic sites (Fig. 3D). Additionally, CTC lines BRx50 and BRx42, which metastasized to the brain, presented higher SEMA4D expression levels compared to CTC lines with no brain tropism in mice (BRx07 and BRx68) (Fig. 3E; Supplementary Fig. S3C and S3D). SEMA4D expression was further increased in BRx50 brain metastatic cells compared to the parental BRx50 CTC line (Fig. 3E). We analyzed a small cohort of breast cancer patient-derived xenografts (PDX), established by subcutaneous implantation of patient samples into the mouse flank (14). Only one case (PDX-CM13) in this cohort showed spontaneous overt brain metastases reliably and corresponded to the strongest SEMA4D staining (Fig. 3F). We also investigated a cohort of 12 human brain metastasis tumor samples directly obtained from surgical resection and found that SEMA4D was expressed in 7 out of 12 patient samples (Supplementary Fig. S4; Supplementary Table S9). However, the corresponding primary tumors or other metastatic sites were unavailable,

thus there were no comparative studies to determinate the significance of SEMA4D expression in brain metastases.

### SEMA4D mediates brain metastasis by promoting CTCs transmigration through the BBB

Semaphorins constitute a large family of secreted or transmembrane proteins and have versatile roles in axonal guidance and the immune system (15). Most of the effects of SEMA4D are mediated by its interaction with Plexin-B1 (PLXNB1), B2 and CD72 receptors. Interestingly, Plexin-B1 is highly expressed in the brain of female NSG mice compared to lung and bone (Supplementary Fig. S3E). Previous reports showed evidence of the proangiogenic role of SEMA4D through its interaction with Plexin-B1 (16). Plexin-B1 expression was observed at a significantly higher level in human brain microvascular endothelial cells (HBMECs) compared to human lung microvascular endothelial cells (HLMECs), pericytes or astrocytes (Supplementary Fig. S3F and S3G). Therefore, we hypothesized that SEMA4D binding to Plexin-B1 on brain endothelial cells affects the transmigration of CTCs through the BBB, a selective barrier composed of endothelial cells, pericytes and astrocytes. We first tested this hypothesis using an *in vitro* BBB assay. SEMA4D overexpression in the non-brain tropic CTC line BRx07 promoted the ability of the cells to cross the BBB (Fig. 4A, 4B; Supplementary Fig. S3H). Furthermore, preventing SEMA4D-Plexin-B1 receptor interaction by incubating the BRx07-SEMA4D cells prior to BBB assay with a small Plexin-B1 peptide—corresponding to the SEMA4D binding domain—was able to reduce BBB transmigration (Fig. 4C). Similarly, co-incubating BBB co-cultures with a recombinant human SEMA4D protein to block the binding of PlexinB1 prior to the assay suppressed CTC transmigration (Supplementary Fig. S3I), indicating the importance of interaction between CTC membrane-bound SEMA4D and endothelial cell Plexin-B1 in mediating the BBB transmigration. In BRx50BrM2, shRNA knockdown of SEMA4D decreased the ability of these cells to cross the BBB (Fig. 4D; Supplementary Fig. S3J). As these CTC lines were derived from luminal type breast cancer patients and our previous global gene expression analysis showed a proximity of CTC lines with the luminal-type cell lines (Supplementary Fig. S2A), we further confirmed the effect of SEMA4D in breast cancer luminal-type cell lines, MCF7 and T47D. SEMA4D overexpression increased BBB transmigration of both cell lines (Fig. 4E–H). Crossing the BBB is a crucial first step in the initiation of brain metastasis. Therefore, we next examined the effect of SEMA4D *in vivo* at early stages of metastasis with a sensitive luciferase assay using whole organ lysate. Normal organ lysate and lysate with defined number of spiked-in cells were analyzed to establish the limit of detection (LOD) (Supplementary Fig. S3K). At 3 and 7 days after inoculation, this analysis showed that SEMA4D significantly promotes BRx07 cells to infiltrate the brain (Fig. 4I), but not the lung or ovaries (Supplementary Fig. S3L and S3M). Conversely, shRNA knockdown of SEMA4D in brain metastasis forming cells (BRx50BrM) reduced the number of cells infiltrating the brain (Fig. 4J; Supplementary Fig. S3N and S3O). We extended these results with MCF7 and T47D cell lines (Fig. 4K and 4L; Supplementary Fig. S3P and S3Q). Intriguingly, in a long-term observation, SEMA4D did not affect metastatic tumor growth after intracardiac injection of BRx07 (Fig. 4M), indicating that SEMA4D is not sufficient to drive overt brain metastasis. To test whether SEMA4D is necessary in brain metastasis capable tumor cells, we used shRNA-mediated suppression of SEMA4D in the highly metastatic breast cancer cell line MDA-MB-231 and

found that shSEMA4D significantly inhibited brain metastasis formation in the long-term assay (Fig. 4N; Supplementary Fig. S3R and S3S). Since SEMA4D has been previously shown to have a role in angiogenesis (16), we directly injected cells into the brain *via* intracranial injection but found no significant changes in metastatic tumor growth of BRx07 (Supplementary Fig. S3T). Additionally, tumor cell proliferation remained unchanged after overexpressing SEMA4D in BRx07 cells *in vitro* (Supplementary Fig. S3U). These data suggest that SEMA4D is necessary for the initial step of crossing BBB but is not sufficient for tumor cell colonization. Other factors are needed for successful colonization in the brain microenvironment, which are absent in BRx07. As indicated in previous studies (13,17), the complexity of brain metastasis may require activation of multiple pathways with distinct roles for both BBB crossing and adaptation.

## MYC promotes brain metastasis by mitigating the oxidative stress elicited by activated microglia

Following this reasoning, we investigated highly expressed genes in brain metastasis compared to other organ sites that would play a complementary role with SEMA4D in promoting brain metastasis. Given that our data suggested the activation of an immune response in the brain after tumor formation (Fig. 2F), we sought factors that could facilitate tumor cell adaptation into activated brain microenvironment. Activation of astrocytes and microglia serves as a major component in the brain immune response and engages the production of cytokines and oxidative stress (18,19). Our second most significantly upregulated gene in brain metastases relative to other metastases is SERPINA5, a Serpin family member that potentially aids disseminated tumor cells to resist the toxic effect of activated astrocytes as previously shown (20) (Supplementary Table S10). Infiltration of activated microglia was observed within regions where BRx07 cells were injected into mouse brains but not in sham injection (Supplementary Fig. S5A). We also confirmed the presence of infiltrating microglia in human brain metastases in a cohort of 11 cancer patients (10 out of 11 patients)(Supplementary Fig. S4). These observations motivated us to investigate the role of proto-oncogene *MYC* in this context because it is the fifth most significantly upregulated gene (following MED1, SERPINA5, HEXIM1, PEBP1 with the highest fold change) in CTC-derived brain metastases relative to other metastases (Fig. 5A; Supplementary Table S10). In addition, amplification of *MYC* (chr8q24.21) was found in brain tropic CTC lines BRx50 and BRx42 but not in BRx07 (Fig. 5B). Indeed, MYC is often expressed in a cohort of brain metastasis tumor samples (9 out of 12) (Supplementary Fig. S4; Supplementary Table S9). Furthermore, MYC expression was higher in the CTC line BRx50 than BRx07 and remained high in BRx50BrM (Fig. 5C) and chromatin accessibility around *MYC* TSS was higher in CTC-derived brain metastases compare to other sites of metastases (Fig. 5D). MYC plays a role in cell cycle progression, survival and cellular transformation (21). We evaluated whether MYC could help tumor cells to mitigate the oxidative stress induced by activated microglia. Studies showed that MYC can promote resistance to oxidative damage by inducing the expression of antioxidant enzymes (22). Indeed, 5 key antioxidant enzymes— Superoxide Dismutase 1, 2 and 3 (SOD1, SOD2 and SOD3), Glutathione Peroxidase 1 (GPX1) and Gamma-Glutamylcysteine Synthetase (GCLC)—were expressed at high levels in MYC-high CTCs and CTC-derived brain metastatic tumor cells (Fig. 5E). Ectopic overexpression of MYC in BRx07 cells

upregulated expression of these antioxidant enzymes (Fig. 5F). Conversely, blocking MYC activity (inhibition of nuclear translocation of MYC) in BRx68 or MDA-MB-231 downregulated expression of these enzymes (Fig. 5G; Supplementary Fig. S5B and S5C). In MCF7 and T47D cell lines we further confirmed MYC-driven upregulation of the GPX1 expression (Supplementary Fig. S5D and S5E), suggesting a direct role of MYC in regulating this gene. We next evaluated the Reactive Oxygen Species (ROS) level in CTCs co-cultured with microglia. Co-culture studies showed that ROS accumulate in CTCs only when microglia is activated after physical contact with tumor cells—not by microglia alone or media exchange—and such accumulation can be mitigated by high levels of MYC in CTCs (Fig. 5H; Supplementary Fig. S5F). Furthermore, to access the contribution of MYC in tumor cell survival under oxidative stress, we compared ROS-induced apoptosis in BRx07 and T47D cells. A reduced apoptosis level was observed in MYC overexpressing cells after treatment with tert-butyl hydroperoxide (TBH), indicating a potential role of MYC in promoting tumor cell survival by mitigating oxidative stress (Fig. 5I and 5J). We evaluated the effect of MYC on tumor growth *in vivo*, by intracranial injection of BRx07 overexpressing MYC. Compared to the control group, despite a similar cell proliferation *in vitro*, MYC stimulated tumor growth in mice brains, suggesting a better adaptation of BRx07 cells in brain microenvironment when MYC was expressed (Fig. 5K; Supplementary Fig. S5G and S5H). Consistent with previous studies (23), we found that MYC dramatically promotes tumor formation of MCF7 and T47D cell lines in the brain (Supplementary Fig. S5I), despite a similar proliferation rate between control and MYC expressing cells *in vitro* (Supplementary Fig. S5H). In cancer patient brain metastases, we assessed a possible association of MYC with proliferation and apoptosis of tumor cells. We detected significantly more *in situ* apoptosis in MYC-negative brain tumors compared to MYC-positive tumors (Fig. 5L; Supplementary Fig. S4; Supplementary Table S9). However, tumor cell proliferation (with Ki67 positivity) in this cohort did not show significant difference associated with MYC status (Fig. 5M; Supplementary Fig. S4; Supplementary Table S9). Finally, to determinate the mechanism underlying MYC-mediated brain metastasis, we tested the effect of GPX1 depletion in MYC overexpressing tumor cells, because of a clear association of GPX1 expression with MYC levels in CTC and breast cancer cell lines (Fig. 5F and 5G; Supplementary Fig. S5C–E), and its upregulation in CTC-derived brain metastases relative to other metastases (5.6 fold change; Supplementary Table S10). ROS-induced apoptosis assay revealed that GPX1 depletion reduced MYC-mediated survival effect in T47D cells (Fig. 5N; Supplementary Fig. S5J). Moreover, GPX1 knockdown decreased the brain tumor activity of T47D-MYC cells *in vivo* (Fig. 5O). These data support a model in which MYC promotes brain metastasis by mitigating oxidative stress via upregulation of GPX1 enzyme.

## Association of SEMA4D and MYC with the brain metastatic phenotype

Next, we examined the contribution of SEMA4D and MYC together in brain metastasis. We explored genetic alteration of both *SEMA4D* and *MYC* genes in brain metastases using a previously reported dataset consisting a cohort of 61 trios of primary tumor, brain metastasis, and matched normal tissue from patients with various cancers (phs000730) (24). Analysis showed that 37 patients (61% of total cases) exhibit *MYC* amplification and 7 patients (12% of total cases) exhibit *SEMA4D* amplification in brain metastasis (Fig. 6A).

Among these patients, 17 out of 37 (46%) and 5 out of 7 (71%) are newly acquired or selected *MYC* and *SEMA4D* amplification in brain metastasis, respectively. In addition, 5 of 7 (71%) brain metastases with SEMA4D amplification also harbor *MYC* amplification (Fig. 6A and 6B). We further evaluated whether expression of both *MYC* and *SEMA4D* can stratify cancer patients. Kaplan-Meier analysis of the data set GSE12276 revealed that high level of SEMA4D and MYC together in primary tumors is significantly associated with brain relapse (Supplementary Fig. S5K). Importantly, we noted that the association of high SEMA4D and GPX1 expression in primary tumors has a pronounced prognostic value for the development of brain metastasis in breast cancer patients (Fig. 6C). Finally, because MYC was amplified and highly expressed in the non-brain tropic CTC line BRx68 (Fig. 5B and 5C), we directly injected BRx68 cells with control or shRNA against MYC into the brain via intracranial injection to test the effect of these cells in colonizing brain. Distinct from BRx07 cells, direct inoculation of BRx68 in brain led to tumor formation with a similar growth rate between control and MYC knockdown within the first 4 weeks (Fig. 6D; Supplementary Fig. S5L). However, shMYC resulted in a reduced brain tumor formation starting from week 5. We then asked whether overexpressing SEMA4D could increase the brain metastatic activity of these cells when introduced into blood circulation. Indeed, expressing SEMA4D in MYC-high BRx68 cells resulted significantly higher brain metastatic activity when injected intracardially, while it had no obvious influence on lung and bone metastasis (Fig. 6E–G). Taken together, these results support the relevance of SEMA4D and MYC in brain metastasis, by contributing to distinct steps of metastasis.

## DISCUSSION

In a direct metastatic tropism analysis, we show that CTCs derived from breast cancer patients exhibit a characteristic tissue tropism by largely recapitulating human metastatic disease in mice, proving the potential utility of CTCs in dissecting metastatic mechanisms and monitoring disease progression. In the case of BRx42, *ex vivo* expanded CTCs—isolated long before the appearance of brain metastasis in patient—showed tropism for brain. To reasonably conclude that CTCs could predict metastasis, more studies with larger sample size would be needed. Despite the limitations of a small patient cohort and an “N of 1” predictive case, this study is a proof of concept that intrinsic molecular features of metastatic precursors amongst CTCs could provide novel insights into the mechanisms of metastasis. Previous studies have characterized molecular features of freshly isolated or *ex vivo* cultured CTCs (25–29) and demonstrated their tumor-initiating properties via direct injection of CTCs into femoral bone or flank of immunodeficient mice (30,31). In this new study, we demonstrated the feasibility of using patient-derived CTCs to identify metastatic precursors that provide novel mechanistic insights into the complicated brain metastasis formation. Although brain is typically considered to be an immune-privileged site, the use of immunodeficient mice in this study has potential limitations. This may be more evident in other metastatic site such as the lung and bone. In future studies it will be important to explore the contribution of immune cells in the metastatic ability of CTCs. Previously, crucial information on mechanisms of brain metastasis was identified from triple-negative breast cancer cell lines (13,32,33). However, molecular mechanisms underlying brain metastasis in the luminal subtype has not been well investigated. We identified members of

protein family previously validated (Serpin A5, Cathepsin L, V and Z) (17,20) (Supplementary Table S8) and novel drivers in brain metastasis. Here, we report a new role of SEMA4D as a mediator of BBB transmigration in CTCs. Beyond its physiologic roles in nervous and immune systems, previous studies report its involvement in tumor progression, including tumor angiogenesis and regulation of tumor invasion (34). As reported previously, SEMA4D enhances bone metastasis via inhibiting the osteoblasts function (35), but no studies have shown its contribution to brain metastasis. Disrupting the interaction between SEMA4D and its receptor, PlexinB1, significantly impaired experimental BBB transmigration, drawing attention to the possibility of therapeutically targeting this interaction in metastatic disease. In addition, we identified the oncogene MYC as a potential cooperating mechanism with SEMA4D to promote brain metastasis. A previous study has shown that MYC can upregulate free-radical scavenging enzymes, such as GCLC, to facilitate cell resistance to oxidative stress from activated immune cells (36). Furthermore, MYC gene amplification is often acquired in lethal distant breast cancer metastases of unamplified primary tumors (37), and is often seen in brain metastases (24). To our knowledge, no one has investigated whether MYC has a direct role for tumor cell adaptation in the brain microenvironment. Based on our data showing that MYC upregulates key redox enzymes, mitigates the accumulation of ROS in tumor cells resulting from activated microglia, promotes tumor cell survival under oxidative stress, and is associated with less apoptosis of tumor cells in cancer patient brain metastases, we propose that MYC may facilitate the disseminated tumor cells that entered the brain parenchyma to resist the initial cytotoxic effect of the activated brain microenvironment. Because MYC itself is a difficult therapeutic target, we investigated MYC-mediating factors involved in tumor cell survival and brain metastasis. In this sense, our study suggests that the role of MYC in brain metastasis may be at least partially mediated by GPX1 enzyme. The association between simultaneous high SEMA4D and GPX1 expression in primary breast tumors and decreased brain MFS implicates these genes as potential therapeutic targets for preventing brain metastasis in patients. In conclusion, these data provides evidence of the promising role of CTC-derived markers as early prognostic factors for organotropic metastases and contribute to the molecular understanding of the distinct, multi-step mechanisms of complex brain metastasis.

## MATERIALS AND METHODS

### Cell culture

CTC lines were derived from patients with luminal breast cancer, as previously reported (8). CTC lines (all female donors) were cultured in ultra-low attachment plates with RPMI 1640 medium, supplemented with EGF (20ng/ml), bFGF (20ng/ml), 1X B27 and 1X antibiotic/antimycotic, in 4% $O_2$ and 5% $CO_2$. CTCs were maintained in culture at constant cell density and because they often show a slow rate of growth, the time spent in culture ranged from 8 to 12 months. Human lung microvascular endothelial cells (HLMEC) were obtained from Lonza in 2016 (certificate of analysis unspecified donor's sex). Primary human cerebral cortex microvascular endothelial cells (HBMEC, male donor) were purchased from Cell Systems in 2016. HBMEC and HLMEC were cultured in endothelial cell media (ECM kit, ScienCell Research Laboratories). Primary human astrocytes (HA) were purchased from

Sciencell in 2016 and cultured on poly-L-lysine-coated cell culture dishes in astrocyte cell media (AM kit, ScienCell Research Laboratories). Primary human pericytes were provided by Dr. Ruchi Bajpai from USC in 2016 (sex of cells unspecified by provider) and cultured in pericyte cell media (PM kit, ScienCell Research Laboratories). Human microglia cell line (HMC3, human fetal brain-derived, sex of cells unspecified by provider) was purchased from ATCC in 2018 and maintain in EMEM medium supplemented with 10% fetal bovine serum. Breast cancer cell lines MCF7 (female donor) was purchased from ATCC in 2018 and T47D (female donor) was provided by Dr. Julie Lang from USC in 2016. Cell lines were maintained in RPMI medium supplemented with 10% fetal bovine serum. Cultures were assayed routinely for mycoplasma contamination using the mycoplasma detection kit (MycoAlert, Lonza Group). Cell lines authentications were performed by autosomal STR profiles by University of Arizona genetics core.

### Patient-derived brain metastases samples

Written informed consents were obtained from cancer patients undergoing neurosurgery. Brain metastases samples were collected under the IRB protocol (HS-18–00450) approved by the institutional review board at Keck School of Medicine of USC (Los Angeles, CA) and in accordance with Belmond Report ethical guidelines. Cancer patients who were scheduled for brain metastasis removal surgery at the Neurosurgery department at Keck School of Medicine Hospital were approached before the surgery to consent for utilizing the removed tumor tissue for research purpose. Besides clinically determined eligibility for removal surgery, there is no selection for cancer type, disease characteristics or patient population for the samples used in this study. Except the cancer type, the patient information is de-identified. Therefore, we only reported the cancer types of these samples (Table S9). Freshly removed brain metastasis tissues were immediately transported to the laboratory and divided into several pieces and fixed in 10% formalin for paraffin embedding or cold 4% paraformaldehyde (PFA) for tissue freezing. Method for tissue freezing consisted in 4% PFA fixation for 4.5 hours. After fixation samples were washed in PBS three times and cryoprotected in 30% sucrose overnight. The next day samples were transferred in OCT solution and promptly frozen in dry ice and then stored at −80°C until cryosectioning.

### Experimental CTC-derived metastatic variants

All animal experiments were carried out in accordance with approved protocols from the Institutional Animal Care and Use Committee of USC. Metastatic tumors were established by inoculation of $1 \times 10^5$ GFP-LUC labeled CTCs in 100 μl of PBS into the left cardiac ventricles of 6–8 weeks old female NSG mice supplemented with subcutaneous slow release estrogen pills. Metastasis formation was monitored every 2 weeks by *in vivo* imaging using IVIS Lumina III (PerkinElmer) following intraperitoneal injection of 100 μl of D-luciferin substrate (Syd Labs). Metastases were confirmed by *ex vivo* bioluminescence imaging and resected under sterile conditions. Brain lesions were placed in brain tumor dissociation medium (Miltenyi Biotec). Lung, ovary and kidney lesions were minced and placed in dissociation medium containing RPMI supplemented with 2 mg/ml collagenase I and 15 U/ml DNAse. Samples were further dissociated into single-cell suspension by automated dissociation using the gentleMACS dissociator. Bone lesions were placed in tumor dissociation media (Miltenyi Biotec) and gently ground 3 times in a mortar. Bone tissues

were further dissociated in tumor dissociation media on an orbital shaker at 37°C for 45 minutes. After dissociation, all tissues were filtered through a cell-strainer (70 μm), and cells were washed twice in PBS, then resuspended in PBS with 1% BSA. GFP+ cells were sorted for further propagation in culture, inoculation in mice, or RNA and ATAC sequencing. GFP- cells were also sorted and used for RNA extraction.

### Generation of CTC-derived brain tumor in NSG mice

Brain tumors were established by inoculation of $5 \times 10^4$ GFP-LUC labeled CTCs in 3 μl of RPMI into brain of 6–8 weeks old female NSG mice supplemented with subcutaneous slow release estrogen pills. The coordinates for injection of tumor cells were used for intracerebral tumor establishment, where a burr hole was drilled 2 mm to the right of the bregma and 1 mm anterior to the coronal suture to a depth of 3 mm. Mice were monitored weekly and bioluminescent signal was normalize by signal from day1. Kaplan-Meier curves were generated with an endpoint defined as the moment a given mouse had reached 10 or 50 tumor signal fold change.

### *Ex vivo* luciferase activity assay

Luciferase activity assay was performed as previously described (38). Briefly, freshly resected organs were snap frozen and individually pulverized in fine powder with dry ice-chilled porcelain pestle and mortar. Pulverized samples were weighted then incubated and vortexed with Promega reporter lysis buffer for 15 min. Alternating freeze and thaw was performed three times with liquid nitrogen and 37°C water bath, and samples were centrifuged at 12,000 x g for 10 min. Each supernatant (20 μl) was mixed with 100 μl of luciferase assay reagent (Promega) and luciferase activity was measured in single transparent tube by Lucetta™ Luminometer (Lonza) with 2 seconds for delay time and 10 seconds for read time. The specificity of the luciferase activity was shown by spiking LUC+ CTCs in defined numbers in tumor-free brain. Presence of metastases was determined by Limit Of Detection (LOD) calculated with formula: $LOD = mean_{blank} + 1.635(SD_{blank}) + 1.635(SD_{lowconcentration\ sample})$, where $mean_{blank}$ and $SD_{blank}$ are the mean and standard deviation of the replicates of a blank sample, and $SD_{low}$ concentration sample is the standard deviation of the replicates of the sample containing the lowest concentration of the cell lysate. LOD for bone and lung were based on luciferase activity from normal (tumor-free) bone and lung organ respectively.

### RNA isolation and gene expression profiling

Sorted GFP positive cells, corresponding to human tumor cells, were collected into a pre-chilled tube maintained at 4°C containing PBS with 1% BSA. RNA was collected from 50,000 sorted cells, according to the manufacturer's instructions (Quick RNA, Zymo). RNA integrity was measured using Bioanalyzer (Agilent). Sequencing libraries were prepared from 100 ng total RNA with KAPA stranded RNA-Seq Kit with RiboErase (Kapabiosystems) according to the protocol supplied by the manufacturer. Single-end 75 bp sequencing was performed using the Illumina NextSeq high output 75 cycle kit at the USC Molecular Genomics Core or CHLA Molecular Pathology Genomics Core. In order to control for the dissociation effects, additional parental CTC libraries of each CTC lines were subjected to the same dissociation protocols used for dissociating metastatic tumor. Tumor

cells sequencing reads were mapped to hg19 (GRCh37) reference using STAR v2.5.2b (39). Genes annotated in the ENSEMBL GRCh37.p13 GTF (release 75) were quantified using HTSeq-count (40). Differential analysis was performed using DESeq2 (41) and genes with FDR ≤ 0.05 were identified as differentially expressed (DE). Benjamini-Hochberg procedure was applied for multiple test correction. Differential genes (DE) across the target metastatic sites were identified after controlling for the cell line, dissociation and culture effects. The DESeq2 normalized mean counts and the shrunken log fold change were used for MA-plots. Over-represented GO molecular terms and KEGG pathway analysis comprising the DE genes were identified using the Bioconductor package GOseq (42). RNA-Seq data of different breast cancer lines were obtained from Daemen et al (11) (GSE48213). Since accurate information of batches of individual samples in the public data is not available, only PCA analysis of CTCs and public data was performed after applying variance stabilization transformation to the normalized read counts. In order to take into account the differences in the variations across the lab, PCA analysis was performed after regressing out lab effect using limma (43). Mice micro-environment samples (sorted GFP-negative cells) were mapped to a custom genome combining hg19 and mm10 (GENCODE GRCm38.p5). Only reads mapped to mm10 genome were quantified and downstream analysis was performed in a similar fashion. For each organ, differentially expressed genes (defined by a FDR ≤ 0.05) in the stroma of CTC inoculated mice were identified by comparing to the stromal cells from the same organ of tumor-free mice after controlling for cell line effects. To compare the mice microenvironment response when exposed to tumor cells to that of activated primary microglia, we compared our results with Das et al. (12) (GSE80304). The differential analysis between LPS activated microglia (4hr) vs control primary microglia was performed using DESeq2 at a FDR level ≤ 0.05. Out of the 181 differentially over-expressed genes in the brain tumor microenvironment, 97 genes are also differentially over-expressed in activated primary microglia. All PCA plots were generated using R with top 1K variable genes across the samples.

## Chromatin accessibility assay

Assay for transposase-accessibility chromatin with high-throughput sequencing (ATAC-seq) was performed, as previously described (44). Briefly, nuclei preparation was generated by resuspension of 25,000 or 50,000 sorted cells in nonionic lysis buffer (10 mM Tris pH7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% (v/v) Igepal CA630). Transposition reaction was performed by using the Tn5 transposase (Nextera kit) at 37°C for 30 minutes. Transposed DNA was further amplified by PCR, and the generated libraries were purified using Agencourt AMPure XP (Beckman Coulter). Library quality was controlled by using a Bioanalyzer high-sensitivity DNA analysis kit (Agilent). Paired-end 75 bp sequencing was performed using a NextSeq 500/500 150 cycles kit at the USC Molecular Genomics Core or CHLA Molecular Pathology Genomics Core. Sequencing reads were trimmed for Nextera adapter sequences with trimgalore and mapped to hg19 reference using Bowtie2 v2.2.8 (45) with parameters -X 2000 –fr –no-discordant –no-mixed –minins 38. Only non-mitochondrial reads with a minimum mapping quality score (≥ 30) were kept for the downstream analysis. Duplicate reads were removed using Picard. All mapped reads were offset by +4 bp for the positive strand and −5 bp for the negative strand (46). The reads mapping to the promoter regions (+/− 2.5Kb of TSS) of all annotated hg19 genes were counted and top 1K variable

sites across samples were used to generate the PCA plot. Accessible sites were identified for each sample using MACS2 (47) with parameters -q 0.01 –shift −100 –extsize 200 –nomodel –nolambda. Peaks intersecting with the ENCODE blacklisted regions were removed. For differential analysis, overlapping peaks across conditions were merged using BEDtools (48) to obtain a union set, and reads aligning to this set were counted using featureCounts (49). Differentially accessible peaks were then identified from this union set using DESeq2 (41). The number of reads mapping to the promoter region (+/− 2.5Kb of TSS) of annotated hg19 genes were quantified and top 1K variable sites were selected for the principal component analysis after the variance stabilizing transformation. The accessibility plots were generated based on the normalized number of Tn5 insertions after pooling the individual groups separately (all the brain metastatic libraries together etc.).

### *In vitro* blood-brain-barrier transmigration assays

*In vitro* BBB was composed with HBMECs (50,000 cells per well) in co-culture with human pericytes (100,000 cells per well) and astrocytes (50,000 cells per well). Artificial BBB was formed for 3 days on a transwell insert with 3 μm membranes coated with gelatin and poly-L-lysine. Three days later, permeability to serum albumin was tested. In the upper chamber, 500ul of Evans blue-conjugated albumin (0.45% in phenol free media) were added and incubated for 30 minutes at 37°C. Absorbance of medium from the bottom chamber was measured at 620 nm. For BBB penetration assays, GFP+ cancer cells (10,000 cells per well) were allowed to transmigrate for 48 hours toward a FBS gradient. Transmigrated tumor cells were quantified in the bottom chamber and in the bottom side of the culture insert. The inserts were washed in PBS once and fixed in 3.7% PFA for 15 minutes. Cells were made permeable with 0.1% Triton X-100 in PBS for 10 minutes, blocked with 5% goat serum (Invitrogen) in PBS with 0.1% Tween 20 for 1 hour at room temperature, and then incubated with GFP primary antibody (ab13970, Abcam) diluted 1:500 in 5% goat serum in PBS overnight at 4°C. Goat anti-chicken Alexa 488 secondary antibody (A-11039, Life Technology) was used at a dilution of 1:500 in 5% goat serum in PBS with 0.1% Tween 20 for 1 hour at room temperature. The membranes from the insert were mounted on a microscope slide. For each experiment, images from 3–5 inserts were taken and the number of GFP+ tumor cells was quantified with 20x objective using Hybrid Cell Count software (BZ-II Analyser, Keyence). Transmigrated tumor cells in the bottom chamber were counted using 20x objective.

### Microglia and tumor cells co-culture

Tumor cells were seeded at a density of $5 \times 10^4$ cells/ml in ultra-low attachment 6 well plates with EMEM medium 10% FBS. On top of the well, co-culture was performed in a transwell insert with 0.4 μm membrane seeded with HMC3 cells ($10^5$ cells per well) or with 50:50 HMC3:tumor cells ratio. Two days after co-culture, Reactive Oxygen Species (ROS) level in cells seeded in lower chamber was evaluated. Cells were incubated with 10 μM of CellRox Orange dye (Invitrogen) at 37°C for 30 min. To verify the specificity of the CellRox reagent, ROS was activated by treatment with 50 μM Luperox® TBH70X just before cell staining. In some experiments, co-cultures were incubated with 5mM N-Aceytyl-L-cysteine. 7-AAD was added to exclude dead cells during recording. Fluorescent cells were detected by FACS ARIAII.

### Annexin V detection assay

Apoptotic cells were detected with an annexin V apoptosis detection kit (BD, Biosciences) and according manufacturer's instruction. Cells were treated with 50 μM Luperox® TBH70X for 24 hours prior annexin V detection. Annexin V positive cells were detected by FACS LSRII.

### Virus production and infection

The pCDH-puro-cMyc and pCDH-puro-control plasmids were purchased from Addgene. The shRNA sequence targeting human SEMA4D used was: 5'-CCGGCCTGAACTTAACATCCTTTAACTCGAGTTAAAGGATGTTAAGTTCAGGTTTTTG-3' (targeting coding sequence) and was cloned into pLKO.1 vector. The pLKO.5-non target shRNA control, shMYC (TRCN0000039642), shGPX1-#2 (TRCN0000418594) and shGPX1-#5 (TRC0000046231) plasmids were purchased from Sigma. For lentiviral particles production, 293T cells were co-transfected with either Lenti-Luc-GFP or pCDH-puro-cMyc in combination with third-generation lentivirus packing vectors (VSVG, PMDL, REV) or second-generation lentivirus packing vectors (VSVG, pCMV-dR8.91), respectively, using TransIT-LT1 transfection reagent (Mirus). Similarly, 293T cells were co-transfected with pLKO.1 SEMA4D-targeting shRNA vectors in combination with second-generation lentivirus packing vectors (VSVG, pCMV-dR8.91). Growth medium was collected 48 and 72 hours post-transfection, then viruses were concentrated with Lenti-X concentrator (Clontech), and viral pellets were resuspended in 300μl PBS. SEMA4D-overexpressing cell were established by transduction with retroviral particles. Retroviral-based packaging vectors—pBabe, Ampho and VSVG—were purchased. The cDNA for full-length human SEMA4D was obtained by RT-PCR and cloned into pBabe with BamHI and EcoRI restriction enzymes. For retroviral particle production, 293T cells were co-transfected with pBabe-SEMA4D, Ampho and VSVG vectors using TransIT-LT1 transfection reagent. Viral supernatant were harvested at 48 and 72 hours post transfection and concentrated with Retro-X concentrator (Clontech), and viral pellets were resuspended in 300μl PBS. CTC lines were infected overnight with 100μl lentiviral particles in 8μg/ml polybrene. Two days later, the cells were selected with 2μg/ml of puromycin for 3 days.

### Western blotting

Cells were harvested following wash with PBS. Cells were lysed by suspension in buffer containing 10 mM TrisHCl (pH 7.4), 5 mM EDTA, 1 % Triton and a protease inhibitor cocktail. Total proteins (10 μg) were run on SDS-polyacrylamide gels and transferred onto a PVDF membrane. Primary antibodies used in this study are: anti-CD100/SEMA4D (610670, BD Biosciences), anti-GAPDH (ABS16, Millipore), anti-HSP90 (ab13492, abcam) and anti-c-MYC (ab39688, abcam). Membranes were probed overnight at 4°C, followed by incubation with an anti-IgG polyclonal antibody conjugated to peroxidase (BioRad) for 1 hour at room temperature. Bands were detected after exposition to a chemiluminescent substrate (Pierce ECL Western Blotting Substrate, ThermoFisher).

### qRT-PCR

Total RNA was purified from various cell samples by using the Quick RNA kit (Zymo), according to the manufacturer's instructions. The cDNA was generated by reverse transcription using the iScript master mix cDNA Synthesis Kit (Bio-Rad) in a 20 μl reaction containing 100 ng of total RNA. Real-time PCR was performed using iQ™ SYBR® Green Supermix (Bio-Rad) on a CFX iCycler real-time PCR instrument (Bio-Rad). The primer sequences used are: human SEMA4D (forward, 5'-AAGCAGCATGAGGTGTATTGG-3'; reverse 5'-AGTTGAGGCACTCTGTCTGTT-3'); human 36B4 (forward, 5'-GTGTTCGACAATGGCAGCAT-3'; reverse, 5'-AGACACTGGCAACATTGCGGA-3'); human GAPDH (forward, 5'- GGAGCGAGATCCCTCCAAAAT-3'; reverse, 5'-GGCTGTTGTCATACTTCTCATGG-3'); human PLXNB1 (forward, 5'-ACCAACTGCATTCACTCCCAA-3'; reverse, 5'-GCACTCATCAGGCATCACAG-3'); human PLXNB2 (forward, 5'-AGCCTCTTCAAGGGCATCTG-3'; reverse, 5'-GCCACGAAAGACTTCTCCCC-3'); mouse plxnb1 (forward, 5'-GTGGTCGTTACGGTCTTATCCA-3'; reverse, 5'- AAGCTGCAAAGAGTACGTCCC-3'); and human MYC (forward, 5'-GGCTCCTGGCAAAAGGTCA-3'; reverse, 5'-CTGCGTAGTTGTGCTGATGT-3'); human SOD1 (forward, 5'-GGTGGGCCAAAGGATGAAGAG-3'; reverser, 5'-CCACAAGCCAAACGACTTCC-3'); human SOD2 (forward, 5'- GCTCCGGTTTTGGGGTATCTG-3'; reverse, 5'-GCGTTGATGTGAGGTTCCAG-3'); human SOD3 (forward, 5'-ATGCTGGCGCTACTGTGTTC-3'; reverse, 5'- CTCCGCCGAGTCAGAGTTG-3'); human GCLC (forward, 5'- GGAGGAAACCAAGCGCCAT-3'; reverser, 5'-CTTGACGGCGTGGTAGATGT-3'); human GPX1 (forward, 5'-CAGTCGGTGTATGCCTTCTCG-3'; reverse, 5'- GAGGGACGCCACATTCTCG-3'). The thermal cycling program was 95°C for 3 minutes followed by 40 cycles of denaturing at 95°C for 15 seconds, annealing at 60°C for 15 seconds, and extension at 72°C for 15 seconds. Specificity of the products was assured by melting curve analysis. The relative transcript level was calculated with the ΔΔCq method and normalized with reference gene 36B4 or GAPDH identified as the most suitable reference genes for RT-qPCR analysis of target gene expression in our samples.

### CNV analysis

Genomic DNA was isolated from bulk sample with Quick-DNA extraction kit (Zymo). Single cells were sorted by FACS into individual wells of 96-well PCR plates, using the FACS Aria II single-cell sorting protocol with specific adjustments. Each well of 96-well PCR plates was preloaded with 2μl volume of lysis buffer (50mM DTT and 200 mM KOH). Whole genome amplification (WGA) of single cells was carried out using the WGA4 Genomeplex Single Cell Whole Genome Amplification Kit (Sigma Aldrich Cat#. WGA4). Briefly, the cells in lysis buffer were incubated for 2 minutes at 95°C. A master mix containing 6.5 μl 10mM Tris-HCl-EDTA pH 8.0 per reaction and 1 μl of the 10x Single Cell Lysis and Fragmentation Buffer was added to the cold reaction. The samples were incubated for 4 minutes at 99°C. Further library preparation and amplification were carried out, according the manufacturer's protocol. Amplified DNA was purified using a QIAquick PCR Purification Kit (Thermo Fisher, Cat#. K210012). Concentration of amplified and purified DNA was quantified with Qubit Fluorometric Quantification (Thermo Fisher). Amplified

DNA was sheared using sonication (Covaris S2/E210 focused-Ultrasonicater) with the microtube setup and the 200bp target size protocol for DNA shearing. Fifteen ng of amplified and sonicated DNA from single cells was used for library construction using the NEBNext Ultra DNA Library Preparation Kit for Illumina (New England Biolabs, Cat#. E7370L). The constructed library DNA concentration was quantified with Qubit (Thermo Fisher), and the expected library size distribution of 300–500bp was confirmed using the Agilent 2100 Bioanalyzer (High-Sensitivity DNA Assay and Kit, Agilent Technologies, Cat#. 5067–4626). The individual libraries from barcoded single cells were pooled. The pooled libraries were cleaned using AMPure XP Beads (Beckman Coulter Inc., Cat# A63882). Libraries were sequenced using the Illumina NextSeq 500 or the HiSeq2500 SR50 generating fastq files. Thirty bp were trimmed off the 5' end of each read to remove the WGA4 adapter sequence before alignment to the hg19 reference genome using the Bowtie algorithm. The resulting BAM file was sorted and PCR duplicates were removed using SAMtools. The number of reads falling into each of 5000 'bins' comprising the entire UCSC reference genome, was calculated using a Python script. Finally, an R script utilizing the Bioconductor package, DNAcopy_1.26.0, was used to normalize and segment the bin counts across each chromosome generating a genome-wide CNV profile. The raw sequencing data phs000730 was obtained from NCBI dbGAP (24). The FASTQ files were aligned to hg19 (GRCh37) reference using bwa-mem (50). Sambamba (51) was used to remove duplicates and pileup was created using samtools mpileup v1.5 (52). Relative copy number was estimated between tumor and matched healthy samples using copynumber tool from VarScan2 v2.3 (53) with default parameters. The differences in the sequencing depth was taken into account using data-ratio parameter. After adjusting for GC content using VarScan2 copyCaller, circular binary segmentation was performed and copy number plots were generated using Bioconductor package DNAcopy (R package version 1.52.0.). Segments with logR ≥ 0.2 were considered as amplified.

## Immunofluorescence

Immunofluorescence staining was performed, as previously described (8). CTC or metastatic variant cells were spun onto poly-L-lysine glass slides with Spintrap for 10 minutes at 800 g. Cells were fixed in 4% PFA in PBS for 10 minutes at room temperature and made permeable in PBS with 0.1% Triton X-100 for 10 minutes. Cells were stained with nuclear 4,6-dianidino-2-phenylindole (DAPI), SEMA4D (A38812, Sigma), Ki67 (clone 7B11, Invitrogen). Fluorophore-conjugated secondary antibody (Molecular Probes) were used at a dilution of 1:500 in 5% goat serum in PBS with 0.1% Tween 20. Staining was measured using the Keyence BZ-9000 fluorescent microscope. Images are representative of at least 3 independent images per sample.

## Patient-derived xenografts models

Paraffin-embedded PDX tissues slides were a gift from Dr. Bodour Salhia (USC Norris Comprehensive Cancer Center). Patient was consented for tissue collection under an IRB approved protocol at Geisinger Health System (GHS, Danville, PA). Procedure to establish PDX models has been previously described (14).

### Targeted mutation analysis

Key drivers mutations in CTC lines were previously identified (ref 8; Supplementary Table S1) and stability of mutations allele frequency was accessed in BRx68 line over time in culture with the Ovation® Target Enrichment system (NuGEN). Genomic DNA was isolated from bulk BRx68 with Quick-DNA extraction kit (Zymo) that have been cultured for different duration of time. The library preparation was proceeded with 400 ng genomic DNA according manufacturer's instruction and sequenced at the USC Molecular Genomics Core. Mutation variants allele frequency was identified using PartexFlow® package.

### Immunohistochemical staining

Organs were fixed with 10% formalin overnight and sectioned by USC's histology laboratory service. Sections of formalin-fixed organs were de-paraffinized and rehydrated, and antigen retrieval was performed in 10 mM citrate buffer (pH6) for 15 minutes. Sections were washed and blocked for 1 hour at room temperature. Primary SEMA4D (ab134128, abcam), mitochondria (ab92824, abcam), Iba1 (ab5076, abcam), GFAP (ab53554, abcam), Cytokeratin (349205, BD, Biosciences), MYC (ab39688, abcam), Ki67 (14–5699-82, Invitrogen) or active caspase-3 (ab2302, abcam) antibodies were diluted in antibody diluent (DAKO), and sections were incubated for 15 minutes at room temperature. Sections were incubated with HRP anti-mouse or anti-rabbit (EnVision, DAKO) for 30 minutes at room temperature. Samples were incubated for 5 minutes with DAB (Vector Laboratories) and counterstained with haematoxylin for 45 seconds before mounting. In situ apoptosis detection was performed on paraffin embedded tissues sections with abcam detection kit (ab20638). Assay was performed according manufacturer's instruction and for analysis, dark brown signal was considered as positive staining for apoptosis. Images represent at least 3 independent fields per sample. For each sample, images from 3 fields were taken and the number of dark brown+ (in situ apoptosis assay) or Ki67+ tumor cells was quantified with 20x objective using Hybrid Cell Count software (BZ-II Analyser, Keyence).

### Statistical analysis

Data are presented as mean ± standard deviation (SD), unless otherwise specified. For *in vivo* experiments, sample sizes are noted in the corresponding figures. Statistical analyses were performed using GraphPad Prism Pro7. Numeric data were analyzed using unpaired two-tailed student's *t*-test unless otherwise noted. For patient progression-free survival analysis, normalized gene expression data GSE12276 was used and imported in Partek Genomics Suite 6.6. Kaplan Meier survival curves were generated with patients split into 2 quantiles, below and above the median (low 50%, high 50%) of SEMA4D, MYC or GPX1 expression and *P* values were generated using Log-Rank statistic. $P<0.05$ was considered statistically significant.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

We thank USC Flow Core, USC Histology Core and USC Molecular Genomics Core for their excellent technical support, and D. Ruble and members of CHLA Molecular Pathology Genomics Core for technical assistance. We thank R. Bajpai and C. Griffin for providing human primary pericyte cells; J. Lang for providing T47D cell line; M. Pratt for assisting with glycosylation analysis; K. Shen for providing *ex vivo* luciferase assay protocol. We are grateful to C. Lytal for editing our manuscript. We thank M. Serowoky for his contribution in optimizing retrovirus production protocol. This research was supported by the following: The National Institutes of Health (NIH) grant DP2 CA206653 (M. Yu), the Donald E. and Delia B. Baxter Foundation (M. Yu), the Stop Cancer Foundation (M. Yu), the PEW Charitable Trusts and the Alexander & Margaret Stewart Trust (M. Yu), the SC CTSI pilot grant UL1TR001855 and UL1TR000130 (M. Yu and F. Attenello), a California Institute for Regenerative Medicine (CIRM) postdoctoral fellowship (R. Klotz), and a CIRM Bridges award EDUC2-08381 to CSU, Channel Islands (N. Izadian as an intern). The project described was supported in part (core facilities) by award number P30CA014089 from the National Cancer Institute.

Financial support: The National Institutes of Health (NIH) grant DP2 CA206653 (M. Yu), the Donald E. and Delia B. Baxter Foundation (M. Yu), the Stop Cancer Foundation (M. Yu), the PEW Charitable Trusts and the Alexander & Margaret Stewart Trust (M. Yu), the SC CTSI pilot grant UL1TR001855 and UL1TR000130 (M. Yu and F. Attenello), a California Institute for Regenerative Medicine (CIRM) postdoctoral fellowship (R. Klotz), and a CIRM Bridges award EDUC2-08381 to CSU, Channel Islands (N. Izadian as an intern). The project described was supported in part (core facilities) by award number P30CA014089 from the National Cancer Institute.

## REFERENCES

1. Maher EA, Mietz J, Arteaga CL, DePinho RA, Mohla S. Brain metastasis: opportunities in basic and translational research. Cancer Res. 2009;69(15): 6015–20. [PubMed: 19638593]
2. Fidler IJ. The Biology of Brain Metastasis: Challenges for Therapy. Cancer J. 2015;21(4): 284. [PubMed: 26222080]
3. Quail DF, Joyce JA. The Microenvironmental Landscape of Brain Tumors. Cancer Cell. 2017;31(3): 326–41. [PubMed: 28292436]
4. Zhang C, Yu D. Advances in decoding breast cancer brain metastasis. Cancer Metastasis Rev. 2016;35(4): 677–84. [PubMed: 27873078]
5. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586): 298–306. [PubMed: 26791720]
6. Pantel K, Alix-Panabières C. Liquid biopsy and minimal residual disease — latest advances and implications for cure. Nat Rev Clin Oncol. 2019;16(7):409–424. [PubMed: 30796368]
7. Yu M, Stott S, Toner M, Maheswaran S, Haber DA. Circulating tumor cells: approaches to isolation and characterization. J Cell Biol. 2011;192(3): 373–82. [PubMed: 21300848]
8. Yu M, Bardia A, Aceto N, Bersani F, Madden MW, Donaldson MC, et al. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. Science. 2014;345(6193): 216–20. [PubMed: 25013076]
9. Gkountela S, Aceto N. Stem-like features of cancer cells on their way to metastasis. Biol Direct. 2016;11: 33. [PubMed: 27457474]
10. Network TCGAR, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics. 2013; 45:1113–1120. [PubMed: 24071849]
11. Daemen A, Griffith OL, Heiser LM, Wang NJ, Enache OM, Sanborn Z, et al. Modeling precision treatment of breast cancer. Genome Biol. 2013;14(10): R110. [PubMed: 24176112]
12. Das A, Kim SH, Arifuzzaman S, Yoon T, Chai JC, Lee YS, et al. Transcriptome sequencing reveals that LPS-triggered transcriptional responses in established microglia BV2 cell lines are poorly representative of primary microglia. J Neuroinflammation. 2016;13(1): 182. [PubMed: 27400875]
13. Bos PD, Zhang XH-F, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. Nature. 2009;459(7249): 1005–9. [PubMed: 19421193]
14. Tew BY, Legendre C, Schroeder MA, Triche T, Gooden GC, Huang Y, et al. Patient-derived xenografts of central nervous system metastasis reveal expansion of aggressive minor clones. Neuro-Oncol. 2019

15. Worzfeld T, Offermanns S. Semaphorins and plexins as therapeutic targets. Nat Rev Drug Discov. 2014;13(8): 603–21. [PubMed: 25082288]
16. Basile JR, Barac A, Zhu T, Guan K-L, Gutkind JS. Class IV semaphorins promote angiogenesis by stimulating Rho-initiated pathways through plexin-B. Cancer Res. 2004;64(15): 5212–24. [PubMed: 15289326]
17. Sevenich L, Bowman RL, Mason SD, Quail DF, Rapaport F, Elie BT, et al. Analysis of tumour- and stroma-supplied proteolytic networks reveals a brain-metastasis-promoting role for cathepsin S. Nat Cell Biol. 2014;16(9): 876–88. [PubMed: 25086747]
18. Liddelow SA, Barres BA. Reactive Astrocytes: Production, Function, and Therapeutic Potential. Immunity. 2017;46(6): 957–67. [PubMed: 28636962]
19. Ransohoff RM, Perry VH. Microglial physiology: unique stimuli, specialized responses. Annu Rev Immunol. 2009;27: 119–45. [PubMed: 19302036]
20. Valiente M, Obenauf AC, Jin X, Chen Q, Zhang XH-F, Lee DJ, et al. Serpins Promote Cancer Cell Survival and Vascular Cooption in Brain Metastasis. Cell. 2014;156(5): 1002–16. [PubMed: 24581498]
21. Stine ZE, Walton ZE, Altman BJ, Hsieh AL, Dang CV. MYC, Metabolism, and Cancer. Cancer Discov. 2015;5(10): 1024–39. [PubMed: 26382145]
22. Kfoury A, Armaro M, Collodet C, Sordet-Dessimoz J, Giner MP, Christen S, et al. AMPK promotes survival of c-Myc-positive melanoma cells by suppressing oxidative stress. EMBO J. 2018;e97673. [PubMed: 29440228]
23. Lee HY, Cha J, Kim SK, Park JH, Song KH, Kim P, et al. c-MYC Drives Breast Cancer Metastasis to the Brain, but Promotes Synthetic Lethality with TRAIL. Mol Cancer Res MCR. 2019;17(2): 544–54. [PubMed: 30266755]
24. Brastianos PK, Carter SL, Santagata S, Cahill DP, Taylor-Weiner A, Jones RT, et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. Cancer Discov. 2015;5(11): 1164–77. [PubMed: 26410082]
25. Zhang L, Ridgway LD, Wetzel MD, Ngo J, Yin W, Kumar D, et al. The identification and characterization of breast cancer CTCs competent for brain metastasis. Sci Transl Med. 2013;5(180): 180ra48.
26. Boral D, Vishnoi M, Liu HN, Yin W, Sprouse ML, Scamardo A, et al. Molecular characterization of breast cancer CTCs associated with brain metastasis. Nat Commun. 2017;8(1): 196. [PubMed: 28775303]
27. Ozkumur E, Shah AM, Ciciliano JC, Emmink BL, Miyamoto DT, Brachtel E, et al. Inertial Focusing for Tumor Antigen–Dependent and –Independent Sorting of Rare Circulating Tumor Cells. Sci Transl Med. 2013;5(179): 179ra47.
28. Zhang Z, Shiratsuchi H, Lin J, Chen G, Reddy RM, Azizi E, et al. Expansion of CTCs from early stage lung cancer patients using a microfluidic co-culture model. Oncotarget. 2014;5(23): 12383–97. [PubMed: 25474037]
29. Cayrefourcq L, Mazard T, Joosse S, Solassol J, Ramos J, Assenat E, et al. Establishment and characterization of a cell line from human circulating colon cancer cells. Cancer Res. 2015;75(5): 892–901. [PubMed: 25592149]
30. Baccelli I, Schneeweiss A, Riethdorf S, Stenzinger A, Schillert A, Vogel V, et al. Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. Nat Biotechnol. 2013;31(6): 539–44. [PubMed: 23609047]
31. Hodgkinson CL, Morrow CJ, Li Y, Metcalf RL, Rothwell DG, Trapani F, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med. 2014;20(8): 897–903. [PubMed: 24880617]
32. Chen Q, Boire A, Jin X, Valiente M, Er EE, Lopez-Soto A, et al. Carcinoma–astrocyte gap junctions promote brain metastasis by cGAMP transfer. Nature. 2016;533(7604): 493–8. [PubMed: 27225120]
33. Boire A, Zou Y, Shieh J, Macalinao DG, Pentsova E, Massagué J. Complement Component 3 Adapts the Cerebrospinal Fluid for Leptomeningeal Metastasis. Cell. 2017;168(6): 1101–1113.e13. [PubMed: 28283064]

34. Jiang H, Chen C, Sun Q, Wu J, Qiu L, Gao C, et al. The role of semaphorin 4D in tumor development and angiogenesis in human breast cancer. OncoTargets Ther. 2016;9: 5737–50.
35. Yang Y-H, Buhamrah A, Schneider A, Lin Y-L, Zhou H, Bugshan A, et al. Semaphorin 4D Promotes Skeletal Metastasis in Breast Cancer. PloS One. 2016;11(2):e0150151. [PubMed: 26910109]
36. Benassi B, Fanciulli M, Fiorentino F, Porrello A, Chiorino G, Loda M, et al. c-Myc phosphorylation is required for cellular response to oxidative stress. Mol Cell. 2006;21(4): 509–19. [PubMed: 16483932]
37. Singhi AD, Cimino-Mathews A, Jenkins RB, Lan F, Fink SR, Nassar H, et al. MYC Gene Amplification is Often Acquired in Lethal Distant Breast Cancer Metastases of Unamplified Primary Tumors. Mod Pathol Off J U S Can Acad Pathol Inc. 2012;25(3): 378–87.
38. Shen K, Luk S, Elman J, Murray R, Mukundan S, Parekkadan B. Suicide Gene-Engineered Stromal Cells Reveal a Dynamic Regulation of Cancer Metastasis. Sci Rep. 2016;6: 21239. [PubMed: 26893143]
39. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013;29(1): 15–21.
40. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma Oxf Engl. 2015;31(2): 166–9.
41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12).
42. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2): R14. [PubMed: 20132535]
43. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7): e47. [PubMed: 25605792]
44. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109: 21.29.1–9.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4): 357–9. [PubMed: 22388286]
46. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12): 1213–8. [PubMed: 24097267]
47. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9): R137. [PubMed: 18798982]
48. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6): 841–2. [PubMed: 20110278]
49. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma Oxf Engl. 2014;30(7): 923–30.
50. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009;25(14): 1754–60.
51. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinforma Oxf Engl. 2015;31(12): 2032–4.
52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25(16): 2078–9.
53. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3): 568–76. [PubMed: 22300766]

## Significance

Interests abound in gaining new knowledge of the physiopathology of brain metastasis. In a direct metastatic tropism analysis, we demonstrated that *ex vivo* cultured CTCs from 4 breast cancer patients showed organotropism, revealing molecular features that allow a subset of CTCs to enter and grow in the brain.

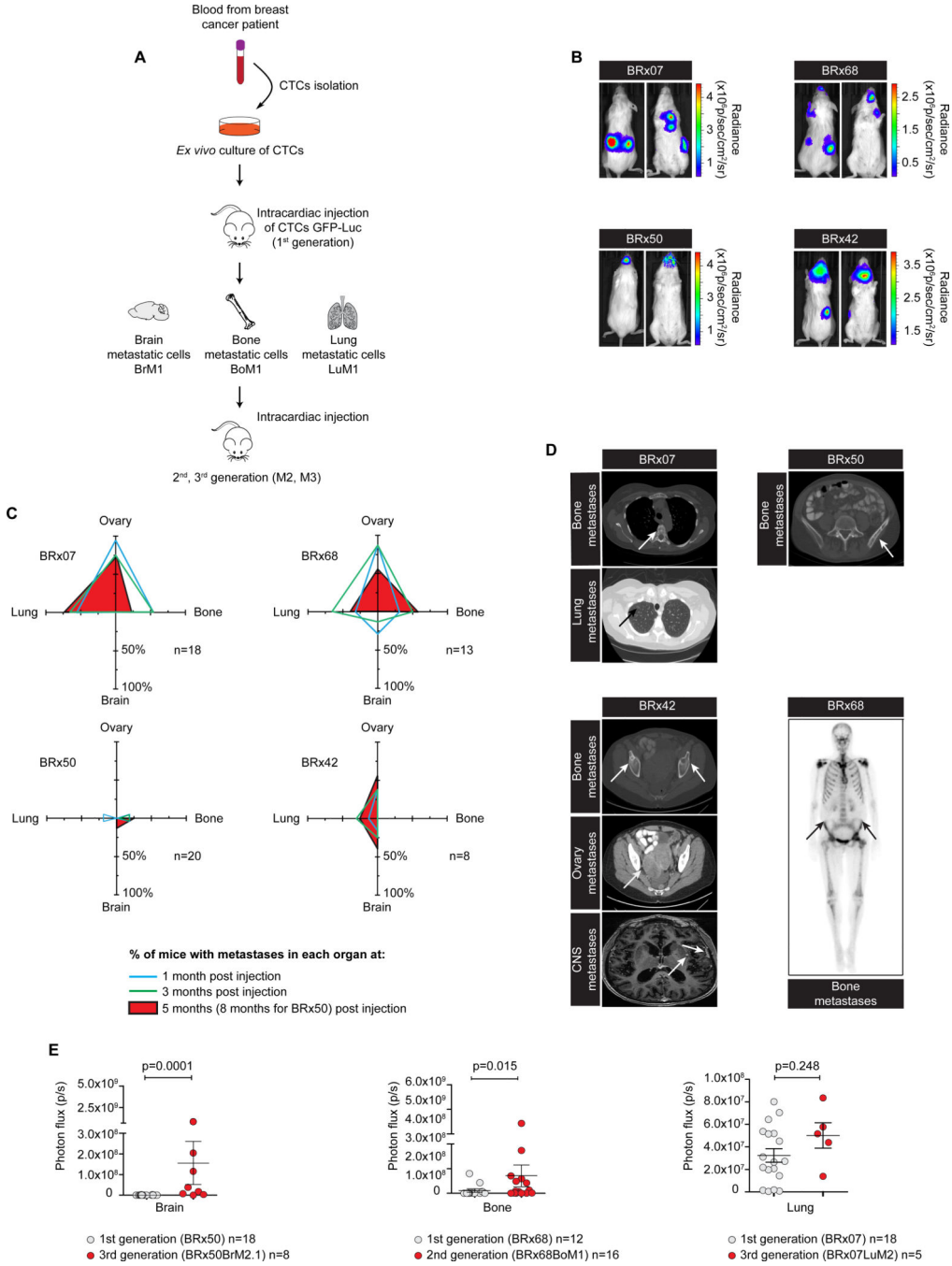


**Figure 1. Patient-derived CTC lines recapitulate human metastases in mice.**

**A**, Schematic showing *ex vivo* expansion of CTCs and *in vivo* inoculating and passaging of CTC lines. **B**, Representative bioluminescence imaging (BLI) data of mice injected with CTC lines at 5 months (BRx07, BRx68 and BRx42) or 8 months (BRx50) after inoculation. Each image shows the back and front of the same mouse. **C**, Plots summarizing the percentages of mice exhibiting ovary, bone, brain and lung metastases at different time point after CTC inoculation, based on *in vivo* and *ex vivo* organ BLI data. n = number of mice. **D**, Scan images of metastatic lesions in patients from whom CTC lines were generated. Arrows

show tumor locations. **E**, Quantification of BLI intensity in brain, bone and lung at 5 months after intracardiac inoculation of CTC or CTC-derived metastatic variants in mice. Circles represent individual mouse, horizontal lines represent the mean ± s.e.m. n = number of mice. *P* values were obtained with Wilcoxon rank sum test.

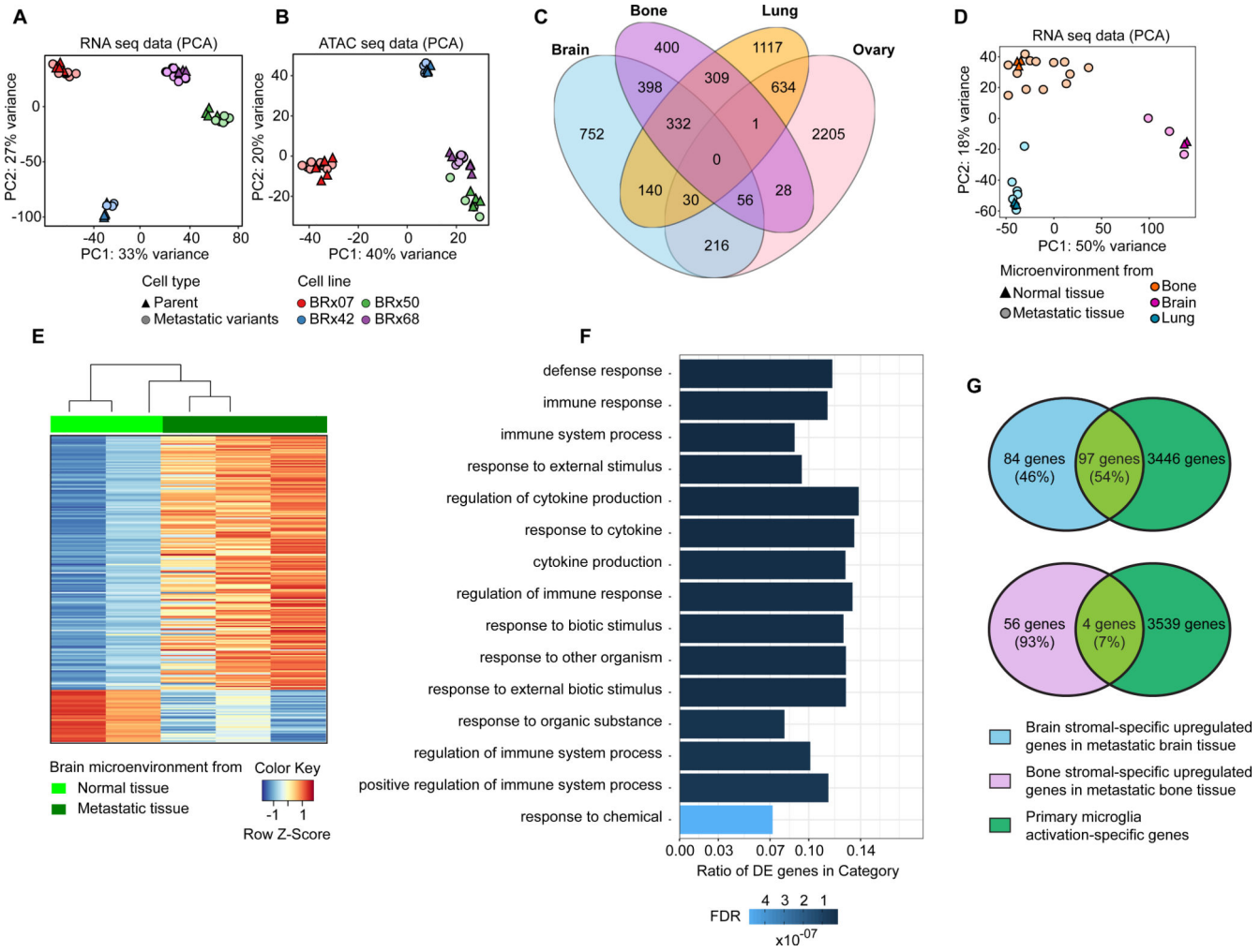


**Figure 2. Global analysis of gene expression and chromatin accessibility in CTC-derived metastases and tumor microenvironment.**

Principal component analysis (PCA) of RNA-seq (**A**) and ATAC-seq (**B**) data in tumor cells. CTC lines are color-coded, and cell types are shape-coded. **C**, Venn diagram depicting differentially expressed genes in the CTC-derived metastatic tumor cells between 4 different metastatic sites. **D**, PCA of RNA-seq data in stromal cells from bone, brain and lung. Organs are color-coded, and conditions are shape-coded. **E,** Heatmap depicting stromal-specific gene expression in normal brain tissue and metastatic brain tissue (differentially expressed genes FDR ≤0.05). **F**, Top enriched gene ontology (GO) pathways of upregulated genes in metastatic brain stromal cells after tumor formation relative to normal brain stromal cells. FDR value is color-coded and the ratio of differentially expressed (DE) genes in each category is represented in the bar graph. **G**, Venn diagram depicting differentially expressed gene observed in each comparison. Primary microglia activation-specific genes were derived from Das et al (12).

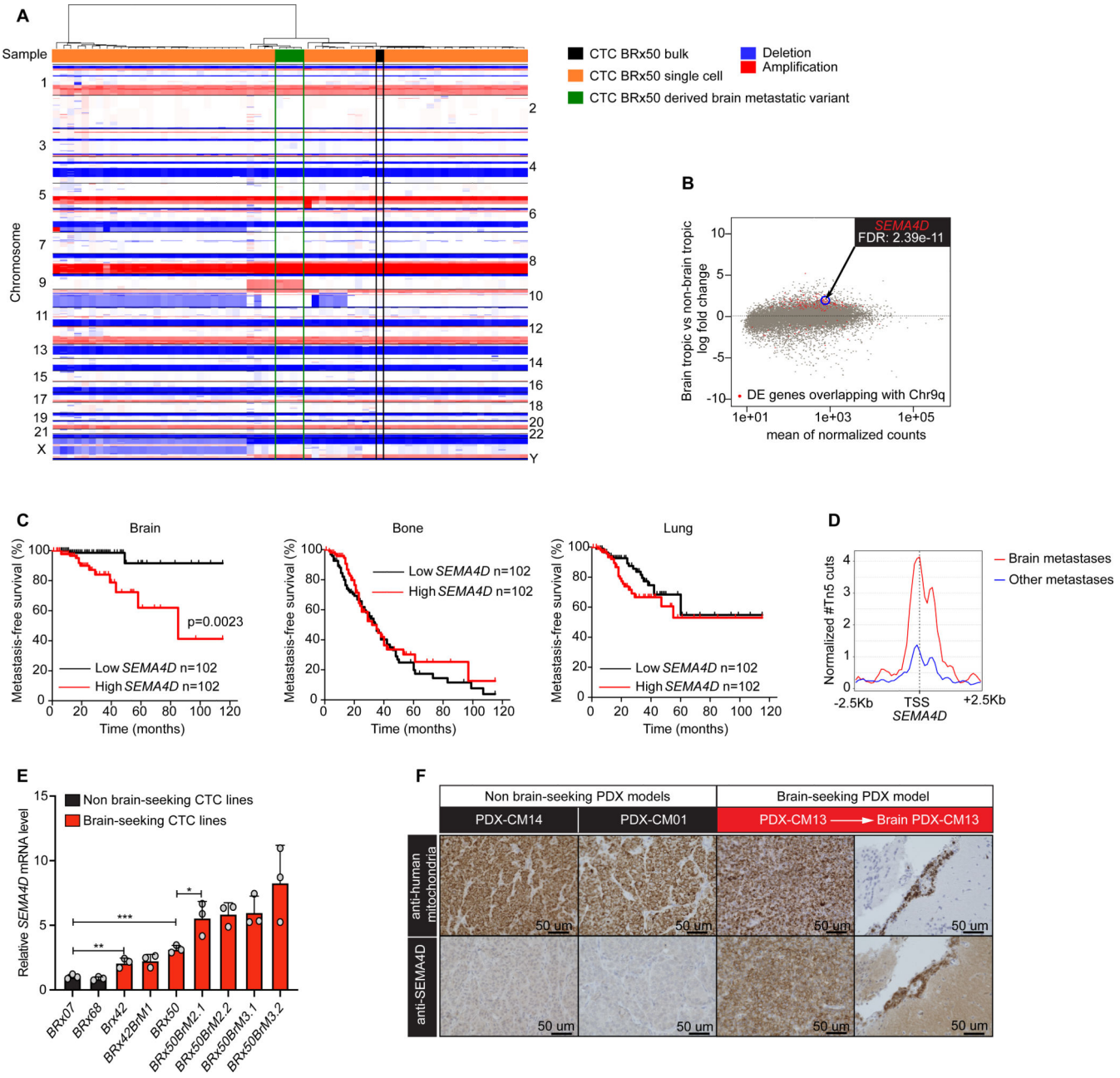


**Figure 3. Identification of SEMA4D associated with brain metastasis.**

**A**, Heatmap of copy number variation (CNV) data for a panel of 61 BRx50 single cells, BRx50 bulk and 4 different BRx50-derived brain metastatic variants. Columns show samples, and rows show chromosomes. **B**, MA plot (log ratio versus abundance) depicting gene expression change in brain tropic tumor cells (BRx50, BRx42 and CTC-derived brain metastases) compared to non-brain tropic tumor cells (BRx07, BRx68 and CTC-derived lung, bone and ovary metastases). Genes differentially expressed (FDR < 0.01) in brain tropic tumor cells and located on chromosome 9q amplified region are highlighted in red. **C**, Kaplan-Meier curves showing metastasis free survival analysis of brain, bone and lung in 204 breast cancer patients (GSE12276). Patients were separated into two equal quantiles of

low and high *SEMA4D* expression. *P* values were determined with log-rank test. **D**, Metaplot comparing chromatin accessibility around *SEMA4D* TSS between CTC-derived brain metastases and CTC-derived other metastases. **E**, Histogram representing *SEMA4D* mRNA relative expression level (mean ± s.e.m., two-tailed unpaired *t*-test with 3 independent replicates). **F**, Representative images of immunohistochemistry staining with antibodies against human mitochondria (top panel) and SEMA4D (bottom panel) in 3 breast cancer brain metastasis PDX tumors in the flank (left 3 columns) and brain metastatic lesion generated by CM13 (Brain PDX-CM13, right column). * *P* value < 0.05; ** *P* value < 0.01; *** *P* value < 0.001.

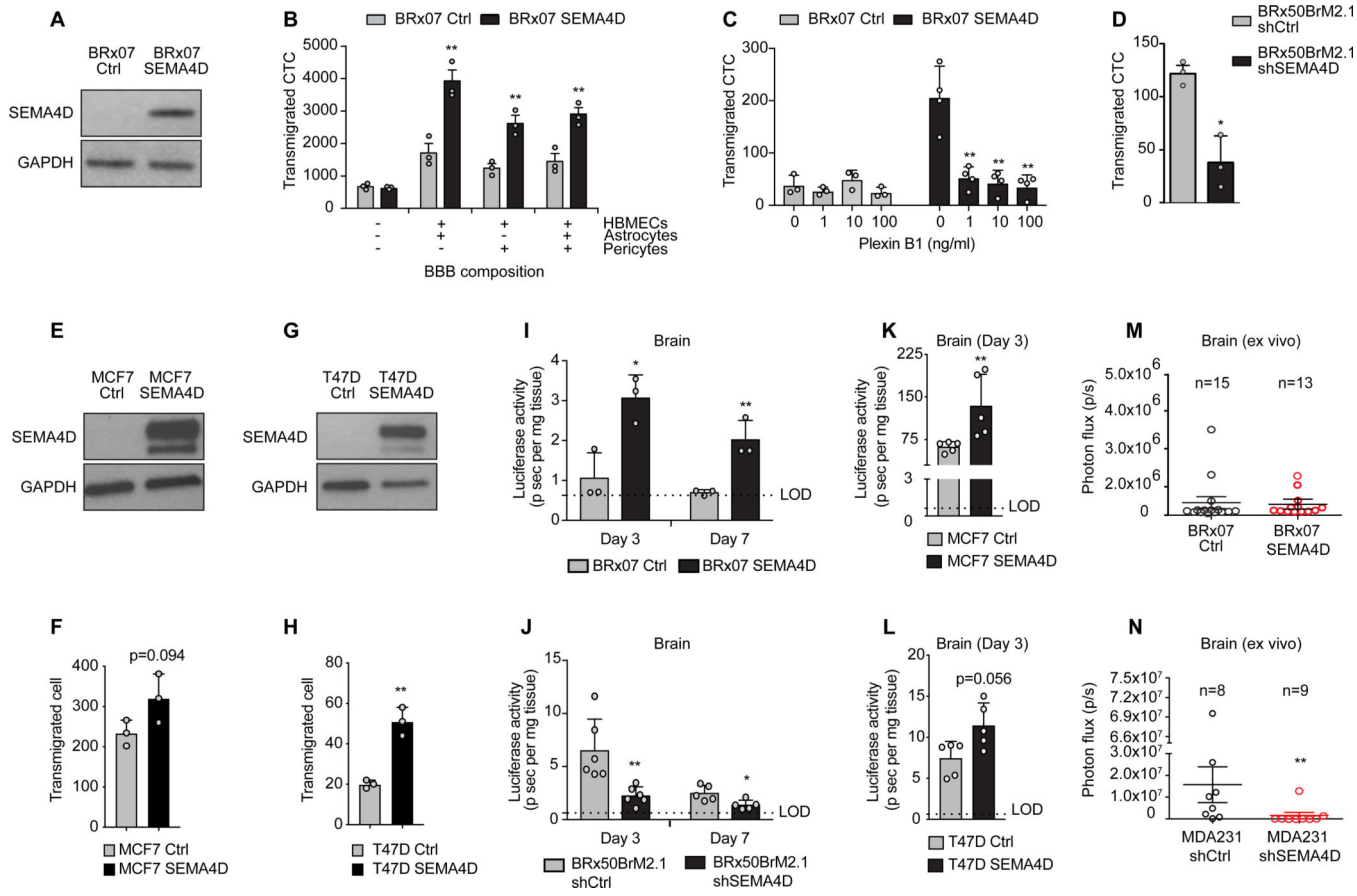


**Figure 4. SEMA4D mediates brain metastasis by promoting CTCs transmigration through the BBB.**

**A**, Immunoblot analysis with antibodies against SEMA4D and GAPDH in BRx07 cells control (BRx07 Ctrl) and BRx07 overexpressing SEMA4D cells (BRx07 SEMA4D). **B**, Bar graph showing quantification of transmigrated GFP-positive cells in *in vitro* BBB assay. *P* values were obtained with two-tailed unpaired *t*-test. **C**, Bar graph showing quantification of transmigrated GFP-positive cells pre-incubated with increasing concentration of Plexin B1 peptide. *P* values were obtained with two-tailed unpaired *t*-test. **D**, Bar graph showing number of transmigrated GFP-positive BRx50BrM2.1 control cells (BRx50BrM2.1 shCtrl) and GFP-positive BRx50BrM2.1 SEMA4D knockdown cells (BRx50BrM2.1 shSEMA4D). *P* values were obtained with two-tailed unpaired *t*-test. **E,** Immunoblot analysis with antibodies against SEMA4D and GAPDH in MCF7 cells control (Ctrl) and MCF7 overexpressing SEMA4D cells. **F,** Bar graph showing number of transmigrated GFP-positive cells. *P* values were obtained with two-tailed unpaired *t*-test. **G,** Immunoblot analysis with antibodies against SEMA4D and GAPDH in T47D cells control (Ctrl) and T47D overexpressing SEMA4D cells. **H**, Bar graph showing number of transmigrated GFP-positive cells. *P* values were obtained with two-tailed unpaired *t*-test. **I-L** *Ex vivo* luciferase signal from mice brain extracted at day 3 and 7 (I-J) or day 3 (K-L) after intracardiac injection of BRx07 (**I**), BRx50 (**J**), MCF7 (**K**) or T47D (**L**) cells with control or high/low expression of SEMA4D (mean ± s.e.m, two-tailed Wilcoxon Rank Sum test). LOD = Limit Of Detection. **M-N**, Quantification of *ex vivo* BLI intensity in brain after intracardiac

inoculation of (**M**) BRx07 or (**N**) MDA-MB-231 cells in mice at 3 months or 7 weeks respectively. Circles represent individual mouse (n= number of mice, mean ± s.e.m, two-tailed Wilcoxon Rank Sum Test). * *P* value < 0.05; ** *P* value < 0.01.

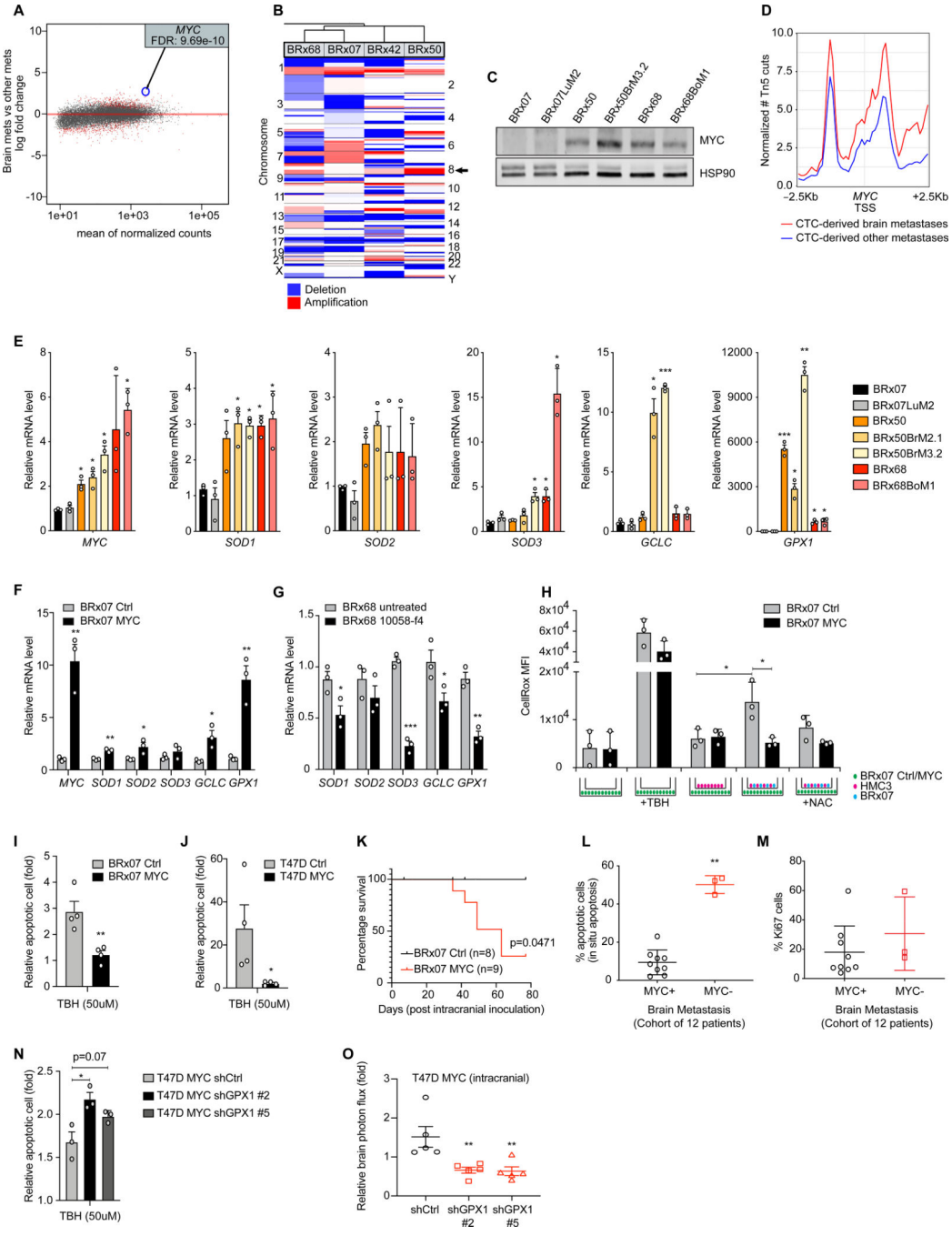


**Figure 5. MYC promotes brain metastasis by mitigating the oxidative stress elicited by activated microglia.**

**A**, MA plot depicting gene expression change in CTC-derived brain metastases compared to other CTC-derived metastases. Genes differentially expressed (FDR <0.01) in brain metastatic tumor cells are highlighted in red. **B**, Heatmap representing copy number alterations for a panel of CTC lines. Arrow shows chromosome 8. **C,** Immunoblot analysis of MYC expression level. **D**, Metaplot comparing chromatin accessibility around MYC TSS domain in CTC-derived brain metastases and other metastases. **E-G,** Histogram representing

*MYC*, *SOD1*, *SOD2*, *SOD3*, *GCLC* and *GPX1* mRNA relative expression levels in (**E**) CTC lines and CTC-derived metastatic tumor cells, or (**F**) BRx07 control (BRx07 Ctrl) and overexpressing MYC (BRx07 MYC) cells, or **(G)** BRx68 untreated and treated with 100 μM of MYC inhibitor (10058-f4) (two-tailed unpaired *t*-test). **H**, Control or BRx07-MYC cells were co-cultured (on the bottom of the well) with human microglia (HMC3) with or without presence of BRx07 cells (on the top of the chamber insert with 0.4 um pores on membrane to allow media exchange). Intracellular ROS levels were measured in BRx07-control and BRx07-MYC cells on the bottom using CellRox orange (CellRox) and median fluorescence intensity (MFI) was quantified by FACS. 50 μM tert-Butyl hydroperoxide (TBH) and 5 mM N-Acetyl Cysteine (NAC) were used as positive controls for inducing and reducing ROS levels, respectively (two-tailed unpaired t-test). Schematic of *in vitro* culture conditions is displayed at the bottom. **I-J**, Quantification of apoptotic cells (annexin V positive cells) after treatment with 50 μM of tert-Butyl hydroperoxide (TBH). BRx07 (**I**) and T47D (**J**) cells were treated for 24h and compared to untreated cell (fold change, two-tailed unpaired t-test). **K**, Kaplan-Meier curves showing brain tumor progression free survival analysis of mice injected with BRx07 control or MYC overexpressing cells, using 10 fold growth increase as threshold. n= number of mice (two-sided log-rank test). **L-M**, Quantification of (**L**) *in situ* apoptosis level (TUNEL) or (**M**) Ki67 positive cells by immunohistochemical staining of MYC-positive (MYC+) or MYC-negative (MYC-) brain metastases, with at least 3 independent fields per sample. Circles represent individual sample (mean ± s.e.m, two-tailed Wilcoxon Rank Sum Test). **N**, Quantification of apoptotic cells (annexin V positive cells) in T47D overexpressing MYC control cells (shCtrl) or GPX1 knockdown cells (shGPX1 #2 and #5) after treatment with 50 μM of tert-Butyl hydroperoxide (TBH). Cells were treated for 24h and compared to untreated cell (fold change, two-tailed unpaired t-test). shGPX1 #2 and #5 refer to two different hairpin sequences. **O**, Quantification of BLI intensity in brain at 3 weeks after intracranial inoculation. Circles represent individual mouse (mean ± s.e.m, n = number of mice, two-tailed Wilcoxon Rank Sum Test). In bar graphs, error bars represent s.e.m and circles represent each independent experiment. * $P$ value < 0.05; ** $P$ value < 0.01; *** $P$ value < 0.001.

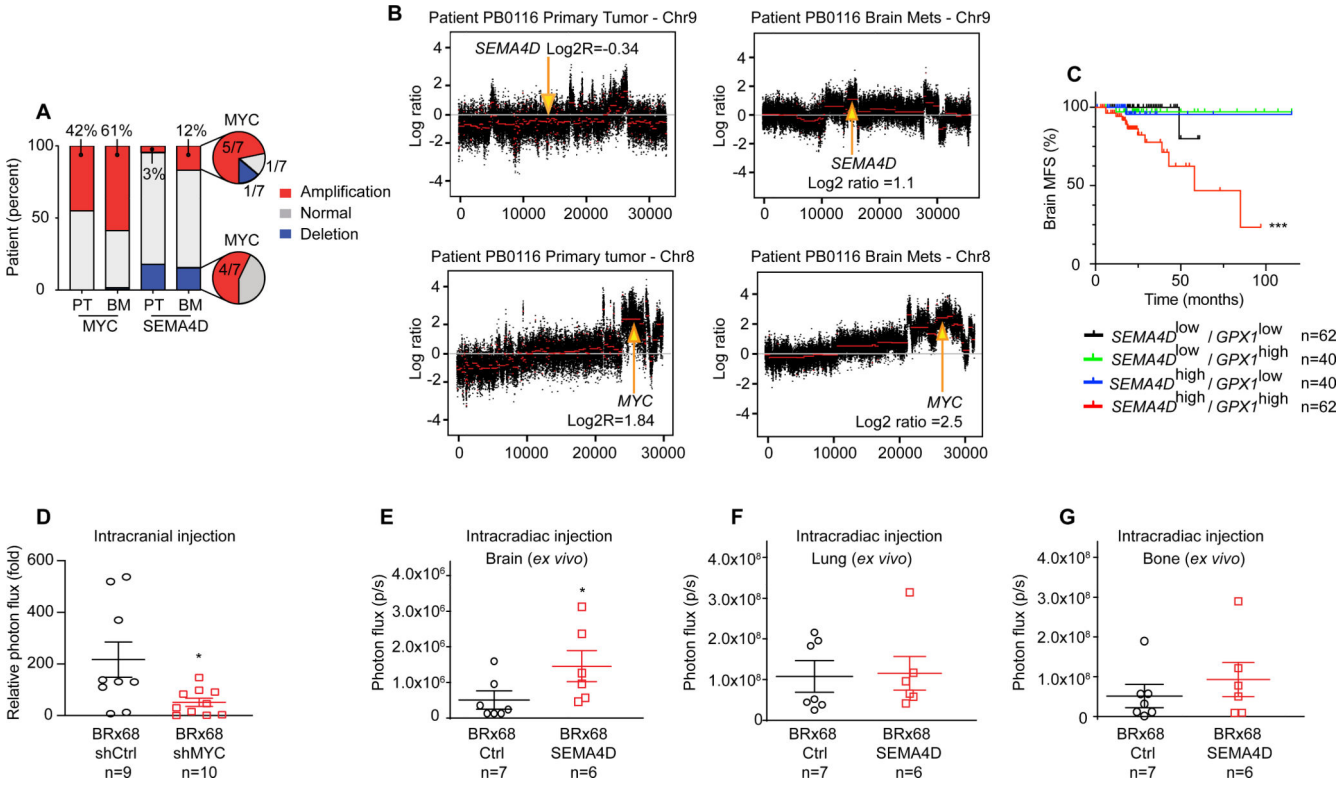


**Figure 6. Association of SEMA4D and MYC with the brain metastatic phenotype.**

**A**, Histogram depicting distribution of MYC and SEMA4D alterations across the cohort of patients (NCBI dbGAP phs000730). Copy number variation is compared between matched primary tumor (PT) and brain metastasis (BM). Percent or number of total cases with gene amplification are given (total n = 62). **B**, Segmented exome CNV analysis of chr8 and chr9 in matched primary breast tumor and brain metastasis of patient PB0116 (NCBI dbGAP phs000730). X-axis represents position in chromosome and Y-axis the log2 copy number. Red segmentation corresponds to the mean of reads across the chromosome. Arrows show *MYC* and *SEMA4D* gene locations. Log2 ratios are given for segments overlapping with *MYC* and *SEMA4D* genes. Data are normalized with matched normal tissue. **C**, Kaplan-Meier curves showing brain metastasis free survival analysis in 204 breast cancer patients (GSE12276) based on *SEMA4D* and *GPX1* expression in primary tumors. *P* values is given for $SEMA4D^{high}/GPX1^{high}$ group compared to each group and was determined with log-rank test. **D**, Quantification of BLI intensity in brain at 6 weeks after intracranial inoculation. Circles represent individual mouse, horizontal lines represent the mean ± s.e.m. n = number of mice. *P* values were obtained with two-tailed Wilcoxon Rank Sum Test. **E-G**, Quantification of *ex vivo* BLI intensity in (**E**), brain, or (**F**) lung, or (**G**), bone at 10 weeks after intracardiac inoculation of BRx68 cells in mice. Circles represent individual mouse (n= number of mice), horizontal lines represent the mean ± s.e.m. *P* values were obtained with two-tailed Wilcoxon Rank Sum Test. * *P* value < 0.05; ** *P* value < 0.01; *** P value < 0.001.

**Table 1.**
**Identification of genes associated with brain metastasis-free survival (MFS).**

The top 20 upregulated genes in brain tropic tumor cells and located on the chromosome 9q amplified region were selected based on their expression change adjusted P value. Genes were analyzed for association with MFS for brain, lung and bone using data from a public dataset (GSE12276). Brain MFS P values were determined with log-rank test.

| Gene symbol | Log2foldchange brain tropic vs non brain tropic | padj | Brain MFS pvalue | MFS description |
|---|---|---|---|---|
| SEMA4D | 1.80 | 2.39e-11 | 0.0023 | decreased brain MFS |
| FAM120AOS | 1.34 | 2.41e-13 | 0.0024 | increased brain MFS |
| FAM189A2 | 5.53 | 4.13e-16 | 0.006 | increased brain MFS |
| SLC44A1 | 1.41 | 1.16e-13 | 0.0101 | increased brain MFS |
| CTSL | 2.91 | 5.52e-11 | 0.019 | decreased brain and lung MFS |
| NOL8 | 1.74 | 6.71e-11 | 0.046 | increased brain MFS |
| UBQLN1 | 1.57 | 2.95e-23 | 0.0566 | increased brain MFS |
| PHF2 | 1.29 | 4.85e-10 | 0.078 | increased brain MFS |
| FXN | 2.97 | 2.44e-17 | 0.088 | |
| CDC14B | 3.31 | 2.61e-15 | 0.1217 | |
| NAA35 | 1.23 | 1.96e-11 | 0.2351 | |
| OSTF1 | 1.38 | 1.67e-10 | 0.3012 | |
| RNF20 | 1.92 | 4.92e-21 | 0.5576 | |
| MFSD14B | 1.41 | 6.79e-10 | 0.6678 | |
| TDRD7 | 2.68 | 1.80e-14 | 0.7787 | |
| BICD2 | 1.24 | 3.00e-10 | 0.786 | |
| SLC35D2 | 1.72 | 5.39e-11 | 0.8438 | |
| TMEM245 | 1.47 | 3.52e-10 | 0.8911 | |
| SPTLC1 | 1.89 | 2.75e-31 | 0.92 | |
| C9orf64 | 1.76 | 4.94e-14 | 0.9235 | |