

Multi-organ Site Metastatic Reactivation Mediated by Non-canonical Discoidin Domain Receptor 1 Signaling

Hua Gao^{1,2,8*}, Goutam Chakraborty^{2,8}, Zhanguo Zhang^{2,9}, Intissar Akalay², Mayur Gadiya^{2,3}, Yaquan Gao¹, Surajit Sinha², Jian Hu⁴, Cizhong Jiang⁴, Muzaffar Akram⁵, Edi Brogi⁵, Birgit Leitinger⁶, and Filippo G. Giancotti^{2,7*}

¹ Metastasis Research Institute, Breast Cancer Center, Key Laboratory of Arrhythmias of the Ministry of Education of China, Research Center for Translational Medicine, Shanghai East Hospital, and Shanghai Key laboratory of signaling and Disease Research, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China.

² Cell Biology Program and Center for Metastasis Research, Sloan Kettering Institute for Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.

³ Sloan Kettering Division, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065.

⁴ Shanghai Key Laboratory of Signaling and Disease Research, School of Life Sciences and Technology, Collaborative Innovation Center for Brain Science, Tongji University, Shanghai 200092, China.

⁵ Department of Pathology, Memorial Hospital, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.

⁶ Molecular Medicine Section, National Heart and Lung Institute, Imperial College London, London SW7 2AZ, United Kingdom.

⁷ Department of Cancer Biology and David H. Koch Center for Applied Research of Genitourinary Cancers, UT MD Anderson Cancer Center, Houston, TX 77054.

⁸ Co-first authors.

⁹ Present address: Hepatic Surgery Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430030, China.

*Correspondence: fggiancotti@mdanderson.org (F.G.G), gaoh@tongji.edu.cn (H.G.)

SUMMARY

Genetic screening identifies the atypical tetraspanin TM4SF1 as a strong mediator of metastatic reactivation of breast cancer. Intriguingly, TM4SF1 couples the collagen receptor tyrosine kinase DDR1 to the cortical adaptor syntenin 2 and, hence, to PKC α . The latter kinase phosphorylates and activates JAK2, leading to the activation of STAT3. This non-canonical mechanism of signaling induces the expression of *SOX2* and *NANOG*, sustains the manifestation of cancer stem cell traits, and drives metastatic reactivation in the lung, bone, and brain. Bioinformatic analyses and pathological studies corroborate the clinical relevance of these findings. We conclude that non-canonical DDR1 signaling enables breast cancer cells to exploit the ubiquitous interstitial matrix component collagen I to undergo metastatic reactivation in multiple target organs.

INTRODUCTION

Experiments in mouse models and analysis of clinical data suggest that disseminated breast cancer cells enter into a protracted state of proliferative dormancy before undergoing reactivation and causing metastatic relapse (Aguirre-Ghiso, 2007). In spite of their clinical relevance, tumor dormancy and reactivation remain poorly understood. Notably, it is unclear if they are mediated by core signaling pathways and transcriptional circuits that operate in multiple target organs.

The metastasis-initiating cells enter into dormancy and eventually undergo reactivation in response to niche signals that are part of the extracellular matrix, embedded in it, or exposed on the surface of adjacent stromal or epithelial cells (Giancotti, 2013). Dormancy-inducing signals include Bone Morphogenetic Protein 4 (BMP4), which is produced by lung-resident cells and suppresses the self-renewal capability of breast cancer cells that have infiltrated this organ. Tumor cells eventually avert BMP4 inhibition by producing the TGF- β ligand inhibitor Coco, which reinstates their proliferative capacity (Gao et al., 2012). Similarly, BMP7 and TGF- β 2 have been implicated in dormancy of prostate and squamous cell carcinoma, respectively (Bragado et al., 2013; Kobayashi et al., 2011). Stimulatory and inhibitory signals can be physically segregated along the abluminal surface of blood vessels, as mature endothelial cells produce thrombospondin, which induces dormancy, whereas angiogenic endothelial cells produce periostin, tenascin C, versican and TGF- β , which promote metastatic outgrowth (Ghajar et al., 2013). As most of the reactivation signals have been studied in models of breast cancer metastasis to lung, it remains unclear if they can function across organ sites.

We recently developed mouse genetic screens that enable the isolation of genetic entities involved in breast cancer reactivation (Gao et al., 2014). We here show that the top hit from one such screen, the tetraspanin TM4SF1, promotes reactivation in multiple organs and elucidate the underlying mechanism.

RESULTS

TM4SF1 Mediates Breast Cancer Reactivation in the Lung

Tail-vein injection of dormant 4TO7 cells transduced with cDNA libraries from metastatic 4T1 cells enabled the isolation of lung metastatic derivatives (Figure S1A) (Gao et al., 2012). cDNAs encoding the divergent tetraspanin TM4SF1 were recovered from 44 metastases that had arisen in 8 mice. 4TO7 cells re-transduced with TM4SF1 colonized the lung efficiently, corroborating the strong pro-metastatic activity of this hit (Gao et al., 2014). Conversely, silencing of TM4SF1 suppressed the capacity of 4T1, MDA-MB231 and ErbB2-transformed (Guo et al., 2006) breast cancer cells to colonize the lung, pointing to a requirement for TM4SF1 in this process (Figures 1A, S1B - S1E).

Consistent with a role for TM4SF1 in metastatic reactivation, its depletion did not reduce the number of viable tumor cells seeding the lung but suppressed their capacity to resume proliferation (Figures 1B, S1F, and S1G). To directly implicate TM4SF1 in this process, we used a doxycycline-regulated promoter to express it in 4TO7 cells, either immediately or 2 weeks after tail vein injection (Figures 1C and S1H). Whereas immediate expression of TM4SF1 induced metastatic outgrowth around day 7 post-injection, expression beginning at day 14 caused it around day 21 (Figures 1D, 1E, and S1I). Thus, a delay in induction of TM4SF1 causes a similar delay in metastatic outgrowth, directly implicating TM4SF1 in metastatic reactivation.

TM4SF1 is an evolutionarily divergent tetraspanin upregulated in lung, colon, breast, and ovarian carcinomas (Hellstrom et al., 1986; Marken et al., 1994; Marken et al., 1992). High levels of *TM4SF1* mRNA in primary tumors correlated with reduced relapse-free survival in ER- but not ER+ patients (Figure 1F; n=3,455), presumably because ER signaling suppresses *TM4SF1* expression in the latter (Al Saleh et al., 2011; Gao et al., 2014). Staining of clinically annotated Tissue Micro Arrays (TMAs; Table S1) indicated that high levels of expression of TM4SF1 in primary tumors correlate with reduced metastasis-free survival, suggesting that TM4SF1 promotes metastatic relapse in patients (Figures 1G and 1H). Finally, hypothesizing that TM4SF1 had a signaling function, we used its transcriptional program as a predictor of its involvement in metastasis. Consistently, we found that

an 8-gene signature induced by *TM4SF1* strongly predicts early relapse in unfractionated patients from a large dataset (Figures 1I, 1J, and S1J; n=855). Multivariate analysis showed that this discrete signature is an independent predictor of poor prognosis (Table S2). These results are consistent with a role for *TM4SF1* in metastatic reactivation of human breast cancer.

TM4SF1 Promotes Cancer Stem Cell Traits

Silencing of *TM4SF1* inhibited the capacity of 4T1 and ErbB2 cells to form tumor organoids in 3D Matrigel but did not inhibit tumor cell survival, proliferation, or migration under standard culture conditions (Figures 2A, 2B, and S2A-S2D). Moreover, inactivation of *TM4SF1* reduced tumor incidence and tumor growth and prolonged latency after injection of limiting numbers of ErbB2 cells in the mammary fat pad of NSG mice (Figures 2C and 2D left). Statistical analysis confirmed that silencing of *TM4SF1* reduces the frequency of tumor-initiating cells (Table S3). Similar results were obtained in secondary transplantation experiments, in spite of the increase in tumor initiating cells and hence the outgrowth of larger tumors (Figures 2C and 2D right, note different scales; Table S3). Finally, although silencing of *TM4SF1* did not affect the expression of the stem cell markers CD24, CD29, Sca1 and CD49f (Figure S2E), it almost completely suppressed the transcription of the pluripotency factors *Sox2*, *Oct4*, and *Nanog* (Figure 2E). These results indicate that *TM4SF1* promotes the manifestation of cancer stem cell traits

TM4SF1 Responds to Collagen I and Causes Metastatic Reactivation in Multiple Organs

Considering the role of specialized ECM niches in metastatic reactivation, we asked if *TM4SF1* mediates outgrowth in response to a specific matrix component. Intriguingly, treatment with soluble collagen I, but not collagen IV, fibronectin or laminin 1, greatly enhanced the capacity of ErbB2 cells to form primary and secondary tumor spheres (Figure 2F). Silencing of *TM4SF1* reversed the stimulatory effect of collagen I in both cases (Figures 2G and S2F). Similar results were obtained with the MDA-MB231 cells (Figure S2G). Collagen fibers, especially if

extensively crosslinked by lysil-oxidase (LOX), promote integrin-mediated activation of Rho Kinase (ROCK), facilitating FAK/Src signaling (Levental et al., 2009). However, neither inhibition of ROCK with Y-27632 nor of LOX with BAPN or a function blocking antibody suppressed tumor sphere formation in response to collagen I, suggesting that this process does not require integrin-mediated mechanotransduction (Figure 2H).

Since Matrigel contains minimal amounts of collagen I, we asked if it supported the outgrowth of tumor organoids by enabling the retention of tumor-produced collagen I. Consistently, we found that several breast cancer cells produce and deposit collagen I in the pericellular matrix to levels similar or higher as compared to those of Mouse Embryo Fibroblasts (Figure S2H). Moreover, we detected a clear rim of collagen I at the tumor-Matrigel interface in organoids grown in 3D Matrigel (Figure S2I). Finally, antibodies to human collagen I inhibited the ability of MDA-MB231 cells to form tumor spheres in the absence of exogenous collagen I, suggesting that the basal outgrowth of these spheres depends on autocrine collagen I signaling (Figure S2J). These observations suggest that TM4SF1 specifically enhances self-renewal in response to collagen I.

Given the ubiquitous presence of collagen I in the interstitial matrix, we examined if TM4SF1 mediates colonization not only of the lung but also of additional organs. Depletion of TM4SF1 suppressed the capacity of ErbB2 and 4T1 cells to produce bone and brain metastases and severely reduced the ability of the bone-tropic MDA-MB231 derivatives, BoM-1833, to colonize the bone (Figures 3A-3E and S3). Quantification of the solitary tumor cells and metastases present in bone and brain sections 3 weeks after injection indicated that silencing of TM4SF1 selectively blocks the ability of solitary tumor cells to outgrow into metastases (Figure 3F). These findings indicate that TM4SF1 is required for metastatic reactivation in multiple organs.

TM4SF1 Promotes Clustering of DDR1

TM4SF1 has diverged significantly from canonical tetraspanins during evolution, suggesting a distinctive function (Wright et al., 2000). Analysis of Pathway

Commons pointed to a potential interaction of TM4SF1 with the collagen receptor tyrosine kinase DDR1 and the cortical scaffold syntenin 2 (Figure 4A). Consistently, Myc-TM4SF1 combined efficiently with Flag-DDR1, but not Flag-DDR2, in 293T cells lysed in RIPA buffer and endogenous TM4SF1 associated to a lower extent with endogenous DDR1 in ErbB2 cells solubilized under the same conditions (Figures 4B and S4A-S4C). In contrast to canonical tetraspanins, recombinant TM4SF1 did not combine with endogenous α v or β 1 integrins, which include all known collagen-binding heterodimers, in MCF-7 cells (Figure S4D).

DDR1 and 2 are slowly activated by fibrillar and non-fibrillar collagens and exert partially redundant functions during development (Leitinger, 2011). Since *DDR1* is amplified in a fraction of breast, pancreatic and ovarian cancer cell lines (Marcotte et al., 2012) and DDR1 protein is heavily phosphorylated on tyrosine in squamous lung carcinoma (Rikova et al., 2007), we first focused on the DDR1-TM4SF1 interaction. Mutational analysis indicated that the extracellular portion of DDR1 is necessary for interaction with TM4SF1, whereas the transmembrane and cytoplasmic domains of the receptor are dispensable (Figure 4C). Considering the structure of DDR1 and TM4SF1 (Carafoli et al., 2012; Marken et al., 1994; Xu et al., 2014), we presume that the discoidin-like domain or the ~50 amino acid-long, highly flexible membrane-proximal segment of the ectodomain of DDR1 interacts with the extracellular loop connecting the 3rd and 4th transmembrane segments of TM4SF1.

Following engagement by fibrillar collagens, DDR1 dimers undergo multimerization leading to the formation of receptor clusters (Mihai et al., 2009; Noordeen et al., 2006). To study the role of TM4SF1 in this process, we used immunofluorescent staining because a large fraction of DDR1 and TM4SF1 becomes insoluble in multiple detergents - including RIPA buffer - in cells treated with collagen I, leading to an underestimate of the extent of association of DDR1 with TM4SF1 (Figures S4B, S4E, and S4F). The results showed that soluble collagen I induces DDR1 aggregation at the cell surface (Figures 4D top and 4E) and concomitant recruitment of TM4SF1 (Figure 4D bottom). Silencing of TM4SF1 suppressed collagen-induced clustering of DDR1, indicating that TM4SF1 is required for this process (Figures 4E and S4H). Similar results were obtained with a DDR1-

blocking monoclonal antibody (Figure 4F). Notably, this antibody, which binds to the discoidin-like domain of DDR1 and blocks its activation (Carafoli et al., 2012), also inhibited the association of DDR1 with TM4SF1 and increased the solubility of DDR1 (Figure S4G). In contrast, function-blocking antibodies to the $\beta 1$ integrin subunit, inhibition of ROCK with Y-27632, or inhibition of myosin II with Blebbistatin did not interfere with collagen-induced clustering of DDR1, excluding a participation of integrin-mediated signaling or myosin II contraction in this process (Figures 4F and S4I). Comparison of the kinetics of cluster assembly and disassembly revealed that the clusters induced by soluble collagen I persisted for 3 hours or more at the cell surface, but those induced by collagen IV, which forms networks, were short-lived (Figure S4J). Taken together, these findings suggest that collagen I enhances the association of DDR1 with TM4SF1, enabling the formation of large and persistent DDR1-TM4SF1 clusters at the cell surface, whereas collagen IV induces short-lived clusters.

DDR1's Prometastatic Function does not Require its Kinase Activity

In breast cancer cells exposed to collagen I, clustering of DDR1 declined before full activation of its kinase domain, suggesting that these two processes proceed independently (Figures 4G and S4J). Consistently, silencing of TM4SF1, and hence disruption of DDR1 clustering, did not inhibit DDR1 kinase activation (Figure 4G). Conversely, inhibition of DDR1 kinase with Ponatinib or Imatinib (Bantscheff et al., 2007; Canning et al., 2014) blocked DDR1 autophosphorylation but did not interfere with DDR1 clustering (Figures 4H and S4K). These results suggest that DDR1 clustering and kinase domain activation can proceed independently and have distinct consequences.

Depletion of DDR1 inhibits non-small cell lung cancer metastasis to the bone (Valencia et al., 2012). To examine the role of DDR1 in breast cancer metastasis, we silenced DDR1 or overexpressed a kinase-dead mutant. Strikingly, whereas depletion of DDR1, but not DDR2, inhibited the capacity of ErbB2 cells to form tumor spheres in vitro (Figures S4L-S4O), pharmacological or dominant-negative inhibition of DDR1 kinase did not interfere with this process (Figures S4P-S4S). Similarly,

silencing of DDR1, but not expression of kinase-dead DDR1, abrogated the capacity of ErbB2 cells to colonize the lung (Figures 4I and 4J). These results suggest that DDR1 promotes tumor sphere formation and lung colonization through a kinase-independent mechanism.

TM4SF1 Mediates Non-canonical DDR1 Signaling

The cytoplasmic C-terminal Cys residue of TM4SF1 binds to one of the two N-terminal PDZ domains of syntenin 2 (Borrell-Pages et al., 2000). While little is known about syntenin 2, its homologue syntenin 1 has been implicated in the activation of PKC, FAK and NF- κ B (Sarkar et al., 2008). Mutation of the C-terminal cysteine of TM4SF1 (C202G) disrupted its interaction with syntenin 2 (Figure S5A) and its capacity to promote lung colonization by 4TO7 cells (Figure 5A). Conversely, silencing of syntenin 2 suppressed lung colonization by the TM4SF1-proficient ErbB2 cells (Figure 5B). Ectopic expression experiments indicated that syntenin 2 combines with DDR1 only in the presence of TM4SF1 (Figures S5B and S5C). Moreover, silencing of TM4SF1 inhibited the interaction of endogenous DDR1 with syntenin 2 in ErbB2 cells (Figure S5D), suggesting that TM4SF1 functions as a membrane adaptor linking DDR1 to syntenin 2.

To gain insight into the signaling pathways activated by TM4SF1 via syntenin 2, we performed overexpression experiments. Wild type TM4SF1 and syntenin 2, but not TM4SF1-C202G, promoted activation of conventional and novel PKC isoforms and phosphorylation of STAT3 in 4TO7 cells; they did not, however, affect the activation of FAK, NF- κ B, and p38 kinase (Figures S5E and S5F). Subsequent experiments showed that soluble collagen I induces two waves of signaling in ErbB2 cells. Conventional PKC isoforms and JAK2 were activated in a biphasic fashion with peaks at 30 minutes and 6 hours following exposure to soluble collagen I. In contrast, DDR1 and STAT3 became phosphorylated at around 3 hours and persisted in an activated form up to at least 12 hours (Figure S5G). Silencing of TM4SF1 suppressed collagen-induced phosphorylation of PKC α , activation of total PKC, and phosphorylation and activation of JAK2 and STAT3, indicating that DDR1 activates PKC α and JAK2/STAT3 signaling through TM4SF1 (Figures 5C, 5D and

S5H). Inhibition of JAK2 with Fedratinib or AZD1480 did not affect collagen I-induced activation of total PKC, suggesting that JAK2 does not function upstream of PKC (Figure 5E). In contrast, inhibition of conventional and novel PKC isoforms with Gö6976 or GF109203X (Bis I) (Wu-Zhang and Newton, 2013) inhibited this process, confirming the involvement of these isoforms in the response to collagen I (Figure 5E). Notably, Gö6976 and Bis I blocked collagen-induced activation of JAK2 and STAT3, placing PKC upstream of JAK/STAT signaling. Consistently, pharmacological inhibition of JAK2 with Fedratinib or AZD1480 suppressed activation of STAT3 but not PKC (Figure 5F). Finally, inhibition of DDR1 kinase activity with Imatinib did not affect activation of PKC, JAK2, and STAT3 in response to collagen I (Figures 5E and 5F). These results suggest that TM4SF1 mediates non-canonical DDR1 signaling by promoting activation of PKC and hence of JAK2 and STAT3.

To confirm the involvement of PKC in activation of JAK2-STAT3 signaling, we examined the effect of PMA, which functions as an allosteric activator of conventional PKC. Treatment with PMA caused rapid and robust activation of JAK2 and subsequent activation of STAT3 in ErbB2 cells, demonstrating that conventional PKC can activate JAK2-STAT3 signaling (Figure 5G). In agreement with the hypothesis that TM4SF1 recruits PKC α through syntenin 2, PKC α combined efficiently with syntenin 2, but not TM4SF1, in 293T cells (Figures S5I and S5J). In contrast, JAK2 interacted with DDR1 and PKC α , but not with TM4SF1 or syntenin 2 (Figure S5K). These results suggest that DDR1, TM4SF1 and syntenin 2 cooperate to recruit and activate PKC α and JAK2.

We next considered the potential role of PKC α in the activation of JAK2. In vitro kinase assays indicated that recombinant PKC α promotes phosphorylation of JAK2 (Figures 5H and S5L). Mass spectrometric analysis revealed that PKC α phosphorylates JAK2 at two evolutionarily conserved residues: T174 in the N-terminal FERM domain and S518 in the segment that adjoins the SH2-like domain to the pseudo-kinase domain (Table S4, Figures S5M and S5N). Mutation of either residue to alanine inhibited the ability of JAK2 to phosphorylate GST-STAT3, suggesting that both sites regulate kinase activity (Figure 5I). Based on existing

structural data, we presume that phosphorylation of T174 and S518 contribute to disrupt negative regulatory interactions that the FERM domain and the SH2-like-pseudokinase domain of JAK2 establish with DDR1 or the pseudokinase-kinase domains of JAK2 (Brooks et al., 2014; Shan et al., 2014; Toms et al., 2013). These results suggest that PKC activates JAK2 and thereby STAT3 by directly phosphorylating T174 and S518. Collectively, these studies indicate that TM4SF1 promotes non-canonical DDR1 signaling by recruiting syntenin 2 and PKC, thus allowing transphosphorylation and activation of JAK2 (Figure 5J).

Non-canonical DDR1 Signaling Sustains Collagen-dependent Outgrowth and Metastasis

Inactivation of PKC, JAK2, or STAT3 suppressed collagen I-induced tumor sphere formation but not DDR1 clustering, implicating non-canonical DDR1 signaling in collagen I-induced self-renewal (Figures 6A, S6A, and S6B). In agreement with the identification of SOX2 as a STAT3 target gene (Chen et al., 2008), transient silencing of TM4SF1 or inhibition of PKC, JAK2, or STAT3 suppressed the expression of *Sox2* but not *Oct4* or *Nanog* (Figures 6B and 6C). Presumably, prolonged inhibition of non-canonical DDR1 signaling via stable depletion of TM4SF1 had downregulated all three pluripotency factors (Figure 3E) because they are connected by self-sustaining interactions (Young, 2011). Notably, transient depletion of *Sox2* or *Nanog* but not *Oct4* suppressed tumor sphere formation, suggesting that these two factors control breast cancer self-renewal in vitro (Figures 6D and 6E).

To dissect the role of the PKC α -STAT3 pathway in metastasis, we tested the effect of kinase-dead forms of PKC α , β II, δ , and ζ on lung colonization. Consistent with a specific involvement of PKC α (Tam et al., 2013), only expression of dominant negative PKC α blocked lung colonization (Figures 6F, 6G, and S6C). Silencing of STAT3 exerted a similar inhibitory effect (Figures 6H, 6I, and S6D-S6G). These findings implicate PKC α -STAT3 signaling in metastatic outgrowth.

To corroborate the involvement of non-canonical DDR1 signaling in cancer stem cell self-renewal and metastatic reactivation, we first verified that silencing of DDR1

or syntenin 2 suppresses the capacity of ErbB2 cells to form tumor organoids in 3D Matrigel (Figure S6H). We then examined lung sections from mice that had been inoculated with DDR1-, syntenin 2-, or STAT3-silenced ErbB2 cells and relative controls. The results indicated that silencing of DDR1, syntenin 2, or STAT3 selectively blocks the ability of solitary tumor cells to exit from dormancy and outgrow into metastatic lesions (Figure S6I). Finally, we investigated the effect of silencing of DDR1 or syntenin 2 on reactivation in other target organs. Depletion of each component suppressed the capacity of ErbB2 cells to colonize the bone and brain (Figures 6J, S6J, and S6K) by preventing solitary tumor cells from resuming proliferation in these organs (Figure S6L). Together, these results implicate several essential elements of non-canonical DDR1 signaling in reactivation from dormancy at multiple metastatic sites.

TM4SF1 Couples DDR1 to Activation of STAT3 during Tumor Progression

To study non-canonical DDR1 signaling during tumor progression and metastasis, we evaluated the deposition of collagen I in the mammary carcinomas and metastatic outgrowths of MMTV-*Neu*(YD) mice, which provide a faithful model of metastatic HER2+ breast cancer (Guo et al., 2006). The pre-neoplastic lesions and mammary intraepithelial neoplasia (MIN) lesions, which were identified by their expression of Neu and the presence of a large percentage of Ki-67-positive cells, were separated from the surrounding interstitial matrix by an intact basal lamina (Figure 7A). Although advanced MIN lesions had a partially disrupted basal lamina and were surrounded by collagen fibers, Masson trichrome and Picrosirius staining revealed large thick bundles of collagen surrounding tumor cell nests only in advanced adenocarcinomas (Figures 7A and S7A). These findings indicate that breast cancer cells come in contact with collagen I after they disrupt the underlying basement membrane and invade into the interstitium.

Staining of lung sections from 5 month-old mice with anti-Neu revealed several micrometastatic lesions, often associated with bronchioles or blood vessels. The micrometastases, and the larger metastases that they generated, did not deposit basement membrane components, such as collagen IV, but were surrounded by

collagen I fibrils (Figure S7B). These fibrils appeared orange under orthogonal polarized light after Picrosirius staining. However, they were not stained in blue by the Masson trichrome method, pointing to an intermediate level of macromolecular organization (Figure S7B). These results suggest that the tumor cells within incipient metastatic lesions or the stromal cells that they recruit can assemble a collagen I-rich metastatic niche. To capture an earlier stage of metastatic outgrowth, we inoculated mice intravenously with ErbB2 cells and subjected lung sections to staining with antibodies to Neu, collagen I, and Ki-67. We found that the majority of solitary tumor cells in contact with collagen I and those within micrometastatic lesions had resumed proliferation, whereas the majority of solitary tumor cells not in contact with collagen I were quiescent (Figures 7B and S7D). These findings are consonant with the hypothesis that contact with collagen I is required for or triggers metastatic reactivation.

Examination of the expression of DDR1 and TM4SF1 revealed that both are upregulated at the transition from advanced MIN lesion to invasive adenocarcinoma in MMTV-*Neu*(YD) mice (Figure S7C). However, whereas the primary tumors of these mice contained a small percentage of P-STAT3-positive cells, the lung metastases had a large number of tumor cells exhibiting strong nuclear staining for P-STAT3 (Figure S7C). This suggests that non-canonical DDR1 signaling to STAT3 is activated during tumor progression and initial metastatic outgrowth.

Finally, we studied the expression of DDR1, TM4SF1 and P-STAT3 in TMAs containing primary cancers and metastatic lesions. Whereas primary tumors and metastases exhibited similar levels of DDR1 and TM4SF1, P-STAT3 was significantly upregulated in metastases (Figure 7C). Analysis of matched samples confirmed this finding (Figure 7D). Consistent with the specific upregulation of STAT3 in metastases, the levels of P-STAT3 staining of primary tumors did not correlate positively with early relapse, but those of metastatic lesions retrospectively correlated with reduced time to progression (Figures 7E and S7E). These findings corroborate the importance of P-STAT3 signaling in human breast cancer metastasis.

DISCUSSION

We have identified a novel signaling mechanism that enables breast cancer cells to exploit the most abundant component of the extracellular matrix, fibrillar collagen, to undergo metastatic reactivation in multiple organs. The tetraspanin TM4SF1 functions in this process by coupling DDR1 to syntenin 2 and hence PKC α . The latter activates JAK2 and its transcriptional effector STAT3, thus impinging on pluripotency factors, which sustain the self-renewal capacity of metastasis-initiating cells. Experiments involving inducible expression of TM4SF1 in dormant cells or inactivation of TM4SF1 and other key components in metastatic cells corroborate the role of non-canonical DDR1 signaling in metastatic reactivation.

Extensive evidence suggests that highly cross-linked collagen fibers promote tumor progression and metastasis by activating an integrin mechanotransduction pathway involving Rho-mediated contraction of the actin cytoskeleton and enhanced FAK/Src signaling (Erler et al., 2006; Levental et al., 2009; Paszek et al., 2005; Samuel et al., 2011). In contraposition, our results indicate that soluble fibrillar collagen I, and presumably other fibrillar collagens, enhance the manifestation of cancer stem cell traits, such as the expression of pluripotency transcription factors, the formation of tumor spheres, and the ability to initiate tumor outgrowth, by activating non-canonical DDR1 signaling independently of integrin engagement. Additional genetic experiments implicate this novel signaling pathway in metastatic reactivation, thus reinforcing the connection between the most basic cellular attribute of cancer stem cells - the capacity for self-renewal - and metastatic reactivation. We speculate that DDR1 signaling and integrin mechanotransduction may not only proceed independently, as suggested previously (Vogel et al., 2000) and in this paper, but also have opposing functions in certain contexts. For example, inhibition of integrin-mediated mechanotransduction with the Rho-kinase inhibitor Y27632 promotes the outgrowth and propagation of normal and tumor organoids in 3D Matrigel (Gao and Chen, 2015). Conversely, DDR1 inhibits integrin $\alpha 2 \beta 1$ signaling in MDCK cells migrating in collagen gels (Wang et al., 2006). These considerations suggest that integrin mechanotransduction and non-canonical DDR1 signaling regulate distinct steps of the invasion-metastasis cascade: the former predominantly

mediating survival and invasive growth through dense extracellular matrices in advanced primary tumors, and the latter enabling cancer stem cell self-renewal and metastatic reactivation in the interstitial matrix of target organs.

Non-canonical DDR1 signaling possesses several distinguishing features, including its dependence on the nucleation of large membrane signalosomes and its complex kinetics (see Supplementary Discussion). Although we cannot exclude the involvement of additional transcriptional effectors, our results have revealed that engagement of DDR1 by collagen I controls self-renewal and metastatic reactivation through activation of STAT3, which prior studies had considered to be only responsive to Stroma-derived IL6 (Yu et al., 2014). Consistent with the identification of SOX2 as a direct target of STAT3 (Chen et al., 2008; Zhao et al., 2014), acute inactivation of TM4SF1 or STAT3 suppressed the expression of SOX2, but not *NANOG* or *OCT4*, and tumor sphere formation, suggesting that non-canonical DDR1 signaling to SOX2 is required for self-renewal in vitro. Acute depletion of *NANOG*, but not *OCT4*, also inhibited tumor sphere formation, implicating both SOX2 and *NANOG* in this process. Since stable depletion of TM4SF1 led to downregulation of all three pluripotency factors presumably because of their participation in a self-sustaining transcriptional network (Young, 2011), we hypothesize that non-canonical DDR1 signaling controls breast cancer stem cell behavior by inducing SOX2 directly and *NANOG* indirectly.

In support of the clinical relevance of non-canonical DDR1 signaling, a discrete 8-gene signature controlled by TM4SF1 exhibited a strong ability to predict metastatic relapse across breast cancer subtypes. In addition, immunohistochemical analysis of matched primary and metastatic samples revealed that the cancer cells within metastases contain significantly higher levels of nuclear P-STAT3 as compared to those in primary tumors. Amongst the metastases, the highest levels of P-STAT3 were present in those that had occurred after a short disease-free interval. Since neither TM4SF1 nor DDR1 levels increased in the metastases of MMTV-*Neu* mice and breast cancer patients, it is possible that additional pathway components are upregulated during tumor progression leading to positive selection of tumor cells in which non-canonical DDR1 signaling can be fully activated upon contact with

collagen I in the stroma of target organ. We presume that this selection occurs within primary cancers because expression of the TM4SF1 signature in primary tumors predicts metastatic relapse.

It has been proposed that inhibition of niche signals may significantly enhance the efficacy of chemotherapy and oncogene-targeted therapy in metastatic disease (Ghajar, 2015; Giancotti, 2013). Non-canonical DDR1 signaling emerges from our study as a general niche signal and a prime target for this novel therapeutic approach. Based on our findings, we propose that antibodies and other biological disrupting the association of DDR1 with TM4SF1 or novel and safe JAK2 kinase inhibitors may display efficacy in the adjuvant setting or, in combination with chemotherapy or oncogene-targeted therapy, in the treatment of metastatic breast cancer.

EXPERIMENTAL PROCEDURES

Animal Studies

All mouse studies were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee of MSKCC. Details are provided in Supplemental Experimental procedures.

Doxycycline-inducible Metastatic Reactivation

4TO7-TGL cells were infected with pCW-Myc-TM4SF1 and inoculated intravenously into syngeneic mice. Doxycycline was administered to mice through the diet (625 mg/kg of food) as well as by intraperitoneal injection (25 mg/kg of body weight; 3 times a week). Doxycycline was administered either immediately after injection of the cancer cells or 14 days later. Bioluminescent imaging was used to verify successful injection and to monitor metastatic outgrowth. Metastatic lesions were confirmed by histological analysis.

Tumor Sphere Assays

Mammosphere assays were performed as previously described (Dontu et al., 2003). Single cells suspension of ErbB2 and MDA-MB231 cells (1000 cells/ml) were plated on ultra low attachment plates and cultured in serum-free MEBM supplemented with

5 µg/ml insulin, 0.5 µg/ml hydrocortisone, 1:50 B27, 20 ng/ml bFGF and EGF, and 4 µg/ml insulin and heparin for 10 days. When indicated, purified collagen I was added at 50 µg/ml. Tumor spheres were visualized under phase contrast microscope, photographed and counted. For serial passage, tumor spheres were collected using 70 µm cell strainers and dissociated with Accutase for 30 min at 37⁰ C to obtain single-cell suspensions.

DDR1 Clustering

293FT transfected with HA-DDR1 and Myc-TM4SF1 or control and TM4SF1-silenced ErbB2 cells were deprived of growth factors for 12 hours and then treated with soluble collagen I or IV for the indicated times followed by fixation with 4% PFA for 10 minutes at 4⁰ C. When indicated, inhibitory mouse monoclonal DDR1 antibodies were added at the indicated concentrations prior to exposure to collagen I. Fixed cells were washed with PBS, permeabilized with 0.2% Triton X-100 for 20 minutes, stained with a rabbit monoclonal antibody cross reactive with both mouse and human DDR1 (clone D1G6), alone or in combination with phalloidin, FM 4-64 membrane dye (Molecular Probes), mouse monoclonal anti-HA antibody (Clone16B12), or rabbit anti-TM4SF1 antibody and counterstained with DAPI. The data were quantified by ImageJ 1.48b and represented graphically.

Human Pathology

Paraffin-embedded tissue microarray sections with multiple cores of primary breast tumors or lung metastatic lesions were obtained from the MSKCC Department of Pathology in compliance with protocols approved by the MSKCC Institutional Review Board and after the subjects gave their informed consent. The levels of expression of TM4SF1 and DDR1 and of activation of STAT3 were determined by immunohistochemical staining. TM4SF1, DDR1, or P-STAT3 immunoreactivity was evaluated and scored by a clinical pathologist (E.B.). Metastasis-free survival was analysed as a function of TM4SF1 or P-STAT3 abundance in breast tumor samples or lung metastatic samples.

DNA Microarray Analysis

Standard methodology was used to extract, label and hybridize the cDNAs to the MouseRef8 gene expression chip (Illumina). Data were analysed in Partek.

Statistical Analysis

Results are reported as mean $\pm$ SD or $\pm$ SEM unless otherwise noted. Comparisons between two groups were performed using an unpaired two-sided Student's *t* test ($p < 0.05$ was considered significant).

Accession Numbers

Gene expression data of 4TO7 and 4TO7-TM4SF1 are deposited at Gene Expression Omnibus with accession number GSE62730.

AUTHOR CONTRIBUTIONS

Study conception: H.G. and F.G.G.; Methodology: H.G., G.C., and F.G.G.; Computation and formal analysis: J.H. and C.J.; Investigation: H.G., G.C., Z.Z., I.A., M.G., Y.G., S.S., and M.A., where I.A., M.G., and Y.G. contributed equally to the investigation and are therefore listed alphabetically; Resources, E.B. and B.L.; Writing – Original draft, review, and editing: H.G., G.C., and F.G.G.; Writing – critical review, commentary or revision: B.L. and F.G.G.; Writing – data presentation: H.G. and F.G.G.; Supervision: H.G. and F.G.G.; Funding acquisition: H.G. and F.G.G.

ACKNOWLEDGEMENTS

We thank R. Levine, J. Massagué, C. McCilloch, F. Miller, and M.J. Tang for reagents, A. Barlas, N. Fan, and M. Turkekul for help with some of the staining experiments, H. Erdjument-Bromage for mass spectrometric analyses, the staff of the Genomics Core and the Molecular Cytology Core for their services, and members of the Giancotti laboratory for discussions. This work was supported primarily by grants from the NIH (R01 CA175712, R35 CA197566, and P01 CA094060 Project 4, to F.G.G.). Additional funding was provided by the Geoffrey Beene Cancer Center of MSKCC (to F.G.G.), the National Basic Research program of China (973 Program, 2015CB964800, to H.G.), the National Natural Science Foundation of China (81372840, to H.G.) and the One Thousand Talents Program of

Shanghai (SH04020, to H.G.) and the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (81221001). H.G. received an Eastern Scholarship from the Program for Professors of Special Appointment at Shanghai Institutions of Higher Learning (2013-14) and Z.Z. was partially supported by the China Scholarship Council.

FIGURE LEGENDS

Figure 1. TM4SF1 Induces Metastatic Reactivation in the Lung and Predicts Breast Cancer Relapse

(A and B) 4T1 cells expressing the bioluminescence reporter TGL (4T1-TGL) were infected with lentiviruses encoding two shRNAs targeting TM4SF1 (sh #1 and #5) or a control shRNA (sh-Co.), analyzed by Q-PCR and immunoblotting (IB) (A, top left), and inoculated intravenously (i.v.) in syngeneic mice. Lung metastasis was measured by bioluminescent imaging (BLI). Normalized photon flux at the indicated times (A, bottom left), representative images (A, right). Error bars, mean $\pm$ SD. *P* values (last time point); Student's *t* test. Lung sections of indicated times were stained with antibodies to GFP (tumor cells) and PECAM-1 (endothelial cells) (B, top). White arrows: solitary tumor cells. Number of GFP-positive solitary tumor cells/high power field (hpf) (blue bars) and metastatic lesions/lung section (red bars) at the indicated times (B, bottom).

(C-E) Schematic representation of inducible expression experiment (C). 4TO7 cells expressing DOX-inducible TM4SF1 were inoculated i.v. in syngeneic mice. DOX was administrated either immediately after injection (Dox day 0) or 14 days after inoculation of the cells (Dox day 14). Lung metastasis was measured by BLI. Normalized photon flux at the indicated time; error bars, mean $\pm$ SE. *P* values; Student's *t* test (D). Representative images (E).

(F) Kaplan-Meier analysis of relapse-free survival of ER+ (left) and ER- patients (right) in publicly available breast cancer datasets (source KM Plotter for breast cancer). Patients were divided according to TM4SF1 expression as indicated. HR: Hazard Ratio.

(G and H) TMA comprising 147 primary breast tumors of MSKCC's patients were subjected to immunohistochemistry with anti-TM4SF1 and counterstaining with Hematoxylin (H). Representative images of cases exhibiting varying levels of TM4SF1 (G). Distribution of cumulative staining intensities across all samples (H, left). Patients were divided according to the intensity of TM4SF1 staining as indicated by the red arrow (left) and metastasis-free survival data were subjected to Kaplan-Meier analysis (H, right).

(I) Hierarchical clustering of genes concordantly up or downregulated (≥ 1.5 fold) in triplicate samples of TM4SF1-overexpressing 4TO7 cells as compared to control cells.

(J) Kaplan-Meier analysis of relapse-free survival in the MSK82, EMC192, EMC286, and NKI295 combined dataset. Patients were divided according to the expression of the 8-gene TM4SF1 signature. Individual genes comprising the signature are listed to the right of the graph.

See also Figure S1.

Figure 2. TM4SF1 Enhances Cancer Stem Cell Traits in Response to Collagen I

(A and B) Control and TM4SF1-silenced mammary tumor cells from MMTV-*Neu*(YD) mice (ErbB2 cells) were cultured in 3D Matrigel. Representative wells (A) and mean number of organoids ($\pm$ SE) per 100 cells from triplicate samples (B). *P* values; Student's *t* test.

(C and D) Control and TM4SF1-silenced ErbB2-TGL cells were admixed with Matrigel and inoculated in the mammary fat pad of NSG mice at the indicated numbers. Mean of primary and secondary tumor volumes at 5 weeks (C; left y axis). Incidence of tumors at 5 weeks (C; right y axis) and tumor latency at indicated cell number (D). *P* values; Student's *t* test. Data point, mean $\pm$ SD

(E) Control and TM4SF1-silenced ErbB2 cells were subjected to Q-PCR analysis. Error bars, mean $\pm$ SD.

(F) ErbB2 cells were subjected to tumor sphere assay with or without collagen (Coll) I (50 μ g/ml), Coll IV (50 μ g/ml), laminin 1 (Lam 1; 10 μ g/ml), or fibronectin (Fn; 50 μ g/ml) for 10 days. Number of tumor spheres per 10^3 cells ($\pm$ SE); *P* values; Student's *t* test.

(G) Indicated ErbB2 cells were subjected to tumor sphere assay in medium supplemented with or without 50 $\mu\text{g/ml}$ collagen I. Number of primary (left) and secondary (right) tumor spheres per 10^3 cells seeded ($\pm\text{SE}$). *P* values; Student's *t* test.

(H) ErbB2 cells were subjected to tumor sphere assay in medium supplemented with or without 50 $\mu\text{g/ml}$ coll I in combination with vehicle (0.25% DMSO), Y-27632 (ROCK1 inhibitor, 10 μM), β -Alaninenitrile (BAPN, 200 μM), or function-blocking anti-Lox1 antibody (5 $\mu\text{g/ml}$). Number of tumor spheres formed per 10^3 cells seeded ($\pm\text{SE}$). *P* values; Student's *t* test.

See also Figure S2.

Figure 3. TM4SF1 Promotes Metastasis to the Bones and Brain

(A) ErbB2-TGL cells were inoculated (3×10^5) intracardially (i.c.) in nude mice. Bone and brain metastases were measured by BLI. Representative images are shown.

(B) Representative X-ray radiography of hind legs. Yellow arrows point to osteolytic lesions.

(C and D) Hematoxylin and Eosin (H&E) and anti-GFP IHC of adjacent sections from bone (C) and brain metastases (D).

(E) The graphs show the cumulative photon flux from both hind legs (top) or the brain (bottom) at day 21 after i.c. injection ($\pm \text{SE}$). *P* values, unpaired Student's *t* test.

(F) Bone (hind limbs) and brain sections of mice were processed for IHC detection of GFP (tumor cells). The graphs show the mean numbers of GFP-positive solitary tumor cells per microscopic field (blue bars) and of metastatic lesions per section (red bars) at end point of the experiment ($\pm\text{SE}$). *P* values; Student's *t* test.

See also Figure S3.

Figure 4. DDR1 Combines with TM4SF1 to Promote Tyrosine Kinase-independent Signaling and Metastasis

(A) Pathway and interaction data are from HPRD, Reactome, NCI-Nature Pathway Interaction Database, and the MSKCC Cancer Cell Map, as derived from Pathway

Commons. Black lines: direct and specific interactions; white spheres: interactions validated by co-immunoprecipitation (IP).

(B) 293FT cells were co-transfected with Myc-TM4SF1, together with Flag-DDR1, or empty vector and IP with anti-Flag. Total lysates (left) and Flag IP (right) were subjected to IB with the indicated antibodies. Asterisks: glycosylated TM4SF1; arrow: TM4SF1 core protein.

(C) 293FT cells were co-transfected with Myc-TM4SF1 together with Flag-DDR1, Flag-TrkA-DDR1 (extracellular domain chimera), Flag-DDR1 609* (intracellular domain truncation), Flag-DDR1 529* (intracellular domain truncation), or empty vector and were IP with anti-Flag. Flag IP and total lysates were subjected to IB with the indicated antibodies.

(D) ErbB2 cells were treated with or without 10 μ g/ml coll I for 30 min, stained with anti-DDR1, and counterstained with the plasma membrane dye FM-4 64 (PM) and DAPI. Representative images (top panels). Myc-TM4SF1 and HA-DDR1 co-transfected 293FT cells were treated with or without 10 μ g/ml collagen I for 30 min and stained with anti-HA and anti-TM4SF1 followed by counterstaining with DAPI. Representative images of co-transfected cells (bottom panels).

(E) ErbB2 cells were treated with or without 10 μ g/ml collagen I for 30 min and stained with anti-DDR1 antibody and DAPI. Representative images (top) and surface fluorescence intensity plots (bottom). The insets show high magnification images of a portion of the plasma membrane adjacent to the boundary of the nucleus demarcated by white dotted lines (bottom).

(F) Control and TM4SF1-silenced MDA-MB231 cells were treated with or without 10 μ g/ml collagen I for 30 min. When indicated, cells were pretreated with 10 μ g/ml anti-DDR1 or anti-ITGB1 (integrin β 1) function-blocking monoclonal antibodies. Number of DDR1 clusters per microscopic field. Mean values ($\pm$ SD); *P* values; Student's *t* test.

(G) ErbB2 cells transfected with control or TM4SF1-specific siRNA SMARTpools were treated with 10 μ g/ml coll I for the indicated times, and IP with anti-DDR1. Immunoprecipitates were subjected to IB with the indicated antibodies

(H) ErbB2 cells were treated with or without 10 $\mu\text{g/ml}$ collagen I for 30 min, alone or in combination with 100 nM Ponatinib or Imatinib, and then subjected to staining with anti-DDR1 followed by DAPI. Number of DDR1 clusters per microscopic field ($\pm$ SD). *P* values; Student's *t* test.

(I and J) Control (sh-co) or DDR1-silenced (sh #6 and #7) (I) and control (vector) or DDR1-P529*-overexpressing (J) ErbB2-TGL cells were inoculated i.v. in nude mice. Lung metastasis was measured by BLI. Normalized photon flux at the indicated times (left ($\pm$ SD) and representative images (right); *P* values; Student's *t* test.

See also Figure S4.

Figure 5. Non-canonical DDR1 Signaling Promotes PKC α -dependent Activation of JAK2 and STAT3

(A) 4TO7-TGL cells were transduced with wild type or mutant (C202G) TM4SF1 or with empty vector, subjected to IB as indicated (left top), and inoculated i.v. in syngeneic mice. Lung metastasis was measured by BLI. Normalized photon flux at the indicated times (left bottom; $\pm$ SD) and representative images (right). *P* values; Student's *t* test.

(B) Control (sh-co) or syntenin 2-silenced (sh #1 and #3) ErbB2-TGL cells were subjected to IB as indicated (left top), and inoculated i.v. in nude mice. Lung metastasis was measured by BLI. Normalized photon flux at the indicated times (left bottom) and representative images (right). Error bars, mean $\pm$ SD. *P* values Student's *t* test.

(C) Control or TM4SF1 specific siRNA (SMART pool) transfected ErbB2 cells were treated with 50 $\mu\text{g/ml}$ collagen I for the indicated times. Total lysates were subjected to IB with the indicated antibodies.

(D-F) ErbB2 cells transfected with control or TM4SF1-specific siRNA SMARTpools were treated with coll I, alone or in combination with Gö6976 (50 nM), Bisindolylmaleimide I (Bis I, 300 nM), Fedratinib (1 μM), AZD1480 (100 nM), or Imatinib (100 nM), lysed, and subjected to a PKC kinase assay (D, E) or immunoblotting (F). Mean values ($\pm$ SD); *P* values Student's *t* test (D, E).

(G) ErbB2 cells were treated with 100 nM PMA for the indicated times and subjected to IB.

(H) Purified recombinant PKC α and JAK2 were admixed in vitro, alone or in combination with PMA (PKC activator, 50 nM) or Gö6976 (PKC inhibitor, 100 nM), as indicated, and subjected to in vitro kinase assay with radioactive ATP.

(I) JAK2-deficient γ 2A cells were transfected with wild type (WT) or mutant forms of Flag-JAK2. Total lysates were subjected to in vitro kinase assay with cold ATP and GST-STAT3 as substrates. Samples were subjected to IB as indicated. Note these cells possess elevated levels of endogenous phosphorylated PKC. Wild type JAK2, but none of phosphorylation site mutants, induces phosphorylation of GST-STAT3 at Y705. An endogenous activity induces phosphorylation of STAT3 at S727, serving as an internal control.

(J) Model of non-canonical DDR1 signaling. Kinase-dependent pathways are shaded grey.

See also Figure S5.

Figure 6. Non-Canonical DDR1 Signaling Promotes Self-renewal in Vitro and Lung Metastasis in Vivo

(A) ErbB2 cells were treated with 50 μ g/ml coll I alone or combination with Gö6976 (50 nM), Bis I (300 nM), Fedratinib (1 μ M), Stattic (2.5 μ M), BP-1-102 (20 μ M), and Ponatinib (100 nM) or they were transduced with DDR1 sh-RNAs (#6 and #7) before treatment with coll I. The graph shows the number of tumor spheres formed per 10^3 cells seeded ($\pm$ SE). *P* values, unpaired Student's *t* test.

(B) ErbB2 cells transfected with control or TM4SF1-specific siRNA SMARTpools were subjected to Q-PCR analysis ($\pm$ SE).

(C) ErbB2 cells were treated with DMSO or Gö6976 (50 nM) (top), Fedratinib (1 μ M) (middle), and Stattic (2.5 μ M) (bottom), and subjected to Q-PCR analysis ($\pm$ SE).

(D and E) ErbB2 cells were transfected with control or Sox2- (si-Sox2), Nanog- (si-Nanog), or Oct4-targeting (si-Oct4) siRNA SMARTpools and subjected to tumor sphere assay with 50 μ g/ml coll I. Number of tumor spheres formed per 10^3 cells

seeded ($\pm$ SE) (D); representative images of tumor spheres (E). *P* values, unpaired Student's *t* test.

(F and G) 4T1-TGL cells were transduced with dominant negative (DN) PKC α , PKC β II, PKC δ , or PKC ζ , or empty vector and inoculated i.v. in syngeneic mice. Lung metastasis was measured by BLI. Normalized photon flux at the indicated times for DN PKC α or PKC β II, PKC δ , or PKC ζ relative to empty vector ($\pm$ SD). *P* values; Student's *t* test (F). Representative images at week 5 (G).

(H and I) Control and STAT3-silenced (#1 and #5) 4T1-TGL cells were inoculated intravenously into syngeneic mice. Lung metastasis was measured by bioluminescent imaging. Normalized photon flux at the indicated times ($\pm$ SD). *P* values; Student's *t* test (H). Representative images at week 5 (I).

(J) Control and DDR1- (#6 and 7) or syntenin 2-silenced (#1 and 3) ErbB2-TGL cells were inoculated i.c. in immunodeficient mice. Bone and brain metastases were detected by bioluminescent imaging. The panels show representative images.

See also Figure S6.

Figure 7. Fibrillar Collagen I and Activated STAT3 Define Metastatic Niches

(A) Representative images of preneoplastic and MIN lesions (top panels) and adenocarcinomas (bottom panels) from 12 and 20 week-old MMTV-*Neu*(YD) mice subjected to IHC with the indicated antibodies followed by Hematoxylin or to Masson trichrome staining. Red arrow: normal duct comprising quiescent cells; purple arrow: pre-neoplastic lesion comprising hyperproliferative cells; green arrow: MIN lesion that has undergone luminal filling (inset).

(B) Lung sections from mice that had been injected intravenously with ErbB2-TGL cells 14 days earlier were subjected to immunofluorescent staining with antibodies to Neu, collagen I, and Ki-67 followed by counterstaining with DAPI. Solitary tumor cells (Neu+) were categorized as closely apposed to coll I (+) or not (-) and as Ki-67+ or Ki-67-. Percentage of Ki-67+ cells among solitary tumor cells in contact or not with coll I ($\pm$ SE). *P* values unpaired Student's *t* test (left). Representative images of single cells from the predominant categories (right).

(C-E) TMAs comprising primary breast tumors and paired lung metastatic lesions from MSKCC patients were subjected to IHC with indicated antibodies. Level of

DDR1 (left), TM4SF1 (middle), and P-STAT3 (right) staining in primary breast tumors and lung metastatic lesions (C). Comparison of P-STAT3 staining in 14 matched primary tumors and metastases (D, left); representative images from one case (D, right). Data are mean ($\pm$ SE and $\pm$ SD), *P* values unpaired student's *t* test. Patients (n=61) were divided based on the level of P-STAT3 staining in their lung metastases and their metastasis-free survival examined by Kaplan-Meier analysis (E). P: primary breast tumors; M: lung metastatic lesion.

See also Figure S7.

REFERENCES

- Aguirre-Ghiso, J.A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. *Nat. Rev. Cancer* 7, 834-846.
- Al Saleh, S., Al Mulla, F., and Luqmani, Y.A. (2011). Estrogen receptor silencing induces epithelial to mesenchymal transition in human breast cancer cells. *PloS One* 6, e20610.
- Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., *et al.* (2007). Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotech.* 25, 1035-1044.
- Borrell-Pages, M., Fernandez-Larrea, J., Borroto, A., Rojo, F., Baselga, J., and Arribas, J. (2000). The carboxy-terminal cysteine of the tetraspanin L6 antigen is required for its interaction with SITAC, a novel PDZ protein. *Mol. Biol. Cell* 11, 4217-4225.
- Bragado, P., Estrada, Y., Parikh, F., Krause, S., Capobianco, C., Farina, H.G., Schewe, D.M., and Aguirre-Ghiso, J.A. (2013). TGF-beta2 dictates disseminated tumour cell fate in target organs through TGF-beta-RIII and p38alpha/beta signalling. *Nat. Cell Biol.* 15, 1351-1361.
- Brooks, A.J., Dai, W., O'Mara, M.L., Abankwa, D., Chhabra, Y., Pelekanos, R.A., Gardon, O., Tunny, K.A., Blucher, K.M., Morton, C.J., *et al.* (2014). Mechanism of activation of protein kinase JAK2 by the growth hormone receptor. *Science* 344, 1249783.
- Canning, P., Tan, L., Chu, K., Lee, S.W., Gray, N.S., and Bullock, A.N. (2014). Structural mechanisms determining inhibition of the collagen receptor DDR1 by selective and multi-targeted type II kinase inhibitors. *J. Mol. Biol.* 426, 2457-2470.
- Carafoli, F., Mayer, M.C., Shiraishi, K., Pecheva, M.A., Chan, L.Y., Nan, R., Leitinger, B., and Hohenester, E. (2012). Structure of the discoidin domain receptor 1 extracellular region bound to an inhibitory Fab fragment reveals features important for signaling. *Structure* 20, 688-697.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106-1117.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* *17*, 1253-1270.

Erler, J.T., Bennewith, K.L., Nicolau, M., Dornhofer, N., Kong, C., Le, Q.T., Chi, J.T., Jeffrey, S.S., and Giaccia, A.J. (2006). Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* *440*, 1222-1226.

Gao, D., and Chen, Y. (2015). Organoid development in cancer genome discovery. *Curr. Opin. Genet. Dev.* *30*, 42-48.

Gao, H., Chakraborty, G., Lee-Lim, A.P., Mavrakis, K.J., Wendel, H.G., and Giancotti, F.G. (2014). Forward genetic screens in mice uncover mediators and suppressors of metastatic reactivation. *Proc. Natl. Acad. Sci. U.S.A* *111*, 16532-16537.

Gao, H., Chakraborty, G., Lee-Lim, A.P., Mo, Q., Decker, M., Vonica, A., Shen, R., Brogi, E., Brivanlou, A.H., and Giancotti, F.G. (2012). The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. *Cell* *150*, 764-779.

Ghajar, C.M. (2015). Metastasis prevention by targeting the dormant niche. *Nat. Rev. Cancer* *15*, 238-247.

Ghajar, C.M., Peinado, H., Mori, H., Matei, I.R., Evason, K.J., Brazier, H., Almeida, D., Koller, A., Hajjar, K.A., Stainier, D.Y., *et al.* (2013). The perivascular niche regulates breast tumour dormancy. *Nat. Cell Biol.* *15*, 807-817.

Giancotti, F.G. (2013). Mechanisms governing metastatic dormancy and reactivation. *Cell* *155*, 750-764.

Guo, W., Pylayeva, Y., Pepe, A., Yoshioka, T., Muller, W.J., Inghirami, G., and Giancotti, F.G. (2006). Beta 4 integrin amplifies ErbB2 signaling to promote mammary tumorigenesis. *Cell* *126*, 489-502.

Hellstrom, I., Beaumier, P.L., and Hellstrom, K.E. (1986). Antitumor effects of L6, an IgG2a antibody that reacts with most human carcinomas. *Proceedings of the Proc. Natl. Acad. Sci. U.S.A* *83*, 7059-7063.

Kobayashi, A., Okuda, H., Xing, F., Pandey, P.R., Watabe, M., Hirota, S., Pai, S.K., Liu, W., Fukuda, K., Chambers, C., *et al.* (2011). Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone. *J Exp. Med.* *208*, 2641-2655.

Leitinger, B. (2011). Transmembrane collagen receptors. *Ann. Rev. Cell Dev. Biol.* *27*, 265-290.

Levental, K.R., Yu, H., Kass, L., Lakins, J.N., Egeblad, M., Erler, J.T., Fong, S.F., Csiszar, K., Giaccia, A., Weninger, W., *et al.* (2009). Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* *139*, 891-906.

Marcotte, R., Brown, K.R., Suarez, F., Sayad, A., Karamboulas, K., Krzyzanowski, P.M., Sircoulomb, F., Medrano, M., Fedyszyn, Y., Koh, J.L., *et al.* (2012). Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov.* *2*, 172-189.

Marken, J.S., Bajorath, J., Edwards, C.P., Farr, A.G., Schieven, G.L., Hellstrom, I., Hellstrom, K.E., and Aruffo, A. (1994). Membrane topology of the L6 antigen and identification of the protein epitope recognized by the L6 monoclonal antibody. *J. Biol. Chem.* *269*, 7397-7401.

Marken, J.S., Schieven, G.L., Hellstrom, I., Hellstrom, K.E., and Aruffo, A. (1992). Cloning and expression of the tumor-associated antigen L6. *Proc. Natl. Acad. Sci. U.S.A* 89, 3503-3507.

Mihai, C., Chotani, M., Elton, T.S., and Agarwal, G. (2009). Mapping of DDR1 distribution and oligomerization on the cell surface by FRET microscopy. *J. Mol. Biol.* 385, 432-445.

Noordeen, N.A., Carafoli, F., Hohenester, E., Horton, M.A., and Leitinger, B. (2006). A transmembrane leucine zipper is required for activation of the dimeric receptor tyrosine kinase DDR1. *J. Biol. Chem.* 281, 22744-22751.

Paszek, M.J., Zahir, N., Johnson, K.R., Lakins, J.N., Rozenberg, G.I., Gefen, A., Reinhart-King, C.A., Margulies, S.S., Dembo, M., Boettiger, D., *et al.* (2005). Tensional homeostasis and the malignant phenotype. *Cancer Cell* 8, 241-254.

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., *et al.* (2007). Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 131, 1190-1203.

Samuel, M.S., Lopez, J.I., McGhee, E.J., Croft, D.R., Strachan, D., Timpson, P., Munro, J., Schroder, E., Zhou, J., Brunton, V.G., *et al.* (2011). Actomyosin-mediated cellular tension drives increased tissue stiffness and beta-catenin activation to induce epidermal hyperplasia and tumor growth. *Cancer Cell* 19, 776-791.

Sarkar, D., Boukerche, H., Su, Z.Z., and Fisher, P.B. (2008). mda-9/Syntenin: more than just a simple adapter protein when it comes to cancer metastasis. *Cancer Res.* 68, 3087-3093.

Shan, Y., Gnanasambandan, K., Ungureanu, D., Kim, E.T., Hammaren, H., Yamashita, K., Silvennoinen, O., Shaw, D.E., and Hubbard, S.R. (2014). Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase. *Nat. Struct. Mol. Biol.* 21, 579-584.

Tam, W.L., Lu, H., Buikhuisen, J., Soh, B.S., Lim, E., Reinhardt, F., Wu, Z.J., Krall, J.A., Bierie, B., Guo, W., *et al.* (2013). Protein kinase C alpha is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell* 24, 347-364.

Toms, A.V., Deshpande, A., McNally, R., Jeong, Y., Rogers, J.M., Kim, C.U., Gruner, S.M., Ficarro, S.B., Marto, J.A., Sattler, M., *et al.* (2013). Structure of a pseudokinase-domain switch that controls oncogenic activation of Jak kinases. *Nat. Struct. Mol. Biol.* 20, 1221-1223.

Valencia, K., Ormazabal, C., Zanduetta, C., Luis-Ravelo, D., Anton, I., Pajares, M.J., Agorreta, J., Montuenga, L.M., Martinez-Canarias, S., Leitinger, B., *et al.* (2012). Inhibition of collagen receptor discoidin domain receptor-1 (DDR1) reduces cell survival, homing, and colonization in lung cancer bone metastasis. *Clin. Cancer Res.* 18, 969-980.

Vogel, W., Brakebusch, C., Fassler, R., Alves, F., Ruggiero, F., and Pawson, T. (2000). Discoidin domain receptor 1 is activated independently of beta(1) integrin. *J. Biol. Chem.* 275, 5779-5784.

Wang, C.Z., Su, H.W., Hsu, Y.C., Shen, M.R., and Tang, M.J. (2006). A discoidin domain receptor 1/SHP-2 signaling complex inhibits alpha2beta1-integrin-mediated signal transducers and activators of transcription 1/3 activation and cell migration. *Mol. Biol. Cell* 17, 2839-2852.

Wright, M.D., Ni, J., and Rudy, G.B. (2000). The L6 membrane proteins--a new four-transmembrane superfamily. *Prot. Sci.* 9, 1594-1600.

Wu-Zhang, A.X., and Newton, A.C. (2013). Protein kinase C pharmacology: refining the toolbox. *Biochem. J.* 452, 195-209.

Xu, L., Li, Q., Xu, D., Wang, Q., An, Y., Du, Q., Zhang, J., Zhu, Y., and Miao, Y. (2014). hsa-miR-141 downregulates TM4SF1 to inhibit pancreatic cancer cell invasion and migration. *Int. J. Oncol.* 44, 459-466.

Young, R.A. (2011). Control of the embryonic stem cell state. *Cell* 144, 940-954.

Yu, H., Lee, H., Herrmann, A., Buettner, R., and Jove, R. (2014). Revisiting STAT3 signalling in cancer: new and unexpected biological functions. *Nat. Rev. Cancer* 14, 736-746.

Zhao, D., Pan, C., Sun, J., Gilbert, C., Drews-Elger, K., Azzam, D.J., Picon-Ruiz, M., Kim, M., Ullmer, W., El-Ashry, D., *et al.* (2014). VEGF drives cancer-initiating stem cells through VEGFR-2/Stat3 signaling to upregulate Myc and Sox2. *Oncogene*.

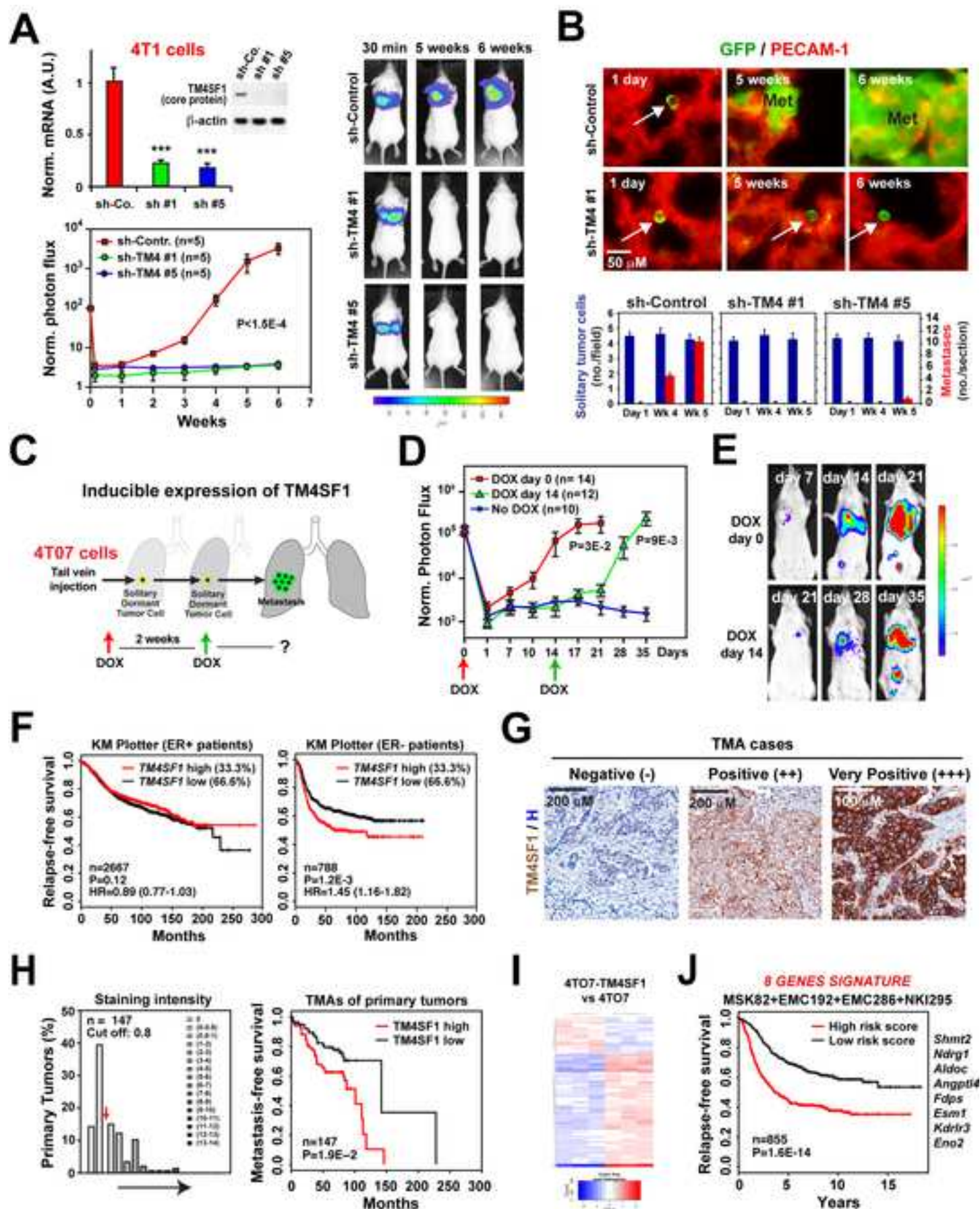


Figure 1

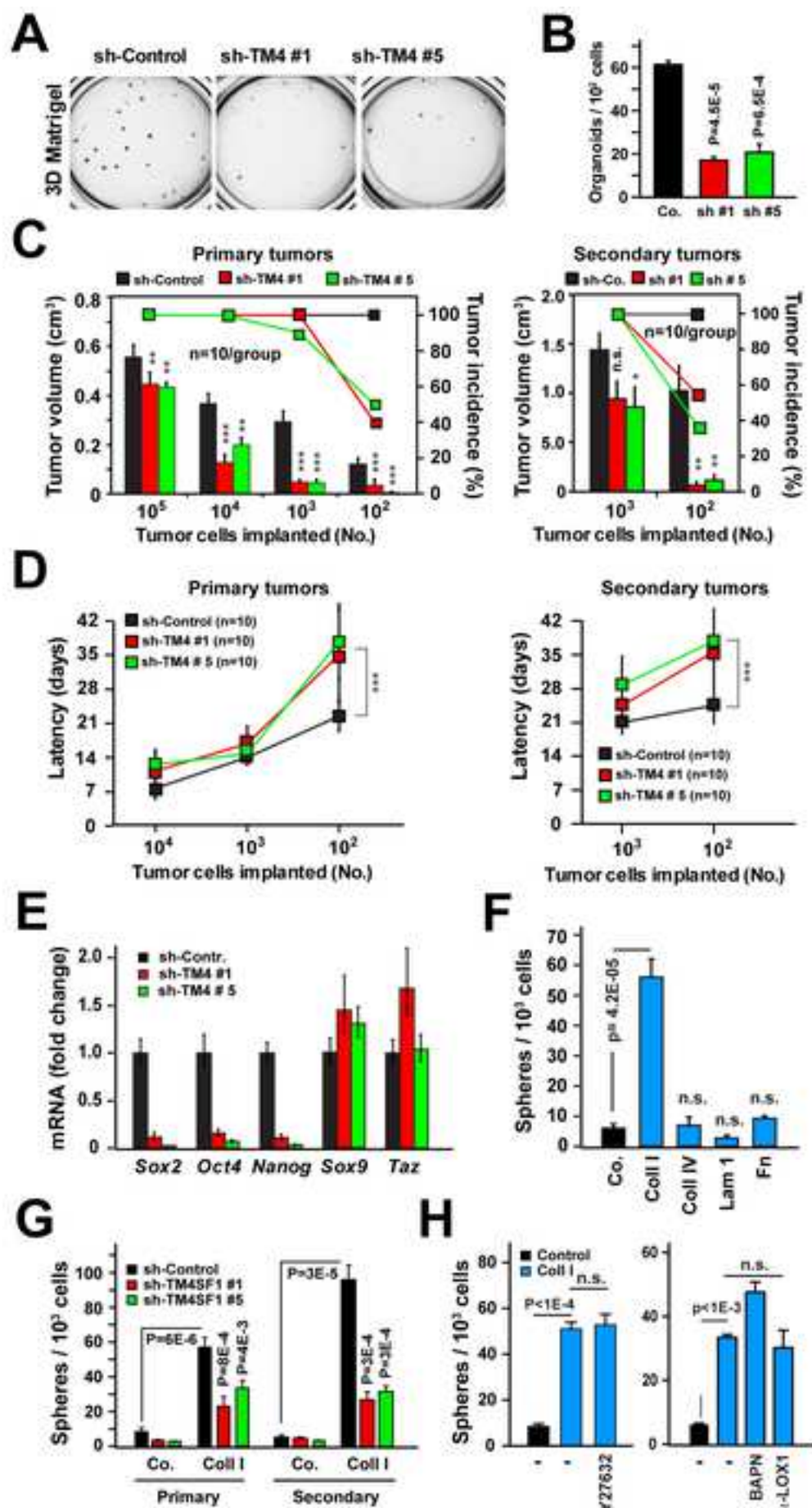


Figure 2

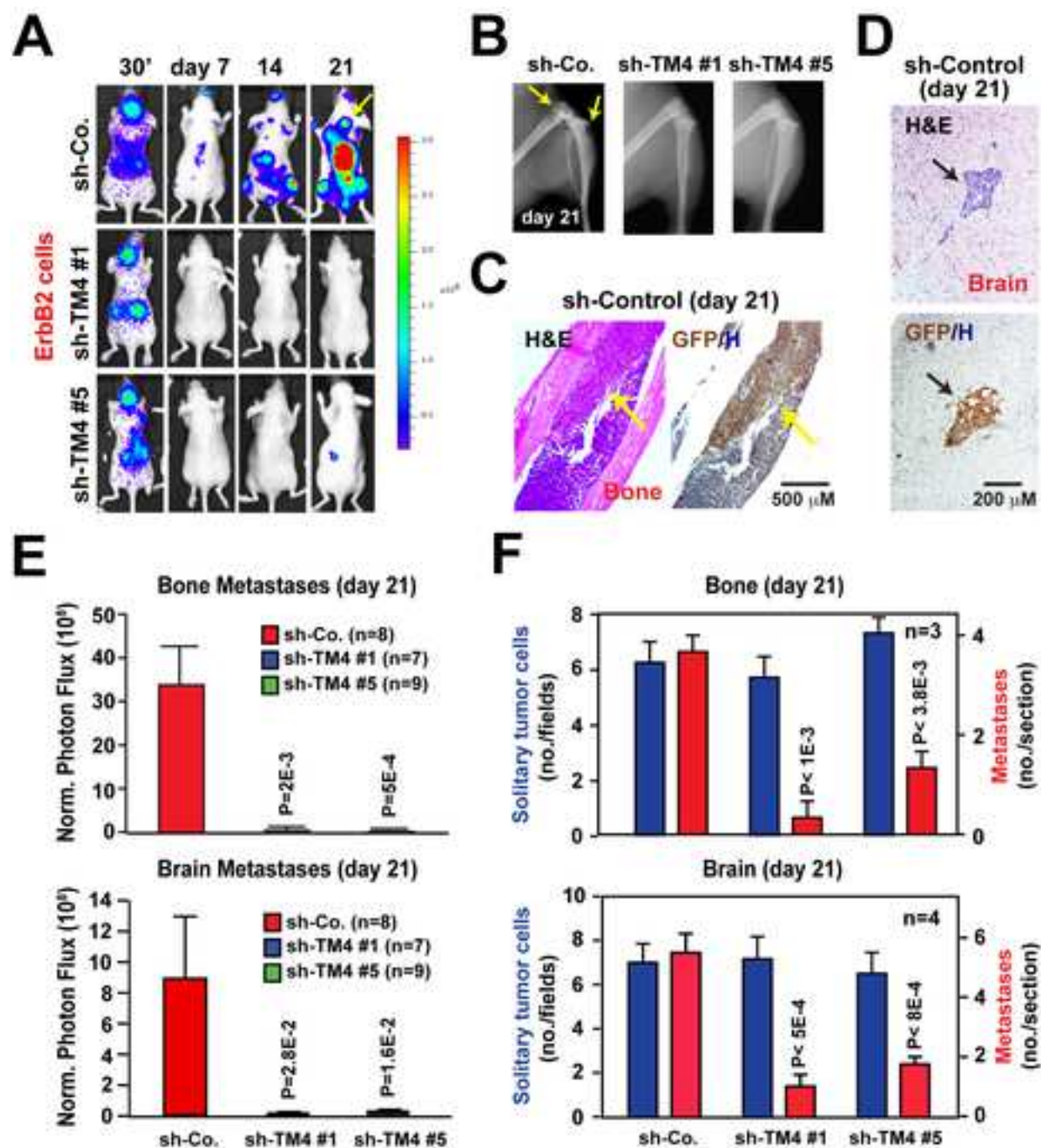


Figure 3

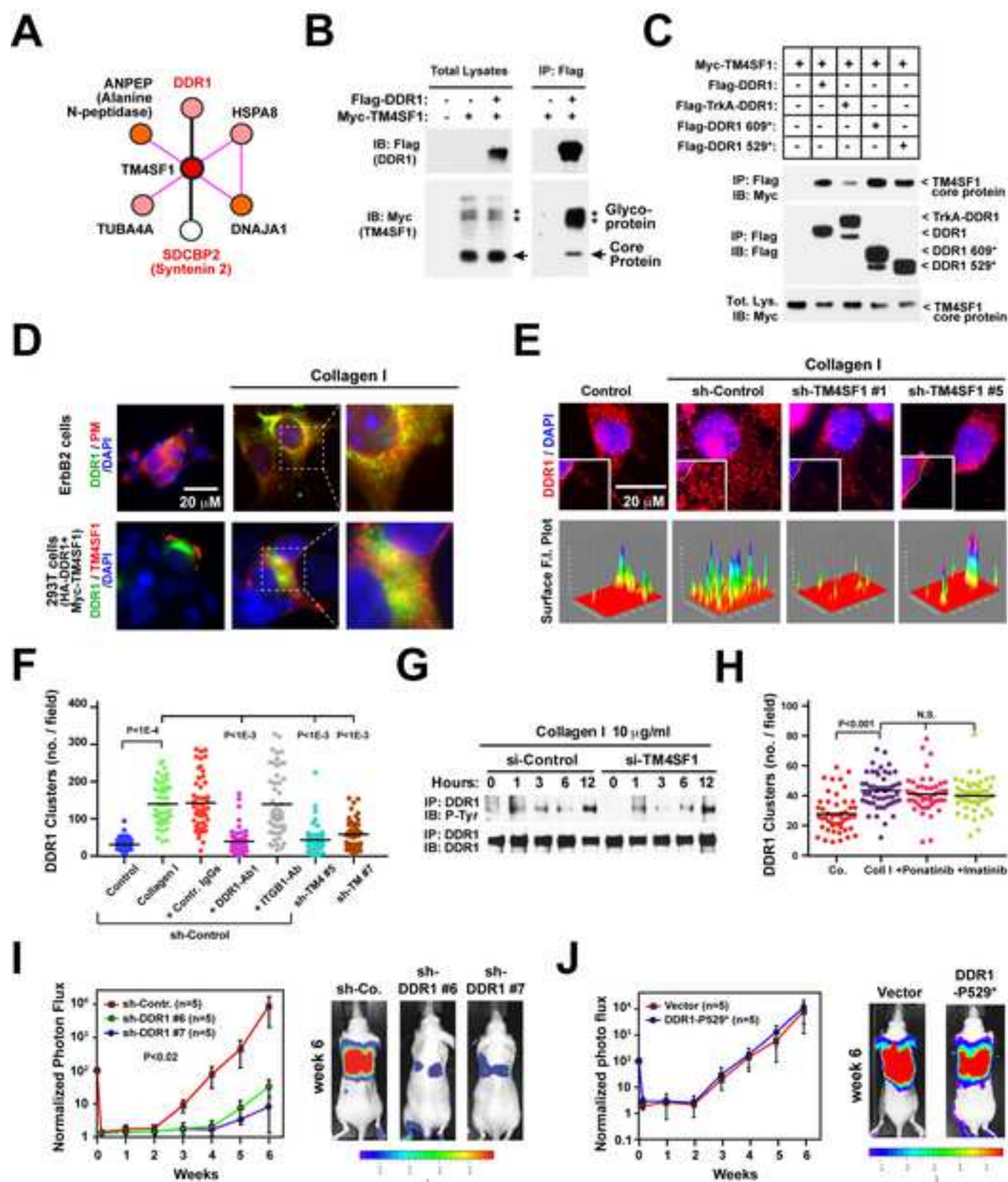


Figure 4

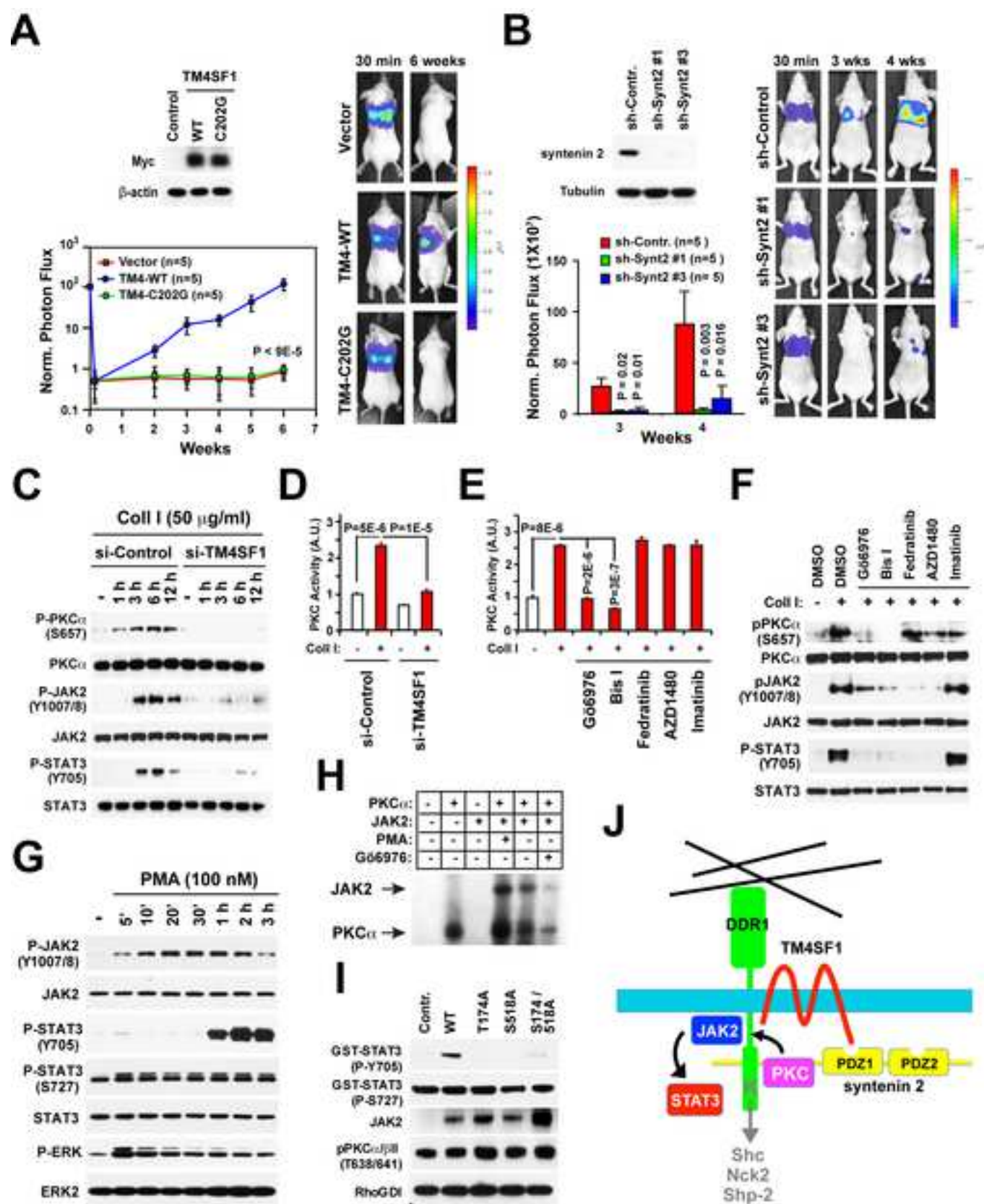


Figure 5

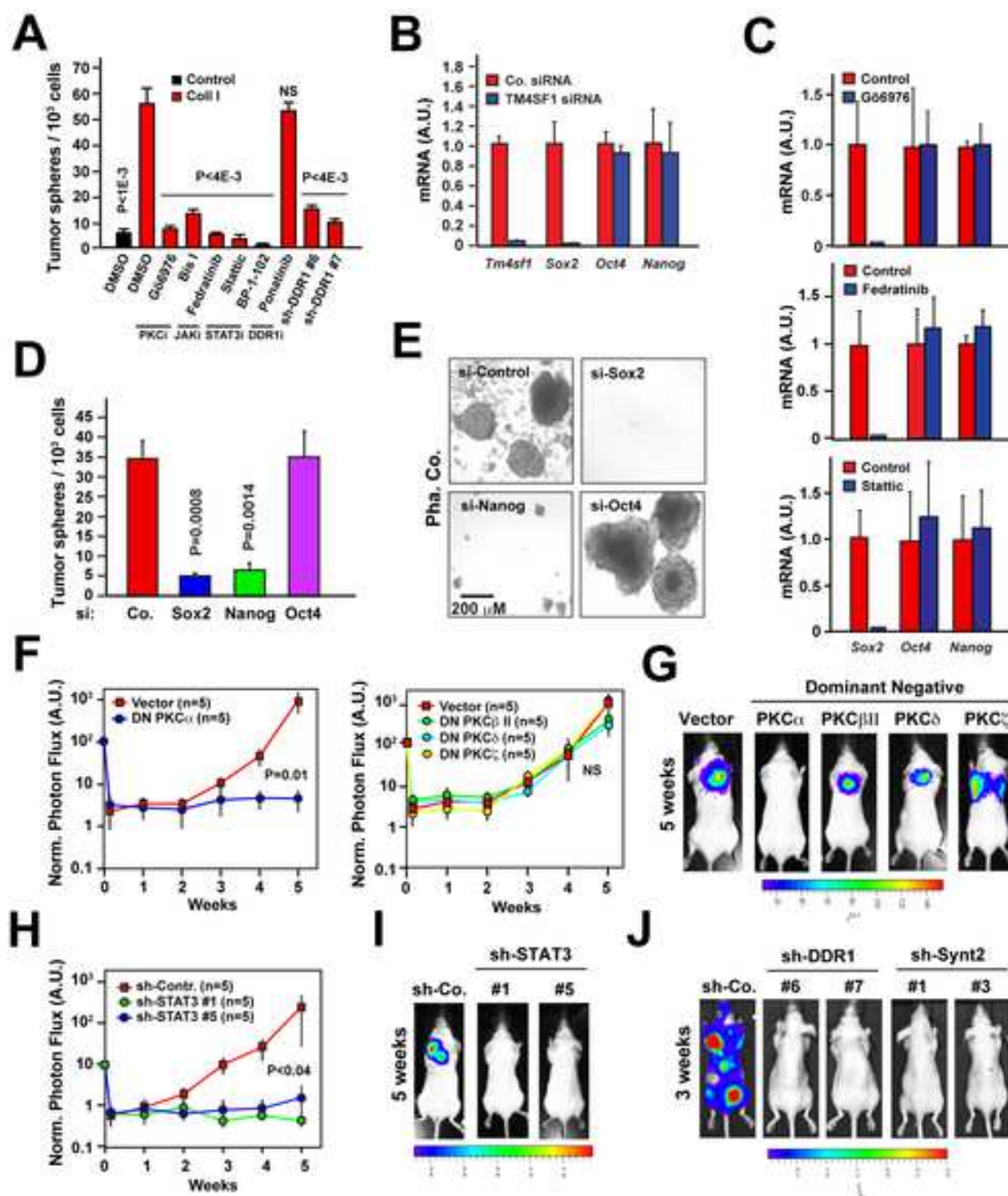


Figure 6

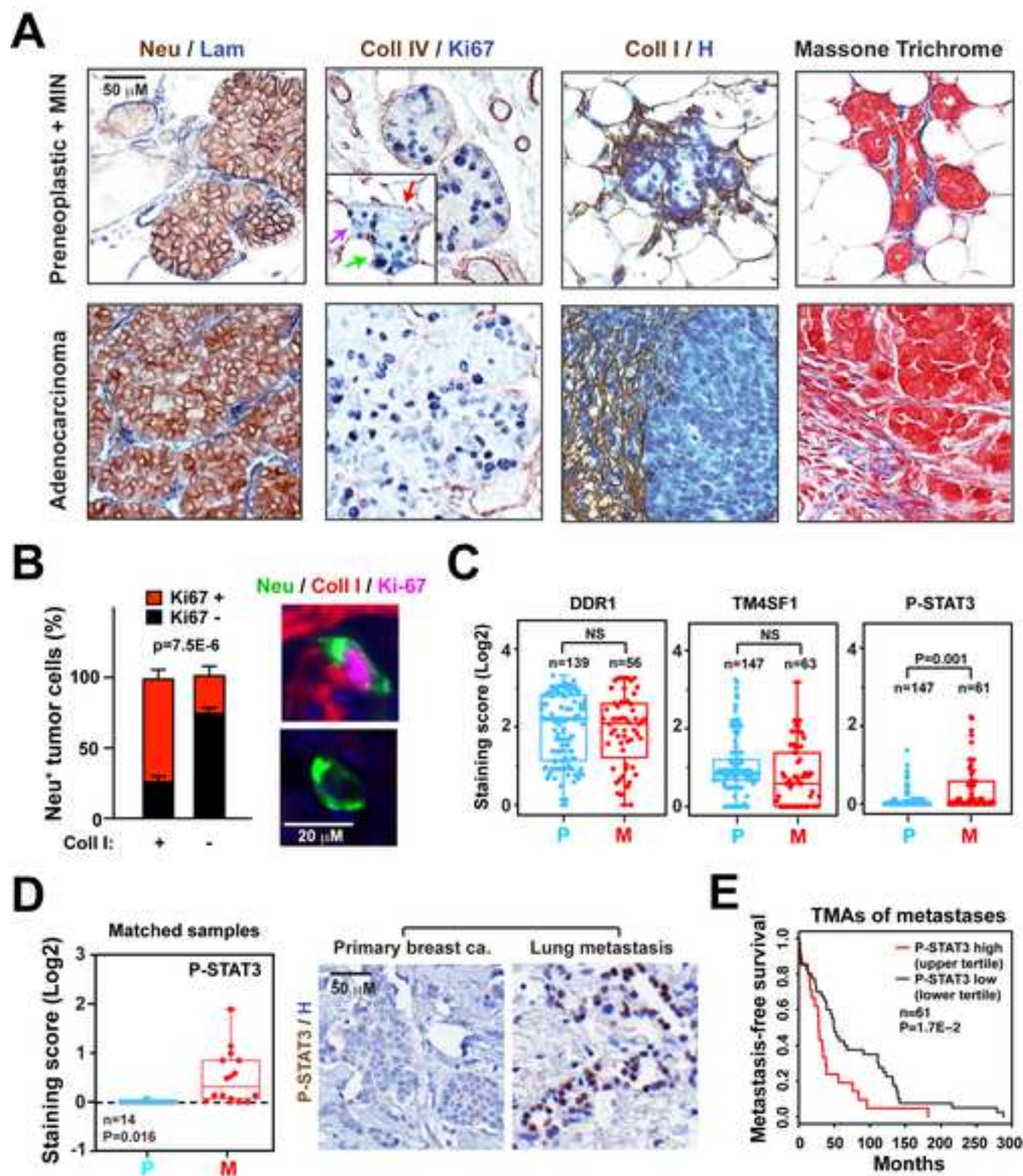


Figure 7

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL DISCUSSION

Non-canonical DDR1 signaling possesses several distinguishing features. Monomeric collagen I bind to DDR1 and induces a conformational change that enables or enhances its association with TM4SF1, leading to the formation of large receptor clusters that persist at the cell surface for up to 3 hours. In response to collagen I - and presumably resulting from the constitutive association of TM4SF1 with syntenin2 – PKC α is recruited near the plasma membrane and phosphorylates and activates JAK2, leading to activation of STAT3 (Figure 5J). In contrast, non-fibrillar collagen IV induces only the formation of small and transient DDR1 clusters, which are presumably unable to sustain non-canonical signaling, suggesting that contact with the basement membrane is insufficient to activate non-canonical DDR1 signaling.

The activation of the tyrosine kinase domain of DDR1 proceeds with unusually long kinetics: maximal activation is observed only several hours after the binding of collagen to the ectodomain of the receptor. In contrast, the aggregation of DDR1 on the plasma membrane occurs rapidly and begins to decline before patent activation of the DDR1 tyrosine kinase domain, suggesting that non-canonical signaling precedes canonical signaling. In agreement with the ability of TM4SF1 to function as a membrane adaptor, mediating both clustering of DDR1 and activation of non-canonical signaling, both PKC α and JAK2 become activated during the initial phase of DDR1 signaling. The second and more pronounced peak of JAK2 activation precedes full activation of STAT3, which persists for the duration of DDR1 signaling (Figure S5G). We presume that the complex kinetics of this signaling system originates from the existence of negative feed back loops as well as amplification steps.

The DDR1 signalosomes are approximately as large as focal adhesions and immunological synapses and may therefore be as complex in composition and function. Given their involvement in the activation of PKC α , it is possible that

these structures also mediate the activation of Fra1 and NF κ B, which have been implicated in cancer stem cell function and metastasis (Das et al., 2006; Tam et al., 2013). In addition, tumor cells adhering to a collagen matrix may form even more complex adhesive junctions, containing both DDR1 and integrin receptors, allowing for cross-communication between the two signaling systems (Shintani et al., 2008).

EXTENDED EXPERIMENTAL PROCEDURES

Reagents

Collagen I (C3867), collagen IV (0543), the PKC activator PMA, the PKC inhibitor Gö6976, the non-muscle myosin II ATPase inhibitor Blebbistatin (B0506), and phalloidin (P5282) were from Sigma. The JAK2 inhibitors Fedratinib (S2736) and AZD1480 (S2162), the Abl and DDR1 inhibitors Ponatinib (S1490) and Imatinib (S1026) were from Selleck. The PKC inhibitor Bisindolylmaleimide I (Bis I, 133052-90-1) was from Cayman. The STAT3 inhibitor Stattic (2798) was from Tocris. The ROCK inhibitor Y-27632 was from Axxora. The FM4-64 membrane dye was purchased from Molecular Probes.

Cell lines

The 4TO7 and 4T1 cell lines were generously provided by Dr. Fred Miller (Wayne State University, Detroit, MI) and cultured as described (Aslakson and Miller, 1992). The bone metastatic BoM-1833 cells were obtained from Dr. Joan Massagué (Cancer Biology and Genetics Program, MSKCC) and cultured in DME-HG with 10% FBS. The JAK2-deficient γ 2A cells (γ 2A) were a kind gift of Dr. Ross Levine (Human Oncology and Pathogenesis Program, MSKCC) and were cultured in DME-HG with 10% FBS. MDA-MB-231, MCF7 cells (from ATCC) and 293FT cells (from Life Technologies) were cultured in the same medium. The mouse ErbB2-transformed mammary tumour cells were derived from MMTV-*Neu*(YD) mice and reconstituted with a wild-type form of rat Neu under the control of the CMV promoter (Guo et al., 2006). For bioluminescent tracking, cell lines were infected with a retroviral vector encoding herpes simplex

virus thymidine kinase 1, green fluorescent protein (GFP) and firefly luciferase (Ponomarev et al., 2004). GFP-positive cells were isolated by FACS.

Antibodies

Antigen	Imm. Animal	Clone	Catalog#	Fluorophore	Dilution	Vendor
Actin beta	mouse	AC-74	A5316		WB 1:10,000	Sigma
CD24	rat	M1/69	561079	PE	FACS (1µg/10 ⁶ cells)	BD
CD29	rat	HMβ1-1	102221	PE/Cy7	FACS (1µg/10 ⁶ cells)	Biolegend
CD49f	rat	GoH3	561893	FITC	FACS (1µg/10 ⁶ cells)	BD
Collagen I	rabbit	Polyclona	Ab34710		WB 1:1000, IHC 1:50 Blocking 5 µg/ml	Abcam
Collagen IV	rabbit	Polyclona	2150-1470		IHC 1:300	AbD Serotec
DDR1	rabbit	D1G6	#5583		WB 1:1,000; IF 1:300; IHC 1:100	CST
DDR1	rabbit	Polyclona	sc-532		WB 1:300	Santa Cruz
DDR1	mouse	1F7			Blocking: 10µg/ml	(Carafoli et al., 2012)
ErbB2	rabbit	A24-V	ALX-810-		IHC 1:100	Enzo
Erk2	rabbit	Polyclona	sc-154		IHC 1:300	Santa Cruz
p-Erk1/2 T202/Y204	rabbit	Polyclona	#9101		IHC 1:500	CST
p-FAK Y397	rabbit	Polyclona	44-624G		WB 1:1,000	Invitrogen
FLAG	mouse	M2	F3165		WB 1µg/ml	Sigma
GFP	chicken	Polyclona	ab13970		IF 1:100; IHC 1:300	Abcam
HA	mouse	12CA5			WB 1µg/ml	MSKCC facility
p-IkBα S32/36	mouse	5A5	#9246		WB 1:300	CST
Jak2	rabbit	D2E12	#3230		WB 1:1,000	CST

p-Jak2 Y1007/1008	rabbit	C80C3	#3776		WB 1:300	CST
Ki67	mouse	MM1	NCL-Ki67- MM1		IHC 1:100	Novocastra
Laminin	rabbit	Polyclona	L9393		IHC 1:30	Sigma
Myc-Tag	mouse	9B11	#2276		WB 1:300	CST
p-p38 T180/Y182	rabbit	Polyclona	#9211		WB 1:1,000	CST
PECAM-1	rat	MEC 13.3	550274		IHC 1:20	BD
PKC α	rabbit	Polyclona	sc-208		WB 1:300	Santa Cruz
p-PKC α S657	goat	Polyclona	sc-12356		WB 1:200	Santa Cruz
p- PKC α / β II T638/641	rabbit	Polyclona	#9375		WB 1:300	CST
p- PKC δ T505	rabbit	Polyclona	#9374		WB 1:300	CST
p- PKC δ / θ S643/676	rabbit	Polyclona	#9376		WB 1:300	CST
RhoGDI	rabbit	Polyclona	Sc#360		WB 1:1000	Santa Cruz
Sca1	rat	D7	108124	PerCP/ Cy5.5	FACS (0.25 μ g/10 ⁶ cells)	Biolegend
Stat3	mouse	124H6	#9139		WB 1:1,000	CST
p-Stat3 Y705	rabbit	D3A7	#9145		WB 1:200; IHC 1:100	CST
p-Stat3 S727	mouse	6E4	#9136		WB 1:300	CST
Syntenin 2	rabbit	Polyclona	10407-1-AP		WB 1:1,000	Protein Tech
TM4SF1	rabbit	Polyclona	C10517		WB 1:200; IF 1: 100 IHC 1:50	Assay BioTech
p-Tyr	mouse	4G10	05-321		WB 1 μ g/ml	Millipore

Constructs, Gene Expression and Silencing

cDNAs encoding full-length mouse TM4SF1, human DDR1, human DDR1 P529*, human DDR2, mouse syntenin 2, human PKC α , human JAK2, mouse JAK2, and Dominant Negative PKCs (human PKC α K368R, mouse PKC β II K371R, mouse PKC δ K376R, rat PKC ζ K281M) were a kind gift from Dr. Christopher A.

McCilloch (University of Toronto, Canada), Dr. Ming-Jer Tang (National Cheng Kung University Medical College, Tainan 701, Taiwan), and Dr. Ross L. Levine (MSKCC, USA), or were obtained from Open Biosystems and Addgene. The TM4SF1 point mutant (C202G), DDR1 truncation mutant (1-609), TrkA-DDR1 chimera (TrkA1-423/DDR1418-913), and JAK2 point mutants (T174A, S518A, and S174A/S518A) were generated by PCR using complementary primers. Wild type and mutant cDNAs were sub-cloned into pBabepuro or pQCXIP (Clontech) retroviral vectors, and verified by sequencing. The lentiviral vector for doxycycline-inducible expression of Myc-TM4SF1 was generated by subcloning the corresponding cDNA between the Nhe1 and BamH I sites of pCW (Wang et al., 2014). pLKO.1 vectors encoding short hairpin RNAs against mouse TM4SF1 (#1, TRCN0000112785 and #5, TRCN0000112787), human TM4SF1 (#1, TRCN0000000286 and #5, TRCN0000000290), mouse DDR1 (#6, TRCN0000023369 and #7, TRCN0000023373), mouse DDR2 (#3, TRCN0000023551 and 4 TRCN0000023552) mouse syntenin 2 (#1, TRCN0000100670 and #3, TRCN0000100672), and mouse STAT3 (#1, TRCN0000071453 and #5, TRCN0000071457), as well as control vectors were obtained from Open Biosystems or Sigma. The psi lenti vector encoding short hairpin RNAs against human TM4SF1 (#7, HSH010941-3-HIVH1 OS271955) were obtained from GeneCopoeia. Retroviral and lentiviral vectors encoding recombinant proteins or shRNAs were generated as previously described (Pylayeva et al., 2007). siRNA smartpools targeting mouse TM4SF1 (L-060359-01-0005), human TM4SF1 (E-010610-00-0005), mouse Sox2 (E-058489-00-0005), mouse Nanog (E-057004-00-0005), mouse OCT4 (E-046256-00-0005), and non-targeting control (D-001810-10-20) were from Dharmacon. Efficiency of knock down was verified by Q-PCR or immunoblotting.

Animal Studies

4TO7 and 4T1 cells were transplanted in 6-12 weeks-old syngeneic BALB/c mice; MDA-MB231, BoM-1833, and ErbB2 cells were xenografted in 6-12 weeks-old nude BALB/c or NOD *scid* gamma (NSG) mice. Orthotopic injections in the mammary fat pad and experimental metastasis assays were performed

essentially as previously described (Gao et al., 2014; Gao et al., 2012; Guo et al., 2006; Pylayeva et al., 2009). The size of primary tumors was measured by vernier calliper. To examine lung colonization, cells transduced with the TGL vector (Ponomarev et al., 2004) were harvested by trypsinization, washed twice in PBS, resuspended at 3×10^5 (4TO7 cells and their derivatives transduced with TM4SF1 or TM4SF1 C202G), 1×10^5 (4T1 cells and their derivatives transduced with shTM4SF1 #1, shTM4SF1 #5, shSTAT3 #1, shSTAT3 #5, DN-PKC α , DN-PKC β II, DN-PKC δ or DN-PKC ζ), or 5×10^5 (ErbB2 cells and their derivatives transduced with shTM4SF1 #1, shTM4SF1 #5, shDDR1 #6, shDDR1 #7, shSyntenin2 #1, shSyntenin2 #3, shSTAT3 #1, shSTAT3 #5, or DDR1-P529*), 2×10^6 (MDA-MB-231 cells and their derivatives transduced with shTM4SF1 #1 or shTM4SF1 #5) in 50 μ l of PBS, and injected in the tail vein. To examine colonization of the bone and brain, 4T1 (3×10^5), ErbB2 cells and their derivatives transduced with shTM4SF1 #1, shTM4SF1 #5, shDDR1 #6, shDDR1 #7, shSyntenin2 #1, shSyntenin2 #3 (3×10^5), and BoM-1833 cells and their derivatives transduced with shTM4SF1 #5, shTM4SF1 #7 (1×10^5) were injected in the left ventricle of nude BALB/c mice in 100 μ l of HBSS. Bioluminescent imaging was used to verify successful injection and to monitor metastatic outgrowth. Metastatic lesions were confirmed by histological analysis.

For tumor initiation experiments, the indicated numbers of TGL-labeled ErbB2 cells were suspended in 100 μ l DME-HG with growth factor-reduced Matrigel (Corning: 354230) in 1:1 ratio and injected to the No. 4 mammary gland of 6-12 weeks-old NSG mice. Primary tumor growth was monitored weekly by taking measurements of tumor length (L) and width (W). Tumor volume was calculated using the formula $\pi LW^2/6$. For serial transplantation, primary tumors were digested using Accutase (Innovative Cell Technologies, AT-104) and GFP-positive cells were sorted and used for secondary tumor initiation experiments.

Bioluminescent and X-ray Imaging

For bioluminescent imaging, mice were anesthetized and injected with 1.5 mg of D-luciferin retro-orbitally at the indicated times. Animals were imaged in an IVIS 100 chamber within 5 min after D-luciferin injection, and data were recorded using Living Image software (Xenogen). To measure lung colonization, photon flux was calculated by using a circular region of interest encompassing the thorax of the mouse. After subtracting a background value obtained from a control mouse injected only with D-luciferin, photon flux was normalized to the value obtained 30 min after injection of the tumor cells. This latter value was arbitrarily set at 100. To measure colonization of other organs after intracardiac injection, mice were imaged as described above except that values were not normalized after subtracting the background. Bone metastases were confirmed by X-Ray imaging. Mice were anesthetized with Ketamine (100 mg/kg) and xylazine (10 mg/kg), placed on digital X ray Film (Scan X) and exposed at 25kV for 15s using a Faxitron instrument (Model MX-20; Faxitron Corp. Buffalo, IL).

Mouse Pathology

MMTV-*Neu*(YD) mice were sacrificed and primary tumors and lungs were collected at ~5 months, when the large majority of these mice exhibit a mixture of micrometastases and macrometastases (Dankort et al., 2001; Guo et al., 2006). Tissue was fixed in phosphate-buffered 4% PFA at 4°C, transferred to 70% ethanol, and blocked in paraffin. Sections were stained with hematoxylin and eosin. Lung metastases were evaluated by microscopic analysis of lung sections. Masson Trichrome and Picrosirius Red staining were performed in the Molecular Cytology Core Facility of MSKCC. Adjacent sections were stained with haematoxylin and eosin. Pathological examination and histological classification of the tumors were performed as described previously (Simpson et al., 2003)

Immunostaining

Immunostaining of mouse lungs was performed as previously described (Guo et al., 2006; Pylayeva et al., 2009). Mice were sacrificed and perfused with PBS through the left ventricle. Lungs were inflated with ice-cold 4% paraformaldehyde through an intra-tracheal injection, fixed overnight, and embedded in OCT or

paraffin. Primary tumors were also fixed with 4% paraformaldehyde overnight and embedded in OCT. To examine extravasation and outgrowth in the lung, 10 μ m-thick non-consecutive sections were cut on a Leica microtome, and immunofluorescence was performed with anti-PECAM-1 by using the Tyramide Signal Amplification Kit (#21, Invitrogen) to visualize lung capillaries and anti-GFP to visualize tumor cells, followed by detection with fluorescently conjugated secondary antibodies (Jackson ImmunoResearch). Solitary tumor cells were randomly counted in 10 fields per section, and metastatic lesions were counted in whole sections (2–3 sections per mouse). Immunohistochemistry on paraffin-embedded sections was performed in Molecular Cytology Core Facility of MSKCC by using the peroxidase-DAB or alkaline phosphatase-BCIP/NBT staining kits (M.O.M. and VECTASTAIN ABC) from Vector Laboratories. Immunohistochemical staining for Neu, laminin, collagen I and IV, Ki67, DDR1, TM4SF1, and P-STAT3 was performed on preneoplastic and MIN lesions, adenocarcinomas, and lung micrometastases and macrometastases as indicated. Immunohistochemical staining for GFP was used to confirm metastatic lesions on bone sections. Human TM4SF1, DDR1, and P-STAT3 immunohistochemical staining was performed using paraffin-embedded tissue microarrays at the Breast Pathology Service of MSKCC. Results were analysed by two investigators in a double blind manner. The expression score was determined by combining staining intensity and the percentage of immunoreactive cells. Tissues with no staining were rated as 0, with a faint or moderate staining in less than 25% of cells as 0.5, with strong staining in less than 25% of cells or with moderate staining in 25-50% of cells as 1, with strong staining in 25-50% of cells or with moderate staining in more than 50% of cells as 1.5, with strong staining in more than 50% of cells as >2.

Paraffin-embedded tissue microarrays containing sections from multiple primary breast tumor cores or lung metastatic lesion cores were obtained from the MSKCC Department of Pathology in compliance with protocols approved by the MSKCC Institutional Review Board and after the subjects gave their informed consent. The levels of TM4SF1 and DDR1 and the degree of activation of STAT3

were determined by immunohistochemistry. TM4SF1, DDR1, and P-STAT3 immunoreactivity were evaluated and scored by a clinical pathologist (E.B.) and by a second investigator in a double blind fashion. Cells with no staining were rated as 0, with a faint staining as 1, with moderate staining as 2, and with strong staining as 3. Tissues were then rated as the sum of the square of staining intensity weighted by the percentage of corresponding immunoreactive cells. Patient samples with no tissue core left were discarded. Metastasis-free survival was examined as a function of TM4SF1 or P-STAT3 abundance in breast tumor samples or lung metastatic samples.

Analysis of protein and mRNA expression

For immunoblotting, cells were washed with PBS and lysed in RIPA buffer unless otherwise noted (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate, and 0.1% SDS) supplemented with protease inhibitors (Calbiochem) and phosphatase inhibitors (PhosSTOP, Roche Life Science). Protein concentrations were measured by using the DC Protein Assay. Other lysis buffers used were OGP buffer (25 mM Hepes·KOH, pH 7.6, 100 mM KCl, 2 mM EGTA, 0.8% octyl β -D-glucopyranoside, 0.2 mM dithiothreitol), Brij-96 buffer (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 1% Brij-96, 0.02% NaN_3), and CHAPS buffer (PBS containing 1% Chaps, 5 mM EDTA) supplemented with protease and phosphatase inhibitors.

Total RNA was extracted using the RNeasy Mini kit coupled with RNase-free DNase set (Qiagen) and reverse transcribed with SuperScript III First-Strand Synthesis SuperMix (Invitrogen). cDNA corresponding to approximately 10 ng of starting RNA was used for one reaction. Q-PCR was performed with Taqman Gene Expression Assay (Applied Biosystems). All quantifications were normalized to endogenous GAPDH. Primers sequence used to amplify genes are listed below.

TaqMan Gene Expression Assays	
Gene Name	Assay ID
TM4SF1	Hs00371997_m1
GAPDH	Hs99999905_m1
TM4SF1	Mm00447009_m1
Syntenin2	Mm01245859_m1
Sox2	Mm03053810-s1
Sox9	Mm00448840_m1
OCT4	Mm00658129_gh
Nanog	Mm02384862_g1
Taz	Mm00504978_m1
GAPDH	Mm99999915_g1

Co-immunoprecipitation assay

For coimmunoprecipitation analysis, cells were lysed in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate, and 0.1% SDS) supplemented with protease inhibitors and phosphatase inhibitors. To immunoprecipitate Flag-tagged protein, total lysates containing 4-5 mg proteins were incubated with 10 µl Flag M2 affinity gel (A2220, Sigma), and the protein–antibody mixture was then eluted with Flag peptide (F4799, Sigma) according to the manufacturer's protocols. To immunoprecipitate DDR1, total lysates containing 4-5 mg proteins were incubated with 20 µl anti-DDR1 rabbit Mab D1G6 (5583, Cell Signaling) and 10 µl Protein G plus agarose (22852, Pierce), and the protein–antibody mixture was then eluted with a synthetic peptide corresponding to residues surrounding Pro600 of human DDR1 protein (Cell Signaling). Immunoprecipitates and total lysates were subjected to immunoblotting with the indicated antibodies.

To demonstrate TM4SF1 mediated endogenous DDR1 and syntenin 2 interaction, total ErbB2 cell lysate containing 1 mg of protein incubated with 2µg of anti-Syntenin2 rabbit polyclonal antibody (10407-1-AP, Proteintech) and 50 µl of

Protein A/G plus agarose (20423, Pierce). Immunoprecipitates and total lysates were subjected to immunoblotting with the indicated antibodies.

Mass Spectrometry

293FT cells in 10 cm diameter plates were transiently transfected with 2 µg of pQCXIN encoding Flag-JAK2 by using linear Polyethylenimine (23966, Polysciences Inc.). Forty eight hours after transfection, DME-HG supplemented with 10% FBS medium was replaced by serum free medium. The cells was stimulated by 100 nM PMA for 20 min at 37⁰ C, and lysed in 1 ml RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate, and 0.1% SDS) supplemented with protease and phosphatase inhibitors (PhosSTOP, Roche Life Science). Total lysates containing 4-5 mg proteins were incubated with mouse IgG agarose (A0919, Sigma) for 1 hour at 4⁰ C. Flag-tagged JAK2 proteins were purified from the supernatant with Flag M2 affinity gel (A2220, Sigma), and the protein–antibody mixture was eluted with Flag peptide (F4799, Sigma) according to the manufacturer's protocols. Proteins were separated on a 4-12% Bis-Tris SDS-PAGE (Invitrogen) and visualized using Gelcode (24590, Pierce). Visible bands were excised and subjected to MALDI-reTOF-MS/MS analysis to identify phosphorylation sites.

3D Matrigel Assays

Cells were cultured in growth-factor-reduced reconstituted basement membrane (Matrigel, Corning) as described previously (Debnath et al., 2003). Images were taken with an inverted microscope.

PKC Activity Assay

To measure total PKC activity, protein lysates were collected and quantified with a Bradford assay. 0.5 µg of lysates were used for each assay. PKC activity was determined by using an immobilized peptide substrate and a sandwich ELISA assay according to the manufacturer's protocol (ADI-EKS-420A, Enzo). Absorbance (450 nm) was measured and normalized to control cells.

In Vitro Kinase Assays

Phosphorylation of JAK2 (TP320503, Origene) by PKC α (P2232, Invitrogen) was assayed *in vitro* using 100 ng of each purified protein in 20 μ l protein kinase assay buffer (PV3189, Invitrogen) containing 10 μ Ci [γ - ^{32} P] ATP, supplemented or not with 100 nM PMA or 100 nM GÖ6976. The reactions were performed for 10 min at 25 $^{\circ}$ C and terminated by adding SDS sample buffer and boiling for 10 min. Reaction products were analyzed by SDS-PAGE and autoradiography.

To assess the ability of wild type or phosphorylation site mutant Jak 2 to induce phosphorylation of STAT3, JAK2-deficient human fibrosarcoma γ 2A cells (a gift from Dr. Ross Levine; MSKCC) were transfected with constructs encoding wild type JAK2 or the indicated phosphosite-mutants. After 48 hours, cells were lysed in ice-cold HEPES lysis buffer (50 mM HEPES pH7.4, 150mM NaCl, 0.5% Triton X-100) containing 30 mg/ml RNase A, 30 mg/ml DNase and protease inhibitors (Protease Inhibitor Cocktail Set III from Calbiochem, 1:200) and phosphatase inhibitors (20 mM NaF, 20 mM glycerophosphate, 0.3 mM sodium vanadate). Lysates were pre-cleared by centrifugation and aliquots consisting of 5 μ l were incubated with GST-STAT3 (10 μ g/ml) in kinase buffer (50 mM Tris HCl [pH 7.5], 100 mM KCl, 50 mM MgCl $_2$, 1 mM Na $_3$ VO $_4$, 1 mM PMSF, 1 mM DTT and 5 mM ATP) at RT for 1 hour. The reactions were terminated with SDS-PAGE sample buffer and immediately boiled for 6 minutes. The mixtures were subjected to SDS PAGE and immunoblotting as indicated.

MTT assay

Control and TM4SF1-silenced 4T1 cells were plated at 1 x 10 ^{3} per well in 96 wells plates for 24 hours. After 24 hours, cells were incubated in 0.5 mg/ml MTT (3-{4,5-dimethylthiazol-2-yl}-2,5-diphenyl tetrazolium bromide; Invitrogen) for 1h at 37 $^{\circ}$ C. MTT crystals were dissolved in isopropanol and absorbance was measured in a plate reader at 570 nM.

Propidium Iodide Staining and Cell Cycle Analysis

Control and TM4SF1-silenced 4T1 cells were cultured for 2 days. After 2 days, single cells suspensions were prepared using Accutase (Innovative Cell Tec.),

washed in PBS, and then fixed in 70% ethanol. Cells were incubated in 2 mg/ml RNAase A and 0.1 mg/ml PI (Sigma) for 45 min at room temperature, washed, and analyzed by flow cytometry.

Wound Scratch Assay

Control and TM4SF1-silenced 4T1 cells were seeded at a density of 150,000 cells per 35 mm cell culture dish. After 48 hours, a scratch was made with a 10 μ l pipette tip in a confluent area of the cell culture dish. Photographs of a selected area of each scratch were taken at 0 and 24 hours after scratching. The wound area was measured using ImageJ software (NIH).

FACS Analysis

Control and TM4SF1-silenced ErbB2-transformed cells were detached with Accutase and washed in blocking solution (HBSS supplemented with 10% FBS). Cell suspensions were incubated with the indicated antibodies for 45 minutes at 4^o C and analysed by FACS.

Bioinformatics Analysis

For each primary breast cancer data set, the gene expression values of each sample were log₂ scaled and then mean centered. Subsequently z-score transformation was performed to normalize gene expression across all samples in each data set. DNA microarray results were analysed with R Console. Class comparison between triplicate 4TO7-TM4SF1 and 4TO7-Vector samples was performed to identify gene expression changes of ≥ 1.5 -fold associated with TM4SF1 overexpression. As a result, we obtained a list of 67 probe sets. This gene set was used to classify samples from human primary breast cancer data sets (MSK82, EMC192, EMC286, and NKI295) into groups with gene expression patterns resembling those of TM4SF1-overexpressing or control 4TO7 cells. Univariate Cox proportional hazards regression was used to identify genes significantly associated (p-value < 0.05) with metastasis-free survival (MFS) and to estimate the regression coefficients among these differentially expressed genes. Risk scores of significant genes were determined by the methods described previously (Gao et al., 2012) while tissue immunohistochemical

staining values were used to present the risk values of DDR1, TM4SF1 and P-STAT3 across all samples in the tissue immunohistochemical staining data sets (TMA8, TMA9, TMA10, TMA11). All immunohistochemical staining data sets were selected as testing sets. For 8-gene-signature, patient samples were assigned to a low risk group if their risk indexes valued less than the 65th percentile of the risk score values and to a high risk group otherwise. For P-STAT3, risk cutoff was defined as the 65th percentile of the risk score values (TMA8 only), as the 70th percentile (TMA9, TMA10, TMA11) for DDR1 and as the 40th (TMA9, TMA10, TMA11) and 60th (TMA8-9 matched) percentile for TM4SF1. Kaplan-Meier survival analysis and log rank tests were performed using the survival package 2.37 in R 3.03 to compare the survival curves of low and high-risk groups. Comparisons between two groups of TMA8-9 matched sets were performed using paired two-sided t test, while unpaired two-sided t test was used between the metastasis-free and metastasis groups ($p < 0.05$ was considered significant). The online resource for breast cancer DNA microarray analysis is available at <http://kmplot.com/analysis> (Gyorffy et al., 2013).

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. TM4SF1 is Required for Metastatic Reactivation in the Lung and Predicts Relapse in Patients, Related to Figure 1

(A) Schematic representation of the tail-vein cDNA screen.

(B and C) MDA-MB231 cells infected with lentiviral vectors encoding a control shRNA (sh-Co.) or two shRNAs targeting TM4SF1 (sh #1 and #5) were subjected to immunoblotting with the indicated antibodies (B), and were inoculated intravenously into nude mice. Lung metastasis was measured by bioluminescent imaging. The graph shows the normalized photon flux at the indicated times (C; left) and the panels show representative images (C, right). Note that the scale for normalized photon flux is logarithmic. *P* values for the last time points were derived from a two-tailed, unpaired Student's *t* test.

(D-G) ErbB2-transformed cells were infected with lentiviral vectors encoding a control shRNA (sh-Control) or shRNAs targeting TM4SF1 (sh-TM4 #1 and #5), subjected to immunoblotting with the indicated antibodies (D), and inoculated intravenously into nude mice. Lung metastasis was measured by bioluminescent imaging. The graph shows the normalized photon flux at the indicated times (E, left) and the panels show representative images (E, right). Mice were sacrificed at the indicated times and lung sections were processed for immunofluorescent detection of GFP (Green; tumor cells) or PECAM-1 (Red; endothelial cells) followed by counterstaining with DAPI. *P* values for the last time points were derived from a two-tailed, unpaired Student's *t* test. White arrows point to solitary tumor cells (F). The graphs show the number of GFP-positive solitary tumor cells per microscopic field (blue bar) and of metastatic lesions (mets) per lung section (red bar) at the indicated times (G).

(H and I) 4TO7 cells were infected with a lentiviral vector encoding doxycycline-inducible Myc-tagged human TM4SF1 and treated with the indicated concentrations of doxycyclin for 48 hours. Cell lysates were subjected to immunoblotting with the indicated antibodies (H). Mice were injected with 4TO7-TGL cells expressing doxycycline-inducible Myc-TM4SF1, treated with doxycycline at the indicated times, and sacrificed at various times post-injection. Lung sections were subjected to double staining with antibodies to GFP and TM4SF1 followed by counterstaining with DAPI (I). Note that the antibodies to TM4SF1 recognize both endogenous and recombinant TM4SF1.

(J) Kaplan-Meier analysis of relapse-free survival in the NKI295 (left), or the MSK82, EMC192, and EMC286 combined dataset (right). Patients were divided according to the expression of the 8-gene TM4SF1 signature (red line, positive; blue line, negative).

Figure S2. TM4SF1 Enhances Cancer Stem Cell Traits in Response to Collagen I, Related to Figure 2

(A) Control and TM4SF1-silenced 4T1 cells were cultured in 3D Matrigel. The pictures show representative wells (left) and the graph the mean number of

colonies ($\pm$ SE) per 100 cells from triplicate samples (right). Error bars, mean $\pm$ SE. *P* values of the indicating time points were derived from unpaired Student's *t* test

(B-D) 4T1 cells stably transduced with lentiviral vectors encoding a control sh-RNA (sh-Control) or two sh-RNAs targeting TM4SF1 (sh-TM4 #1 and #5) were subjected to MTT assay (B), PI staining and FACS analysis (C), or wound scratch assay (D). The graph shows the number of viable cells as measured by MTT staining (B). The pictures show representative images of in vitro wound closure at the indicated times (D, left), and the graph shows average wound closure in Arbitrary Units (AU) ($\pm$ SE) (D, right).

(E) Control and TM4SF1-silenced ErbB2 cells were subjected to FACS analysis with antibodies to SCA1 and CD49f (top panels) or CD29 and CD24 (bottom panels).

(F) Control and TM4SF1-silenced ErbB2 cells were subjected to tumor sphere assay in medium with or without 50 μ g/ml collagen I. The pictures show representative images of tumor spheres.

(G) Control and TM4SF1-silenced MDA-MB231 cells were subjected to tumor sphere assay in medium with or without 50 μ g/ml collagen I. The graph shows the number of tumor sphere per 10^3 cells seeded (top), and the pictures show representative images of tumor spheres (bottom). Error bars, mean $\pm$ SE. *P* values of the indicating time points were derived from unpaired Student's *t* test

(H) Equal amounts of total proteins (15 μ g) from ascorbic acid treated (100 μ M, 48 h) Mouse Embryo Fibroblasts (MEF), 4TO7, 4T1, ErbB2, and MDA-MB231 cells were analyzed by immunoblotting with polyclonal antibodies reacting with the α 1 and α 2 chains of both human and mouse collagen I. Pc and Pn correspond to partially processed and co-migrating α chains.

(I) Paraffin-embedded sections of ErbB2 organoids were subjected to immunohistochemical staining with the indicated antibodies followed by counterstaining with hematoxylin (H).

(J) MDA-MB231 cells (5×10^3) were treated with polyclonal antibodies to collagen I ($5 \mu\text{g/ml}$) or control IgGs ($10 \mu\text{g/ml}$) and subjected to tumor sphere formation assay. The graph shows the number of tumor spheres per 10^5 cells (left) and the pictures shows the representative images (right). Error bars, mean $\pm$ SE. *P* values of the indicating time points were derived from unpaired Student's *t* test.

Figure S3. TM4SF1 Mediates Mice and Human Breast Cancer Metastasis to Bone, Related to Figure 3

(A-C) 4T1-TGL cells infected with lentiviral vectors encoding a control shRNA (sh-Co.) or two shRNAs targeting TM4SF1 (sh-TM4 #1 and #5) were inoculated in the left ventricle (3×10^5) of syngeneic mice. Bone and brain metastases were monitored by bioluminescent imaging and the panels show representative images (A). Bone metastases were confirmed by staining adjacent sections with H&E or anti-GFP followed by counterstaining with Hematoxylin. The pictures show representative images (B). The graph shows the cumulative photon flux signal of metastases in the hind limbs at day 21 (C, top). The graph shows the photon flux signal of metastases in the brains at day 14 (C, Bottom). Error bars, mean $\pm$ SE. *P* values of the indicating time points were derived from unpaired Student's *t* test.

(D-H) Bone metastatic derivative of MDA-MB231 (BoM-1833) cells infected with lentiviral vectors encoding two shRNAs targeting TM4SF1 (sh-TM4 #5 and #7) or a control shRNA (sh-Co.) were subjected to Q-PCR analysis (D) and immunoblotting (E) and inoculated (1×10^5) in the left ventricle of nude mice. Bone metastasis was measured by bioluminescent imaging. The panels show representative images (F). Bone metastases were confirmed by H&E staining at end point (G). The graph shows the normalized photon flux at the indicated times (H). Error bars, mean $\pm$ SD. *P* values of the indicating time points were derived from unpaired Student's *t* test.

Figure S4. TM4SF1 Combines with DDR1 to Promote Tyrosine Kinase-independent Signaling and Metastasis, Related to Figure 4

(A) 293FT cells were co-transfected with Myc-TM4SF1 in combination with Flag-DDR1, Flag-DDR2, or empty vector, and immunoprecipitated with anti-Flag. Flag immunoprecipitates (top and middle) and total lysates (bottom) were subjected to immunoblotting with the indicated antibodies.

(B) 293FT cells were co-transfected of Myc-TM4SF1 in combination with Flag-DDR1 or empty vector, treated with or without 50 μ g/ml collagen I for 6 hours, and lysed in Glucopyranoside (OGP), Brij96, CHAPS, or RIPA lysis buffer. Equal amounts of total proteins were subjected to immunoprecipitation and immunoblotting with the indicated antibodies. Note that stimulation with collagen I render a large part of DDR1 insoluble in Brij96, yet the soluble part becomes with a larger fraction of TM4SF1.

(C) Equal amounts of total proteins from ErbB2 cells were immunoprecipitated by anti-DDR1 or control IgGs. Immunoprecipitates (top) and total lysates (bottom) were subjected to immunoblotting with the indicated antibodies.

(D) MCF-7 cells were transfected of Flag-TM4SF1 or empty vector and immunoprecipitated with anti-Flag. Flag immunoprecipitates and total lysates containing equal amounts of proteins were subjected to immunoblotting with antibodies to DDR1, the integrin β 1 subunit (ITGB1), the integrin α v subunit (ITGAV), or Flag. Note that that α 1 β 1, α 2 β 1, α 10 β 1 and α 11 β 1 integrins, which bind to native collagen, share the β 1 subunit, whereas α v β 3, which binds to denatured collagen, can be recognized with antibodies to the α v subunit.

(E) 293FT cells were co-transfected of Myc-TM4SF1 in combination with Flag-DDR1, treated with 10 μ g/ml collagen I for the indicated times, and lysed in RIPA buffer. Equal amounts of proteins from the soluble and insoluble fraction were subjected to immunoblotting with the indicated antibodies.

(F) 293FT cells were co-transfected of Myc-TM4SF1 in combination with Flag-DDR1 or empty vector, treated with or without 10 μ g/ml collagen I or collagen IV

for 6 hours, and immunoprecipitated with anti-Flag. Flag immunoprecipitates (top and middle) and total lysates (bottom) were subjected to immunoblotting with the indicated antibodies.

(G) 293FT cells were co-transfected of Myc-TM4SF1, together with Flag-DDR1 or empty vector, treated with 10 μ g/ml collagen I alone or in combination with 10 μ g/ml anti-DDR1 monoclonal antibody 1 for 3 hours as indicated, and immunoprecipitated with anti-Flag. Flag immunoprecipitates and total lysates were subjected to immunoblotting with the indicated antibodies.

(H) Control and TM4SF1-silenced ErbB2 cells were treated with or without 10 μ g/ml collagen I for 30 min. After washing, the cells were fixed and stained with anti-DDR1 followed by counterstaining with DAPI. The graph shows the number of DDR1 clusters per microscopic field. P values were derived from a two-tailed, unpaired Student's t test.

(I) ErbB2 cells were treated with or without 10 μ g/ml collagen I for 30 min, alone or in combination with 10 μ M Y-27632 (ROCK1 inhibitor) or 10 μ M Blebbistatin (non-muscle myosin II ATPase inhibitor) as indicated, and then subjected to staining with anti-DDR1 (red) and anti-phalloidin (green) followed by counterstaining with DAPI. The pictures show representative images.

(J) ErbB2 cells were treated with 10 μ g/ml collagen or I (top) IV (bottom) for the indicated times, and then subjected to staining with anti-DDR1 followed by counterstaining with DAPI. The graphs show the number of DDR1 clusters per area.

(K) 293FT cells were co-transfected with Myc-TM4SF1 in combination with Flag-DDR1 or empty vector, treated with 10 μ g/ml collagen I alone or in combination with 1 μ M Imatinib for 3 hours as indicated, and immunoprecipitated with anti-Flag. Flag immunoprecipitates were subjected to immunoblotting with the indicated antibodies.

(L and M) ErbB2 cells infected with lentiviral vectors encoding a control shRNA (sh-Co.) or two shRNAs targeting DDR1 (sh #6 and #7) were subjected to

immunoblotting (L) and tumor sphere assay in medium supplemented with or without 50 $\mu\text{g/ml}$ collagen I (M). The graph shows the number of tumor sphere formed per 10^3 cells seeded. Error bars, mean $\pm$ SE. *P* values were derived from unpaired Student's *t* test.

(N and O) ErbB2 cells infected with lentiviral vectors encoding a control shRNA (sh-Co.) or two shRNAs targeting DDR2 (sh #3 and #4) were subjected to Q-PCR analysis (N) and tumor sphere assay in medium supplemented with or without 50 $\mu\text{g/ml}$ collagen I (O). The graph shows the number of tumor sphere formed per 10^3 cells seeded. Error bars, mean $\pm$ SE. *P* values were derived from unpaired Student's *t* test.

(P) ErbB2 cells were subjected to tumor sphere assay in medium supplemented with or without 50 $\mu\text{g/ml}$ collagen I in the presence of vehicle (DMSO), 0.1 μM Imatinib, or 0.1 μM Ponatinib. The graph shows the number of tumor sphere formed per 10^3 cells seeded. *P* values were derived from unpaired Student's *t* test

(Q) ErbB2-TGL cells transduced with DDR1-P529* or empty vector were analyzed by immunoblotting with the indicated antibodies.

(R) ErbB2 cells were transfected with Flag-DDR1 P529* or empty vector, treated with 10 $\mu\text{g/ml}$ collagen I for the indicated times, and immunoprecipitated with anti-DDR1. Immunoprecipitates were subjected to immunoblotting with the indicated antibodies.

(S) ErbB2 cells transduced with Flag-DDR1 P529* or empty vector were subjected to tumor sphere assay in medium supplemented with or without 50 $\mu\text{g/ml}$ collagen I. The graph shows the number of tumor spheres formed per 10^3 cells seeded. *P* values were derived from unpaired Student's *t* test.

Figure S5. Non-canonical DDR1 Signaling Promotes PKC α -dependent Activation of JAK2 and STAT3, Related to Figure 5

(A) 293FT cells were co-transfected of Flag-syntenin 2, together with Myc-TM4SF1, Myc-TM4SF1 C202G, or empty vector, and immunoprecipitated with

anti-Flag. Flag immunoprecipitates and total lysates were subjected to immunoblotting with the indicated antibodies.

(B) 293FT cells were co-transfected of HA-syntenin 2, together with Flag-DDR1, Flag-DDR2, or empty vector, and immunoprecipitated with anti-Flag. Total lysates and Flag immunoprecipitates were subjected to immunoblotting with the indicated antibodies.

(C) 293FT cells were co-transfected of Myc-TM4SF1, together with the indicated constructs, and immunoprecipitated with anti-Flag. Total lysates and Flag immunoprecipitates were subjected to immunoblotting with the indicated antibodies.

(D) TM4SF1-silenced ErbB2-transformed cells were serum starved for 12 hours and then treated with 50 µg/ml of collagen I for 6 hours, equal amounts of total proteins were immunoprecipitated by anti-Syntenin2 antibody. Immunoprecipitates were subjected to immunoblotting with anti-DDR1 or anti-Syntenin2 antibodies and total lysates with anti-DDR1 antibody.

(E) 4TO7 cells transduced with TM4SF1, TM4SF1 C202G, syntenin 2, or empty vector were treated with or without 10 µg/ml collagen I as indicated. Total lysates were subjected to PKC kinase assay.

(F) 4TO7 cells transduced with TM4SF1, TM4SF1 C202G, syntenin 2, or empty vector were subjected to immunoblotting with the indicated antibodies.

(G) ErbB2 cells were treated with 10 µg/ml collagen I for the indicated times. Anti-DDR1 immunoprecipitates and total lysates were subjected to immunoblotting with the indicated antibodies.

(H) ErbB2 cells were transfected with a SMART pool of siRNAs targeting TM4SF1 or a control non-targeting pool, and subjected to Q-PCR analysis. *P* values were derived from unpaired Student's *t* test.

(I) Model of TM4SF1-mediated recruitment of syntenin 2 and PKC.

(J) 293FT cells were co-transfected of Myc-PKC α , together with Flag-syntenin 2, Flag-TM4SF1, or empty vector, and immunoprecipitated with anti-Flag. Total

lysates (left) and Flag immunoprecipitates (right) were subjected to immunoblotting with the indicated antibodies.

(K) 293FT cells were co-transfected of Myc-JAK2, together with Flag-TM4SF1, Flag-DDR1, Flag-syntenin 2, Flag-PKC α , or empty vector, and immunoprecipitated with anti-Flag. Total lysates (left) and Flag immunoprecipitates (right) were subjected to immunoblotting with the indicated antibodies.

(L) Purified PKC α and JAK2 were incubated with PMA (PKC activator) or Gö6976 (PKC inhibitor) as indicated and subjected to in vitro kinase assay. Proteins were separated by SDS-PAGE and stained with Colloidal Blue. Asterisks point to the purified JAK2 (upper band) and PKC α proteins (lower band).

(M) 293FT cells were transfected of Flag-JAK2, treated with or without PMA, and immunoprecipitated with anti-Flag. Immunoprecipitates were subjected to mass spectrometry. Schematic representation of JAK2 showing JAK homology (JH) domains 1–7, including the kinase domain (JH1) and pseudokinase domain (JH2), SH2-like domains, and FERM domain. Red arrows point to the two phosphorylation sites identified by mass spectrometry (T174 and S518). Black arrow indicates the activating mutation commonly found in myeloproliferative disorders (V617F).

(N) Evolutionary conservation of the two PKC phosphorylation sites of JAK2.

Figure S6. Non-Canonical DDR1 Signaling Promotes Self-renewal in Vitro and Lung Metastasis in Vivo, Related to Figure 6

(A and B) ErbB2 cells were treated with or without 10 μ g/ml collagen I for 30 min alone or in combination with Gö6976 (PKC inhibitor), Bisindolylmaleimide I (Bis I, PKC inhibitor), Fedratinib and AZD1480 (Jak2 inhibitors), Ponatinib, Imatinib, or Stattic (STAT3 inhibitor) as indicated, and then subjected to staining with anti-DDR1 followed by counterstaining with DAPI. The panels show representative

images (A) and the graph the number of DDR1 clusters per surface area (B). *P* values were derived from unpaired Student's *t* test.

(C) 4T1-TGL cells transduced with dominant negative PKC α , PKC β II, PKC δ , or PKC ζ , or empty vector were subjected to immunoblotting with the indicated antibodies.

(D) 4T1 cells were infected with lentiviral vectors encoding a control shRNA (sh-Control) or two shRNAs targeting STAT3 (1 and 5) and subjected to immunoblotting with the indicated antibodies.

(E-G) ErbB2 cells infected with lentiviral vectors encoding two shRNAs targeting STAT3 (sh #1 and #5) or a control shRNA (sh-Control) were subjected to immunoblotting with the indicated antibodies (E) and inoculated intravenously into nude mice. Lung metastasis was measured by bioluminescent imaging. The panels show representative images at week 5 (F). The graph shows the normalized photon flux at week 5 (G).

(H) ErbB2 cells were infected with lentiviral vectors encoding a control shRNA (sh-Co.) or two shRNAs targeting DDR1 (sh #6 and #7) or syntenin 2 (Synt2; sh# 1 and Sh# 3) were subjected to 3D Matrigel culture. The graph shows the mean number of organoids per 100 cells from triplicate samples (right). Error bars, mean $\pm$ SE. *P* values of the indicating time points were derived from unpaired Student's *t* test.

(I) ErbB2-TGL cells were infected with lentiviral vectors encoding a control shRNA (sh-Control) or DDR1 (sh #6 and #7) and syntenin 2 (Synt2; sh# 1 and Sh# 3) and STAT3 (Sh #1 and #3), and inoculated intravenously into nude mice. End point lung sections were processed for immunofluorescent detection of GFP (tumor cells). The graphs show the number of GFP-positive solitary tumor cells per microscopic field (blue bars) and of metastatic lesions per lung section (red bars). Error bars, mean $\pm$ SE. *P* values of the indicating time points were derived from unpaired Student's *t* test

(J-L) ErbB2-TGL cells were infected with lentiviral vectors encoding a control shRNA (sh-Co.) or two shRNAs targeting DDR1 (sh #6 and #7) and syntenin 2 (

Synt2; sh# 1 and Sh# 5) and inoculated intracardiacally into immunodeficient mice. Bone and brain metastases were detected by bioluminescent imaging and represented graphically (J). Bone metastases were confirmed by X-ray radiography (K). Mice were sacrificed at the indicated times and bone (hind limbs) and brain sections were processed for immunohistochemical detection of GFP (tumor cells). The graphs show the number of GFP-positive solitary tumor cells per microscopic field (blue bars) and of metastatic lesions per bone and brain sections (red bars) at the end point of the experiment (L). Error bars, mean $\pm$ SE. *P* values of the indicating time point were derived from unpaired Student's *t* test.

Figure S7. Fibrillar Collagen I and Activated STAT3 Define Metastatic Niches, Related to Figure 7

(A) Paraffin-embedded sections of mammary fat pads from 13 week-old MMTV-*Neu*(YD) mice were subjected to picrosirius red staining. The pictures show representative images of collagen fibril formation with picrosirius red staining, visualized under bright (parallel) and polarized light (orthogonal) in MIN lesions or adenocarcinomas. Green arrows point to the scant collagen fibrils surrounding MIN lesions, and the blue arrow the abundant collagen fibers around a normal duct.

(B) Paraffin-embedded sections of lungs from 5 month-old MMTV-*Neu*(YD) mice were subjected to immunohistochemical staining with the indicated antibodies, to Masson trichrome staining, or to picrosirius red staining. The top left image and mid row shows the representative images of micrometastatic lesions (red arrows) and a macrometastasis with a necrotic core (asterisk) surrounded by collagen I but not collagen IV. The top right image shows the absence of heavily crosslinked collagen fibers stained by Masson trichrome (blue) but the presence of collagen fibrils detected by picrosirius red staining (bottom row), visualized under bright (parallel, red) and polarized light (orthogonal, orange).

(C) Paraffin-embedded sections of mammary fat pads from 12 and 20 week-old MMTV-*Neu*(YD) mice and lung metastases from 20 week-old MMTV-*Neu*(YD)

mice were subjected to immunohistochemical staining with antibodies to DDR1, TM4SF1, or P-STAT3, followed by counterstaining with Hematoxylin. The pictures show representative images of DDR1 (left panels), TM4SF1 (middle panels), and P-STAT3 (right panels) staining in MIN lesions, adenocarcinomas, micrometastatic lesions (red arrows), and macrometastatic lesions (green arrows).

(D) Lung sections from mice that had been injected 14 days earlier with ErbB2-TGL cells were subjected to immunofluorescence staining with the indicated antibodies followed by counter staining with DAPI. Representative image of a micrometastatic lesion.

(E) TMAs from primary breast tumors of MSKCC patients (n=147) were subjected to immunohistochemical staining with anti-P-STAT3 followed by counterstaining with Hematoxylin. Patients were divided according to the level of P-STAT3 in their primary tumors and their metastasis-free survival data were subjected to Kaplan-Meier analysis.

SUPPLEMENTAL TABLES

Table S1

**Percentage of ER, PR, HER2 and Triple Negative Cases in TMA8 and TMA9,
Related to Figure1**

	ER %	PR%	ER/PR%	HER2%	ER/PR/HER2%	Triple negative %
TMA9 (Breast Primary n= 51	52.94	39.21	35.29	15.66	7.84	33.33
TMA 8 (Lung Met n= 80)	37.5	15	15	23.75	1.25	32.5

Table S2. Multivariate Analysis of the Association of the 8-gene Signature with Overall Relapse, Related to Figure 1

4TO7-TM4SF1 8-gene signature			
Dataset	Variable	p-value	HR(95% CI)
NKI295	8-gene (low risk)	0.00354	0.52(0.34,0.81)
	ER Positive	0.31	1.29(0.79,2.08)
	Grade Poorly diff	0.69	1.1(0.68,1.78)
	Grade Well diff	0.37	0.72(0.35,1.47)
	NIH_Risk Low and Med	0.80	1.1(0.52,2.3)
	70_genes Poor	7.63E-05	3.41(1.86,6.26)
	size >2cm	0.028	1.67(1.06,2.63)
EMC286	Variable	p-value	HR(95% CI)
	8-gene (low risk)	0.000130	0.44(0.29,0.67)
	ER Positive	0.3099	1.26(0.81,1.98)
MSK82	Variable	p-value	HR(95% CI)
	8-gene (low risk)	0.00221	0.25(0.1,0.61)
	size >=2cm	0.500	0.68(0.22,2.08)
	Lymph Nodes Positive	0.188	1.81(0.75,4.38)
	HER2 Positive	0.035	0.27(0.08,0.91)

Multivariate analysis with the Cox proportional hazards regression model was used to assess the ability of the 8-gene signature to predict overall survival independently of commonly used pathological and clinical parameters.

Table S3. Frequency of Tumor-initiating Cells (T-ICs) in Primary and Secondary Transplantation Experiments, Related to Figure 2

Primary Tumor-Initiating Capacity				
ErbB2 cells	No. of Cells	Tumor Incidence	T-IC Frequency	Latency (Days)
Sh-Control	100	10/10	1	22.4±3.0
	1000	10/10		14±0
	10000	10/10		7.7±2.2
Sh-TM4SF1 #1	100	4/10	1/200	35±9.3
	1000	10/10		16.8±3.6
	10000	10/10		11.2±3.6
Sh-TM4SF1 #5	100	5/10	1/1000	37.7±7.7
	1000	9/10		14.8±2.3
	10000	10/10		12.6±3.0

Secondary Tumor-initiating capacity				
ErbB2 cells	No. of Cells	Tumor Incidence	T-IC Frequency	Latency (Days)
Sh-Control	100	10/10	1	24.5±3.7
	1000	10/10		21±0
Sh-TM4SF1 #1	100	5/10	1/150	35.7±7.0
	1000	10/10		24.5±5.9
Sh-TM4SF1 #5	100	3/10	1/300	37.8±6.7
	1000	10/10		28.7±6.1

Table S4. Tandem Mass Spectrometry (MS/MS) Analysis of JAK2 Threonine 174 and Serine 518 and Phosphorylated Peptides, Related to Figure S5

VPVTHEpT(174)QEECLGMAVLDMMR														
#	b	b ⁺⁺	b ⁺	b ⁺⁺⁺	b ⁰	b ⁰⁺⁺	Se q.	y	y ⁺⁺	y ⁺	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	100. 0757	50.5 415					V							2 1
2	197. 1285	99.0 679					P	2401. 9896	1201 .498	2384.9 631	1192 .985	2383. 9791	1192.4 932	2 0
3	296. 1969	148. 6021					V	2304. 9369	1152 .972	2287.9 103	1144 .458	2286. 9263	1143.9 668	1 9
4	397. 2445	199. 1259			379.2 340	190. 1206	T	2205. 8684	1103 .437	2188.8 419	1094 .924	2187. 8579	1094.4 326	1 8
5	534. 3035	267. 6554			516.2 929	258. 6501	H	2104. 8208	1052 .914	2087.7 942	1044 .400	2086. 8102	1043.9 087	1 7
6	663. 3461	332. 1767			645.3 355	323. 1714	E	1967. 7619	984. 3846	1950.7 353	975. 8713	1949. 7513	975.37 93	1 6
7	844. 3601	422. 6837			826.3 495	413. 6784	T	1838. 7193	919. 8633	1821.6 927	911. 3500	1820. 7087	910.85 80	1 5
8	972. 4186	486. 7130	955.3 921	478. 1997	954.4 081	477. 7077	Q	1657. 7052	829. 3563	1640.6 787	820. 8430	1639. 6947	820.35 10	1 4
9	1101 .461	551. 2343	1084. 4347	542. 7210	1083. 4507	542. 2290	E	1529. 6467	765. 3270	1512.6 201	756. 8137	1511. 6361	756.32 17	1 3
10	1230 .503	615. 7556	1213. 4773	607. 2423	1212. 4933	606. 7503	E	1400. 6041	700. 8057	1383.5 775	692. 2924	1382. 5935	691.80 04	1 2
11	1333	667. 2601	1316. 4865	658. 7469	1315. 5024	658. 2549	C	1271. 5615	636. 2844	1254.5 349	627. 7711	1253. 5509	627.27 91	1 1
12	1446 .597	723. 8022	1429. 5705	715. 2889	1428. 5865	714. 7969	L	1168. 5523	584. 7798	1151.5 258	576. 2665	1150. 5417	575.77 45	1 0
13	1503 .618	752. 3129	1486. 5920	743. 7996	1485. 6080	743. 3076	G	1055. 4682	528. 2378	1038.4 417	519. 7245	1037. 4577	519.23 25	9
14	1634 .659	817. 8332	1617. 6325	809. 3199	1616. 6485	808. 8279	M	998.4 468	499. 7270	981.42 02	491. 2137	980.4 362	490.72 17	8
15	1705 .696	853. 3517	1688. 6696	844. 8384	1687. 6856	844. 3463	A	867.4 063	434. 2068	850.37 97	425. 6935	849.3 957	425.20 15	7
16	1804 .764	902. 8859	1787. 7380	894. 3726	1786. 7540	893. 8806	V	796.3 692	398. 6882	779.34 26	390. 1750	778.3 586	389.68 29	6
17	1917 .848	959. 4279	1900. 8221	950. 9147	1899. 8381	950. 4227	L	697.3 008	349. 1540	680.27 42	340. 6407	679.2 902	340.14 87	5
18	2032 .875	1016 .941	2015. 8490	1008 .428	2014. 8650	1007 .936	D	584.2 167	292. 6120	567.19 01	284. 0987	566.2 061	283.60 67	4
19	2179 .911	1090 .459	2162. 8844	1081 .945	2161. 9004	1081 .453	M	469.1 898	235. 0985	452.16 32	226. 5852			3
20	2326 .946	1163 .976	2309. 9198	1155 .463	2308. 9358	1154 .971	M	322.1 544	161. 5808	305.12 78	153. 0675			2
21							R	175.1 190	88.0 631	158.09 24	79.5 498			1

TNGIpS(518)DVQISPTLQR														
#	b	b ⁺⁺	b [*]	b ⁺⁺⁺	b ⁰	b ⁰⁺⁺	Se q.	y	y ⁺⁺	y [*]	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	102.0 550	51.5 311			84.04 44	42.5 258	T							15
2	217.0 819	109. 0446	200.0 553	100. 5313	199.0 713	100. 0393	N	1510.7 860	755. 8966	1493. 7594	747. 3834	1492 .775 4	746. 8914	14
3	274.1 034	137. 5553	257.0 768	129. 0420	256.0 928	128. 5500	G	1395.7 591	698. 3832	1378. 7325	689. 8699	1377 .748 5	689. 3779	13
4	387.1 874	194. 0974	370.1 609	185. 5841	369.1 769	185. 0921	I	1338.7 376	669. 8724	1321. 7110	661. 3592	1320 .727 0	660. 8672	12
5	456.2 089	228. 6081	439.1 823	220. 0948	438.1 983	219. 6028	S	1225.6 535	613. 3304	1208. 6270	604. 8171	1207 .643 0	604. 3251	11
6	571.2 358	286. 1216	554.2 093	277. 6083	553.2 253	277. 1163	D	1156.6 321	578. 8197	1139. 6055	570. 3064	1138 .621 5	569. 8144	10
7	670.3 042	335. 6558	653.2 777	327. 1425	652.2 937	326. 6505	V	1041.6 051	521. 3062	1024. 5786	512. 7929	1023 .594 6	512. 3009	9
8	798.3 628	399. 6850	781.3 363	391. 1718	780.3 523	390. 6798	Q	942.53 67	471. 7720	925.5 102	463. 2587	924. 5261	462. 7667	8
9	911.4 469	456. 2271	894.4 203	447. 7138	893.4 363	447. 2218	I	814.47 81	407. 7427	797.4 516	399. 2294	796. 4676	398. 7374	7
10	998.4 789	499. 7431	981.4 524	491. 2298	980.4 684	490. 7378	S	701.39 41	351. 2007	684.3 675	342. 6874	683. 3835	342. 1954	6
11	1095. 5317	548. 2695	1078. 5051	539. 7562	1077. 5211	539. 2642	P	614.36 20	307. 6847	597.3 355	299. 1714	596. 3515	298. 6794	5
12	1196. 5794	598. 7933	1179. 5528	590. 2800	1178. 5688	589. 7880	T	517.30 93	259. 1583	500.2 827	250. 6450	499. 2987	250. 1530	4
13	1309. 6634	655. 3353	1292. 6369	646. 8221	1291. 6529	646. 3301	L	416.26 16	208. 6344	399.2 350	200. 1212			3
14	1437. 7220	719. 3646	1420. 6955	710. 8514	1419. 7114	710. 3594	Q	303.17 75	152. 0924	286.1 510	143. 5791			2
15							R	175.11 90	88.0 631	158.0 924	79.5 498			1

Tryptic peptides at experimental masses 2500.044 (VPVTHEpTQEECLGMAVLDMMR) and 1708.8029 (TNGIpSDVQISPTLQR) were analyzed. Under each spectrum, matching y ions (C-terminal fragments) or b ions (N-terminal fragments) are colored in red within the tables. Signature ions with increase in mass difference (97.9769 amu from neutral loss of phosphate) confirmed phosphorylation at Threonine 174 for peptide encompassing 168-188 and Serine 518 for peptide encompassing 514-528.

References:

- Aslakson, C.J., and Miller, F.R. (1992). Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res.* 52, 1399-1405.
- Carafoli, F., Mayer, M.C., Shiraishi, K., Pecheva, M.A., Chan, L.Y., Nan, R., Leitinger, B., and Hohenester, E. (2012). Structure of the discoidin domain receptor 1 extracellular region bound to an inhibitory Fab fragment reveals features important for signaling. *Structure* 20, 688-697.
- Dankort, D., Jeyabalan, N., Jones, N., Dumont, D.J., and Muller, W.J. (2001). Multiple ErbB-2/Neu Phosphorylation Sites Mediate Transformation through Distinct Effector Proteins. *J. Biol. Chem.* 276, 38921-38928.
- Das, S., Ongusaha, P.P., Yang, Y.S., Park, J.M., Aaronson, S.A., and Lee, S.W. (2006). Discoidin domain receptor 1 receptor tyrosine kinase induces cyclooxygenase-2 and promotes chemoresistance through nuclear factor-kappaB pathway activation. *Cancer Res.* 66, 8123-8130.
- Debnath, J., Muthuswamy, S.K., and Brugge, J.S. (2003). Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* 30, 256-268.
- Gao, H., Chakraborty, G., Lee-Lim, A.P., Mavrikis, K.J., Wendel, H.G., and Giancotti, F.G. (2014). Forward genetic screens in mice uncover mediators and suppressors of metastatic reactivation. *Proc. Natl. Acad. Sci. U.S.A* 111, 16532-16537.
- Gao, H., Chakraborty, G., Lee-Lim, A.P., Mo, Q., Decker, M., Vonica, A., Shen, R., Brogi, E., Brivanlou, A.H., and Giancotti, F.G. (2012). The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. *Cell* 150, 764-779.
- Guo, W., Pylayeva, Y., Pepe, A., Yoshioka, T., Muller, W.J., Inghirami, G., and Giancotti, F.G. (2006). Beta 4 integrin amplifies ErbB2 signaling to promote mammary tumorigenesis. *Cell* 126, 489-502.
- Gyorffy, B., Surowiak, P., Budczies, J., and Lanczky, A. (2013). Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PloS One* 8, e82241.
- Ponomarev, V., Doubrovin, M., Serganova, I., Vider, J., Shavrin, A., Beresten, T., Ivanova, A., Ageyeva, L., Tourkova, V., Balatoni, J., *et al.* (2004). A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Euro. J. Nuc. Med. Mol. Imag.* 31, 740-751.
- Pylayeva, Y., Gillen, K.M., Gerald, W., Beggs, H.E., Reichardt, L.F., and Giancotti, F.G. (2009). Ras- and PI3K-dependent breast tumorigenesis in mice and humans requires focal adhesion kinase signaling. *J. Clin. Invest.* 119, 252-266.
- Pylayeva, Y., Guo, W., and Giancotti, F.G. (2007). Analysis of integrin signaling in genetically engineered mouse models of mammary tumor progression. *Met. Enzym.* 426, 439-461.
- Shintani, Y., Fukumoto, Y., Chaika, N., Svoboda, R., Wheelock, M.J., and Johnson, K.R. (2008). Collagen I-mediated up-regulation of N-cadherin requires cooperative signals from integrins and discoidin domain receptor 1. *J. Cell Biol.* 180, 1277-1289.

Simpson, P.T., Shoker, B.S., Barraclough, R., Halliwell, N., Rudland, P.S., Sibson, D.R., and Davies, M.P. (2003). Examination of tumour histopathology and gene expression in a neu/S100A4 transgenic model of metastatic breast cancer. *Int. J. Exp. Path.* 84, 173-184.

Tam, W.L., Lu, H., Buikhuisen, J., Soh, B.S., Lim, E., Reinhardt, F., Wu, Z.J., Krall, J.A., Bieri, B., Guo, W., *et al.* (2013). Protein kinase C alpha is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer cell* 24, 347-364.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80-84.

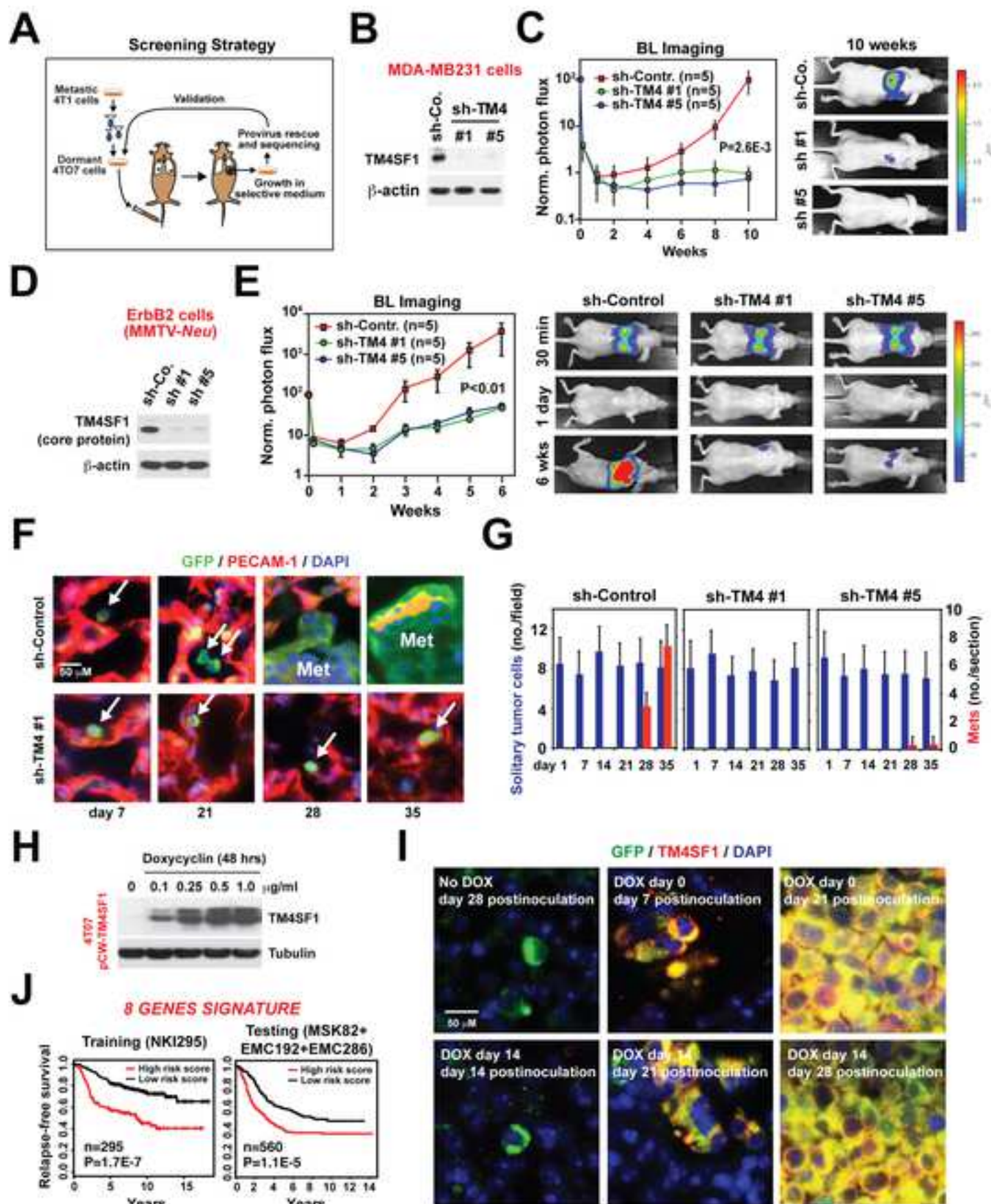


Figure S1

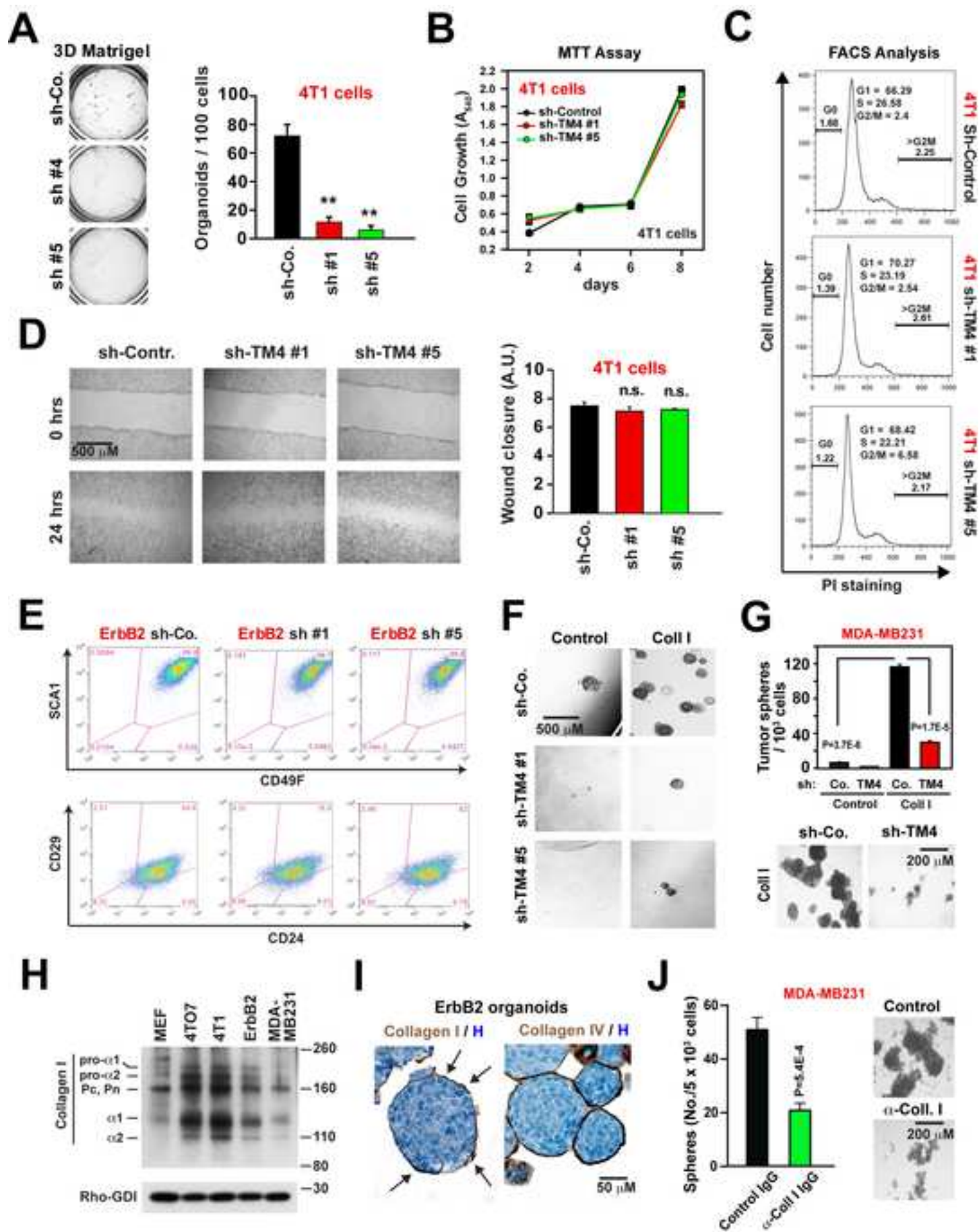


Figure S2

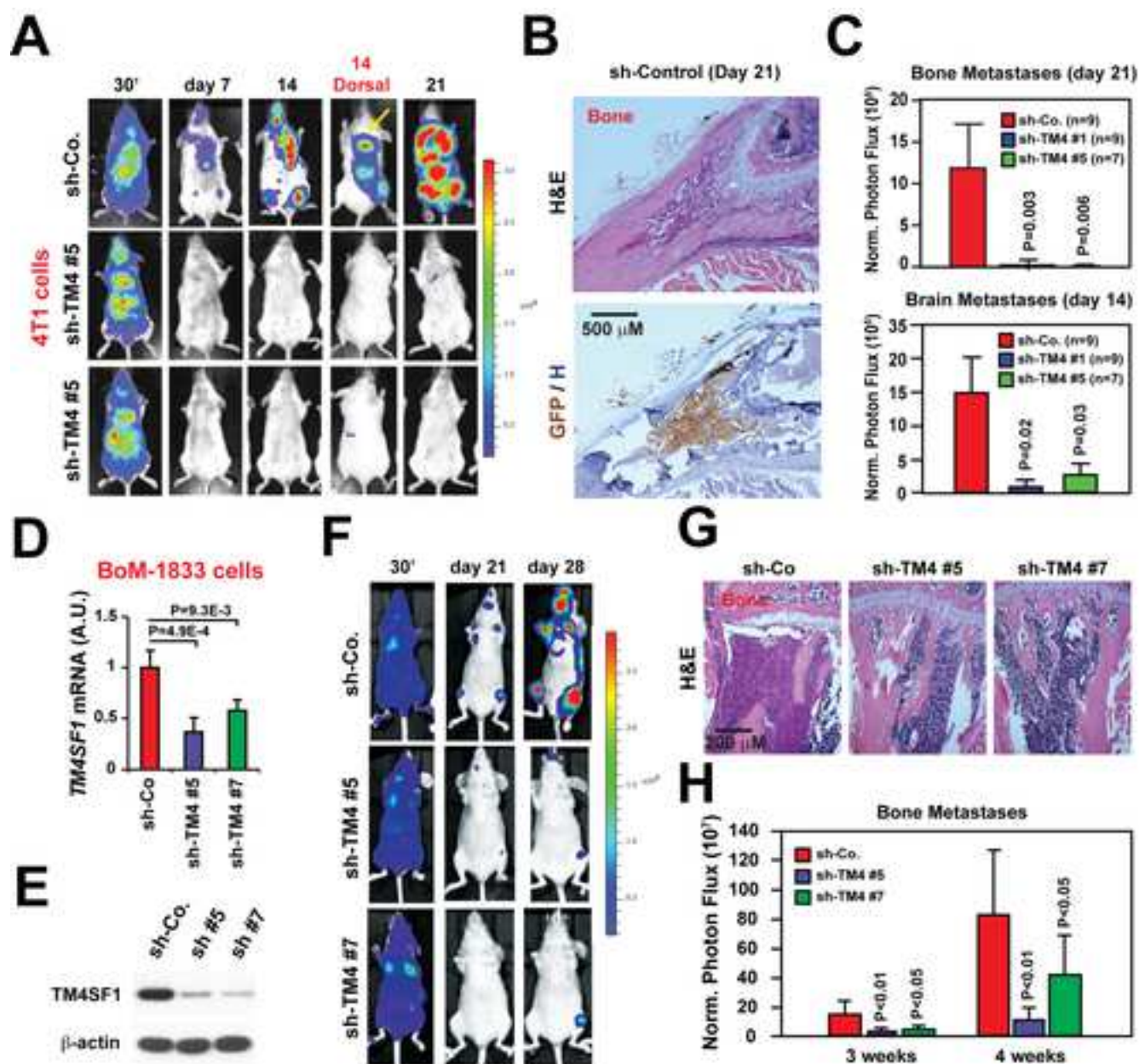


Figure S3

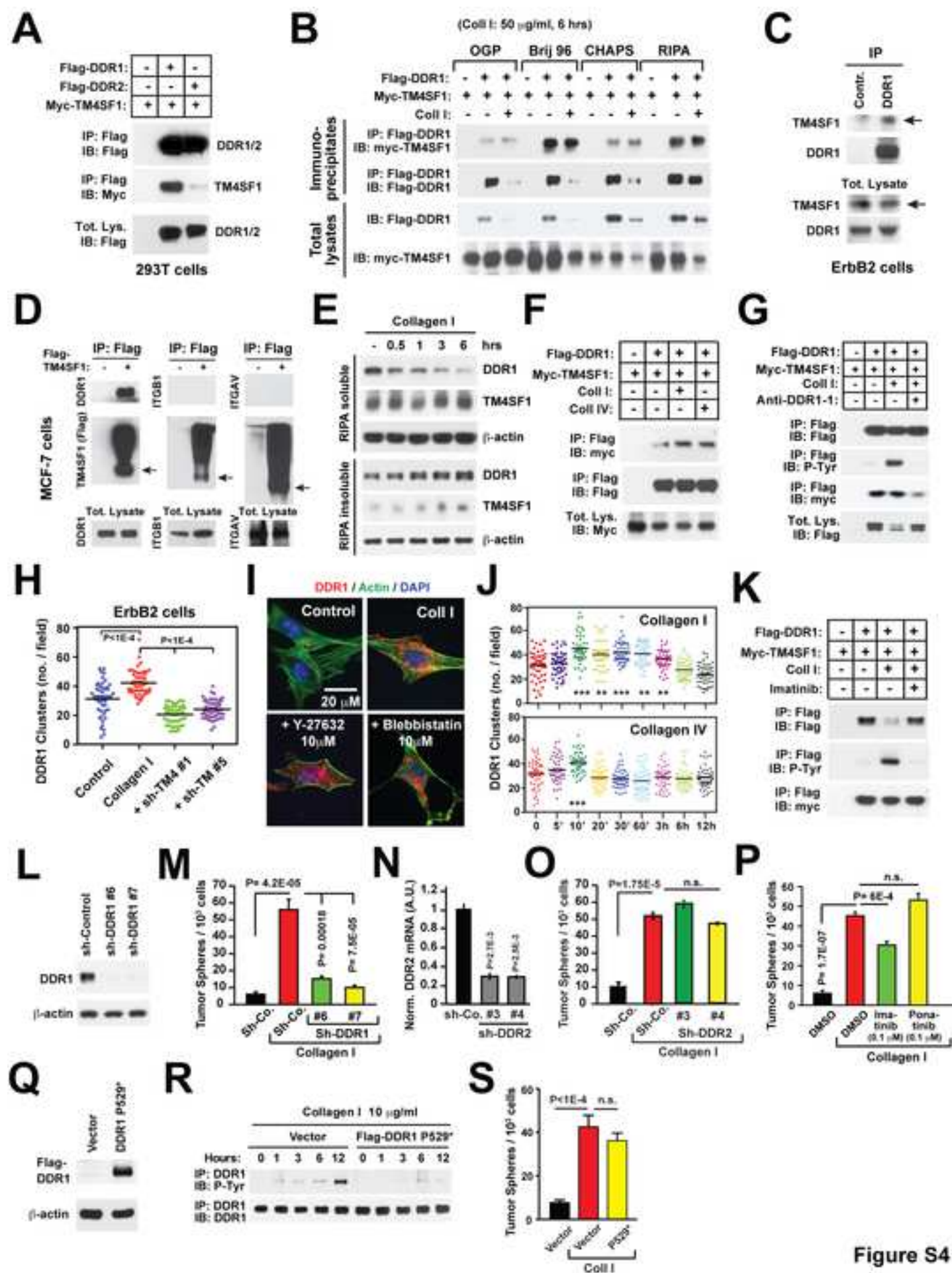


Figure S4



Figure S5

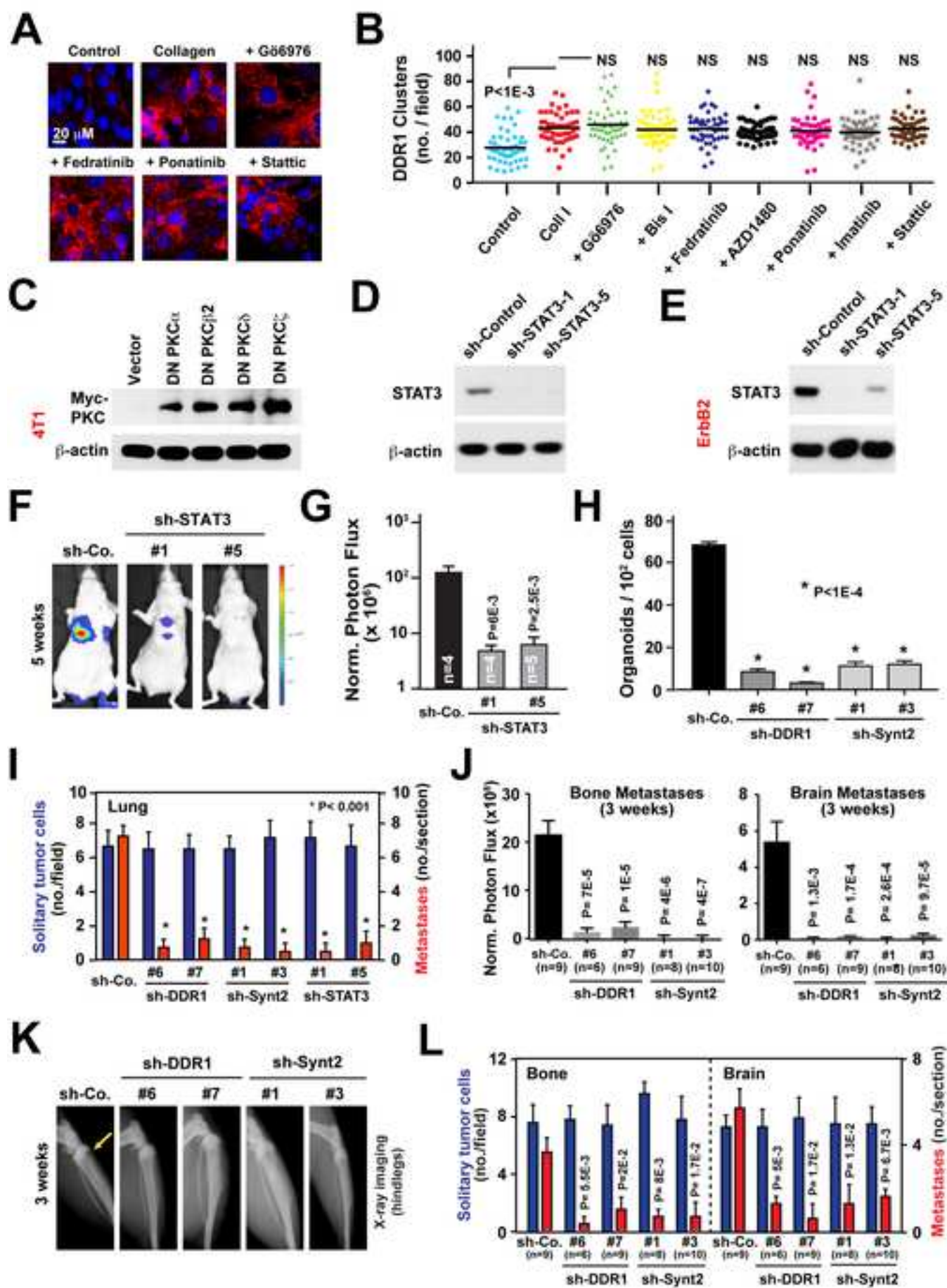


Figure S6

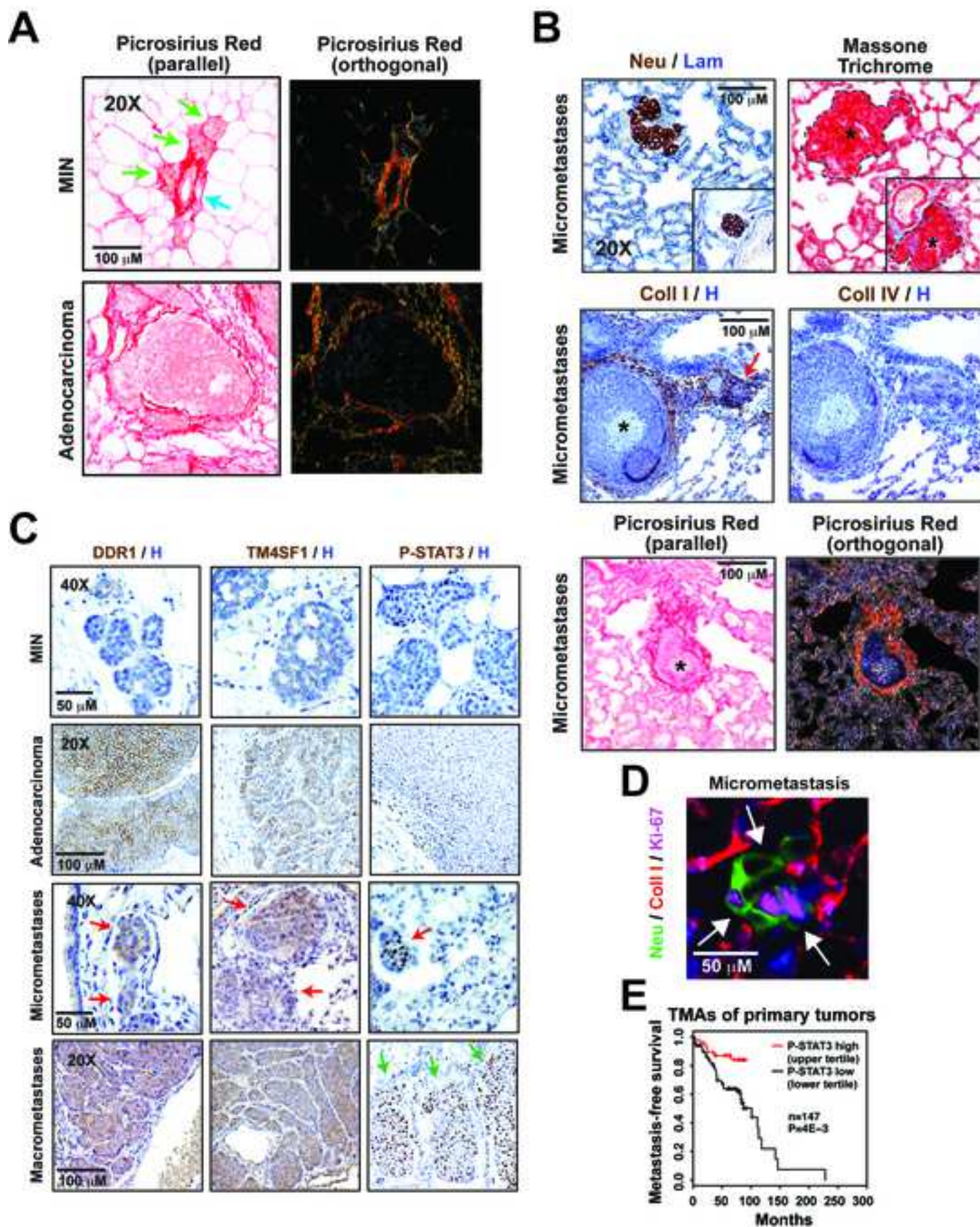


Figure S7