Title

Molecular definition of a metastatic lung cancer state reveals a targetable CD109—Janus kinase—Stat axis

Journal

Nature Medicine, 23(3)

ISSN

1078-8956

Authors

Chuang, Chen-Hua
Greenside, Peyton G
Rogers, Zoë N
et al.

Publication Date

2017-03-01

DOI

10.1038/nm.4285

Peer reviewed

HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2019 April 08.

Published in final edited form as:

Nat Med. 2017 March ; 23(3): 291–300. doi:10.1038/nm.4285.

Molecular definition of a metastatic lung cancer state reveals a targetable CD109–Janus kinase–Stat axis

Chen-Hua Chuang[1], **Peyton G Greenside**[2], **Zoë N Rogers**[1], **Jennifer J Brady**[1], **Dian Yang**[3], **Rosanna K Ma**[1], **Deborah R Caswell**[3], **Shin-Heng Chiou**[1], **Aidan F Winters**[1], **Barbara M Grüner**[1], **Gokul Ramaswami**[1], **Andrew L Spencley**[3], **Kimberly E Kopecky**[4], **Leanne C Sayles**[5], **E Alejandro Sweet-Cordero**[3,5], **Jin Billy Li**[1], **Anshul Kundaje**[1,6], and **Monte M Winslow**[1,3,7]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA.

[2]Biomedical Informatics Training Program, Stanford University School of Medicine, Stanford, California, USA.

[3]Cancer Biology Program, Stanford University School of Medicine, Stanford, California, USA.

[4]Department of Surgery, Stanford University School of Medicine, Stanford, California, USA.

[5]Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA.

[6]Department of Computer Science, Stanford University, Stanford, California, USA.

[7]Department of Pathology, Stanford University School of Medicine, Stanford, California, USA.

Abstract

Lung cancer is the leading cause of cancer deaths worldwide, with the majority of mortality resulting from metastatic spread. However, the molecular mechanism by which cancer cells acquire the ability to disseminate from primary tumors, seed distant organs, and grow into tissue-destructive metastases remains incompletely understood. We combined tumor barcoding in a mouse model of human lung adenocarcinoma with unbiased genomic approaches to identify a transcriptional program that confers metastatic ability and predicts patient survival. Small-scale *in vivo* screening identified several genes, including *Cd109,* that encode novel pro-metastatic factors. We uncovered signaling mediated by Janus kinases (Jaks) and the transcription factor Stat3 as a critical, pharmacologically targetable effector of CD109-driven lung cancer metastasis. In summary, by coupling the systematic genomic analysis of purified cancer cells in distinct

Correspondence should be addressed to M.M.W. (mwinslow@stanford.edu).

AUTHOR CONTRIBUTIONS

C.-H.C. and M.M.W. conceived the study; C.-H.C. designed, conducted, and analyzed the majority of experiments; P.G.G. conducted the majority of bioinformatics analyses; Z.N.R., J.J.B., D.Y., R.K.M., D.R.C., S.-H.C., A.F.W., B.M.G., A.L.S., and K.E.K. conducted experiments, analyzed data and contributed to the Discussion; G.R. and J.B.L. (SNPiR), L.C.S. and E.A.S.-C. (pLKO library and human lung cell lines), and A.K. (bioinformatics) provided crucial reagents and discussion; and M.M.W. and C.-H.C. wrote the manuscript with comments from all authors.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

malignant states from mouse models with extensive human validation, we uncovered several key regulators of metastatic ability, including an actionable pro-metastatic CD109–Jak–Stat3 axis.

Most cancer patients die of complications resulting from metastases, but the molecular and cellular changes that endow cancer cells with the ability to leave the primary tumor, survive during transit through the blood, and establish new tumors in secondary organs remain incompletely understood[1]. Cell-intrinsic alterations and external signals from the tumor microenvironment alter the cancer cell state and enhance the likelihood that a cancer cell will overcome the multiple barriers that limit metastatic spread[2]. A better understanding of the molecular changes that enable cancer cells to overcome the barriers imposed during the metastatic process could aid in the diagnosis, prevention, and treatment of metastatic cancer[1,2].

Much of our understanding of the molecular mechanisms that drive metastatic ability has been generated from experimental systems that relied on the use of cancer cell lines. However, these cell lines are unlikely to maintain all of the molecular features of patients' tumors growing within their unperturbed native environment[3–7]. Conversely, direct molecular analyses of primary tumors and metastases from patients almost always entails the analysis of bulk tumor samples; thus, the molecular changes within the cancer cells themselves can be difficult to glean. Genetically engineered mouse models of metastatic human cancer represent experimental systems with which the natural history of malignant progression can be investigated. In these models, autochthonous primary tumors develop entirely within their natural environment and can evolve to gain metastatic proclivity that is sufficient to produce widespread multi-organ metastatic disease[8–12].

Lung adenocarcinoma is a prevalent type of lung cancer that frequently harbors activating point mutations in the *KRAS* oncogene and inactivation of the p53 tumor suppressor pathway[13–18]. This subtype of lung cancer has been modeled by engineering conditional alleles of *Kras* and *Trp53* (which encodes the tumor suppressor Trp53) in mice[8,19]. Expression of the Cre recombinase in lung epithelial cells from $Kras^{LSL\text{-}G12D/+}$;$Trp53^{flox/flox}$ mice leads to the removal of a stop cassette and the expression of oncogenic $Kras^{G12D}$, as well as inactivation of p53. Although lung tumors in $Kras^{LSL\text{-}G12D/+}$;$Trp53^{flox/flox}$ mice are initiated synchronously, only a small fraction of tumors progress to acquire metastatic ability[3,8]. Thus, the direct analysis of these tumors at distinct stages of metastatic progression could provide unique insights into the molecular mechanisms that drive metastatic progression.

Here we leverage tumor barcoding in a mouse model of human lung adenocarcinoma to identify neoplastic cells at defined stages of metastatic progression. We use unbiased genomic analysis and a small-scale *in vivo* functional screen to uncover novel drivers of metastatic ability. By coupling these analyses with extensive human validation, functional metastasis assays *in vivo*, and molecular analyses, we identify a targetable pro-metastatic CD109–Jak–Stat3 axis.

RESULTS

Isolation of neoplastic cells from distinct stages of metastatic progression

To uncover the gene expression alterations that occur during lung cancer progression, we interrogated the gene expression state of neoplastic cells at different stages of malignant progression. We initiated tumors in *Kras*$^{LSL\text{-}G12D/+}$;*Trp53*$^{flox/flox}$ mice that also contained a *Rosa26*$^{LSL\text{-}tdTomato}$ Cre reporter (hereafter referred to as KPT mice) with a pool of barcoded lentiviral vectors that express Cre recombinase and purified Tomato^{+} cancer cells using FACS (Fig. 1a)[3,8,20]. Five to nine months after tumor initiation, cancer cells were isolated from 3–8 individual primary tumors per mouse, as well as from metastases from various organs including the lymph node, pleura, soft tissues, and liver (Fig. 1b and Supplementary Fig. 1a). Sequencing of the barcode region of the integrated lentiviral vectors established primary tumor–metastasis and metastasis–metastasis relationships (Fig. 1c,d). Tumor barcoding allowed us to distinguish nonmetastatic primary tumors (T_{nonMet}) from those primary tumors that had seeded macrometastases (T_{Met}; Fig. 1e and Supplementary Fig. 1d). We performed RNA sequencing (RNA-seq)-based gene expression profiling on ten T_{nonMet} primary tumors, nine T_{Met} primary tumors, and 24 individual macrometastases (Met) that represented 12 metastatic events (Fig. 1f and Supplementary Fig. 1d). To examine additional earlier stages of lung cancer development, we also analyzed premalignant cells from hyperplasias that developed in KPT mice shortly after tumor initiation (referred to as KPT-early or KPT-E), as well as from tumors from *Kras*^{G12D};*Rosa26*$^{LSL\text{-}tdTomato}$ (KT) mice, which have intact p53 signaling and whose tumors rarely gain metastatic ability (Fig. 1f and Supplementary Fig. 1d)[3,19]. Chimeric transcripts from the lentiviral long terminal repeat (LTR)-genomic DNA junctions and somatic mutations identified from our RNA-seq data confirmed the T_{Met} and Met relationships (Supplementary Figs. 1b,c and 2a–j).

Global analysis of lung cancer progression highlights two distinct transitions during primary tumor development

The clustering of all samples by gene expression clearly distinguished the KPT-E, KT, and T_{nonMet} samples from the Met samples. T_{Met} tumors clustered mostly with the T_{nonMet} samples, while some had gene expression profiles that were intermediate between these two distinct molecular states, consistent with varying contributions of the metastatic subclones within different T_{Met} tumors (Fig. 2a). KPT-E and KT tumors were also distinct from the more advanced T_{nonMet} and T_{Met} primary tumors (Fig. 2a). Finally, metastases clustered by their clonal relationship rather than by their metastatic site (Supplementary Fig. 3a). Removal of the gene signatures associated with each metastatic family still failed to uncover organ-specific differences in clustering, suggesting that the tissue environment at the metastatic site has limited influence on global gene expression programs in established macrometastases (Fig. 2b and Supplementary Fig. 3b–f).

Several hundred genes underwent significant changes in gene expression during the transition between the KPT-E or KT state and the state of the KPT primary tumors at later time points (T_{nonMet} and T_{Met}; Fig. 2c, Supplementary Fig. 3g, and Supplementary Tables 1–3). The T_{nonMet} and Met cancer cell states showed even more dramatic gene expression differences (Fig. 2d,e, Supplementary Fig. 3h and Supplementary Tables 1, 4, and 5). In

particular, the Met state was enriched for gene sets that reflected more aggressive cancer, increased proliferation, extracellular matrix remodeling, and *Kras* signaling (Supplementary Figs. 4 and 5a,b and Supplementary Tables 4 and 5). Previously established regulators of human lung adenocarcinoma metastasis, including lymphoid enhancer binding factor 1 (Lef1) and Hif transcriptional complex activity, were increased, whereas GATA-binding protein 6 (Gata6) activity was decreased in metastases (Supplementary Tables 4 and 5)[5,21,22]. Most of the genes that were previously identified as differentially expressed in cell lines derived from T_{nonMet} and Met tumors from this model[3] were confirmed in the *ex-vivo*-sorted samples (Supplementary Fig. 3i). Notably, the analysis of directly isolated cancer cells identified nearly three times as many differentially expressed genes than were identified in our previous analysis of cell lines from these tumors (Supplementary Fig. 3i)[3]. This highlights the value of analyzing *ex vivo* cancer cells directly from different stages of metastatic progression.

Relapse in patients with lung adenocarcinoma following potentially curative surgery is driven in part by subclinical pre-existing metastatic disease. We investigated whether a metastasis signature derived from the most differentially expressed genes between T_{nonMet} and Met samples could predict patient outcome. We developed a Cox random forest model from the signature using cross-validation and determined the threshold of predicted values that discriminated between long and short survival times among patients. Patients with primary tumors that had a gene expression program similar to that of the metastases from the mouse model had considerably shorter overall survival (Fig. 2f and Supplementary Fig. 5c). The predictive power of our signature was mostly comparable to previously published signatures (Supplementary Fig. 5d–k)[23–25]. Notably, the predictive power of this signature was even greater on the smaller set of patients with oncogenic *KRAS* mutations (Fig. 2g and Supplementary Fig. 5l,m).

Functional interrogation identifies several novel regulators of lung adenocarcinoma metastasis

To enable rapid and quantitative screening of candidate drivers of metastasis, we established a model of spontaneous metastasis, in which subcutaneous growth of a cell line that we generated from disseminated cancer cells from a KPT mouse (889 cells) leads to the development of liver and lung metastases. This model mimics all steps of the metastatic process, including dissemination, transient survival in circulation, invasion, and metastatic seeding and outgrowth in distant organs. We chose 23 candidate genes whose products could promote metastasis (hereafter referred to as pro-metastatic genes) for evaluation in a small-scale functional screen. We considered the magnitude and significance of differential expression between the T_{nonMet} and metastatic state, as well as the ability of each gene to predict outcome in patients with lung adenocarcinoma[23,26–28]. To uncover previously unknown regulators of metastatic ability, we also focused on genes for which there was little pre-existing data to link them to metastatic ability (Supplementary Table 1). We tested between three and ten lentiviral vectors that encoded short hairpin RNAs (shRNAs) specifically targeting the expression of each candidate gene to identify the shRNAs with the best knockdown efficiency. Knockdown of the expression of each candidate pro-metastatic factor in the 889 cell line had only minor, if any, effect on *in vitro* proliferation, no effect on

in vitro cell death, and no more than a 30% reduction in subcutaneous tumor growth (Supplementary Fig. 6b,c,i). Metastatic ability was assessed by enumerating fluorescent metastatic lesions in the lung and liver (Fig. 3a–c and Supplementary Fig. 6a,l,m)[28]. Knockdown of five genes (*Cd109, Lrp8, Kcnn4, Emp3,* and *Dbn1*) reduced metastatic seeding to the lung by >90% and almost completely inhibited metastasis to the liver (Fig. 3a and Supplementary Fig. 6d,j). For six of the highest-ranked candidate pro-metastatic genes (*Cd109, Cbfb, Lrp8, Gpx8, Onecut2,* and *Emp3*), we confirmed that knockdown using a second independent shRNA also reduced lung and liver metastases (Fig. 3b and Supplementary Fig. 6d,j,k).

CD109 is required for metastatic ability

Of our top hits, *CD109* was the strongest predictor of survival for patients with lung adenocarcinoma, and patients within the highest decile of *CD109* expression had approximately threefold shorter survival times than all of the other patients (Supplementary Table 1 and Supplementary Fig. 7a–f). Cd109 is a glycosylphosphatidylinositol (GPI)-linked cell surface protein of the α2-macroglobulin–complement superfamily[29–33] that was dramatically upregulated in Met samples (Supplementary Fig. 7g,h and Supplementary Table 1). *Cd109* knockdown almost completely abrogated the formation of lung and liver metastases from subcutaneous tumors and greatly reduced the number of metastases after intravenous transplantation (Fig. 3a–f and Supplementary Figs. 6d–h, 8a–d,h–l, and 9a–g). *Cd109* knockdown did not reduce the number of cancer cells in the lung at 5 min or 2 h after intravenous transplantation but reduced the number of cancer cells in the lung 2 d after transplantation (Supplementary Fig. 8e–g). The reduced number of cancer cells in the lung was not driven by reduced proliferation of cells in which *Cd109* had been knocked down (Supplementary Fig. 8j,k). *Cd109* knockdown also reduced liver metastasis after intrasplenic transplantation of 889 cells into recipient mice. *Cd109* knockdown reduced the metastatic ability of two additional $Kras^{G12D};Trp53^{-/-}$ mouse lung adenocarcinoma cell lines, 299DTC and 579ST, demonstrating that this cell surface protein is critical for metastasis to diverse organ sites (Fig. 3g and Supplementary Figs. 9c and 10a–i).

Consistent with the variable expression of *CD109* in human lung adenocarcinomas, human lung cancer cell lines also showed diverse CD109 expression (Supplementary Fig. 7i,j). We initially investigated the function of CD109 in a human lung cancer cell line (H460) that harbors oncogenic KRAS, expresses high levels of CD109, and forms liver and lung metastases after intravenous transplantation of H460 cells into recipient mice (Fig. 3h–k). *CD109* knockdown greatly reduced the ability of H460 cells to metastasize after intravenous transplantation, suggesting that CD109 can drive the remarkable metastatic aptitude of human lung adenocarcinoma (Fig. 3h–j and Supplementary Fig. 7k,l).

CD109 controls the activation of Stat3

Given the dramatic effect of *Cd109* knockdown on metastatic ability, we investigated global gene expression changes elicited by *Cd109* knockdown (Supplementary Table 6). This highlighted several potential pathways as effectors of Cd109 function, including Stat3 signaling (Supplementary Fig. 11a and Supplementary Tables 7 and 8)[31,32]. Activation of the Stat3 transcription factor is regulated by phosphorylation, and cells expressing a *Cd109*-

specific shRNA (shCd109) had a consistent and dramatic reduction in Stat3 phosphorylation. Reduced Stat3 phosphorylation was confirmed in *Cd109*-null clones and in two additional mouse lung adenocarcinoma cell lines in which *Cd109* was knocked down (Fig. 4a,b and Supplementary Fig. 11b,c). Expression of exogenous *Cd109* in a mouse embryonic fibroblast cell line also increased phosphorylation of Stat3 (Fig. 4c). Moreover, human lung cancer cell lines that express high levels of CD109 had higher STAT3 phosphorylation than cell lines with low levels of CD109 (Fig. 4d). *CD109* knockdown in human H460 cells, an additional *KRAS*-mutated cell line, and four *KRAS*-WT human lung cancer cell lines reduced STAT3 phosphorylation. Taken together, these data confirm that the regulation of Stat3 activation by Cd109 is conserved in mouse and human cells (Fig. 4e and Supplementary Fig. 11d–h).

Cd109 knockdown reduced the clonal growth ability of multiple mouse lung cancer cell lines under low-density-plating and anchorage-independent conditions (Fig. 4f and Supplementary Fig. 12a–e). Cd109 was also required for efficient migration, suggesting that a combination of these cellular phenotypes likely explains its prometastatic effect *in vivo* (Fig. 4f and Supplementary Fig. 12a–e). Consistent with a conserved molecular mechanism in human lung cancer, CD109 was also required for clonal growth and migration of several human lung cancer cell lines (Fig. 4g and Supplementary Fig. 12f–n).

STAT3 is a key effector of the pro-metastatic function of CD109

High levels of phosphorylated STAT3 (pSTAT3) correlate with poor differentiation and short survival of patients with lung adenocarcinoma; however, it remains unclear whether STAT3 is a key driver of lung cancer metastasis[34–37]. To directly test whether Stat3 activity is pro-metastatic in lung cancer, we determined whether knockdown of *Stat3* expression recapitulated the phenotypes resulting from knockdown of *Cd109* expression. Stat3 was not required for subcutaneous tumor growth, but *Stat3* knockdown greatly reduced liver and lung metastases from subcutaneous tumors, as well as after intravenous transplantation into recipient mice (Fig. 4h–j and Supplementary Fig. 13a–j). *Stat3* knockdown elicited similar cellular phenotypes as *Cd109* knockdown, dramatically reducing clonal growth ability and modestly affecting cell migration (Fig. 4f).

To directly test the epistatic relationship of Cd109 and Stat3, we expressed hyperactivatable Stat3 (Stat3C)[38] in cells expressing shControl or shCd109 (Supplementary Fig. 14a). Stat3C expression almost completely reversed the cell culture phenotypes elicited by *Cd109* knockdown (Fig. 5a). Restoration of Stat3 activity also increased the ability of shCd109-expressing cells to metastasize to the lung and liver from subcutaneous tumors (Fig. 5b–d and Supplementary Fig. 14b–d). Finally, expression of Stat3C also partially restored the ability of shCd109-expressing cells to form metastases after intravenous transplantation into recipient mice (Fig. 5e and Supplementary Fig. 14e–g).

CD109 is best characterized as a negative regulator of signaling involving transforming growth factor (TGF)-β and the Smad family of transcription factors[30,33,39], and we found that pSmad2 was increased in shCd109-expressing cell lines (Supplementary Fig. 15a). Consistent with the metastasis-suppressive role of Smad4 in some settings[40,41], expression of a dominant-negative Smad4 variant (dn-Smad4) in lung adenocarcinoma cells increased

their metastatic ability by approximately threefold (Supplementary Fig. 15b—f). Notably, not only did dominant-negative TGF-β receptor II and dnSmad4 fail to rescue the metastatic ability of shCd109-expressing cells, but dnSmad4 was also incapable of increasing the metastatic ability of these cells (Supplementary Fig. 15d–f). Collectively, these data are consistent with a model in which increased Cd109 expression drives metastatic ability through increased Stat3 activation and in which Stat3 activity is one critical component of the metastatic state.

Jak kinases are required for CD109-induced Stat3 activation

Members of multiple kinase families can phosphorylate Stat3 (refs. 42–44); therefore, we determined whether inhibition of Src, receptor tyrosine kinases or of Jak kinases could reduce Cd109-driven Stat3 phosphorylation. Pharmacologic inhibition of members of the Jak family of kinases, using multiple small-molecule inhibitors, reduced pStat3 levels similar to those observed after knocking down *Cd109* expression in 889 cells and also reduced pSTAT3 levels in H460 human lung cancer cells (Fig. 5g–j and Supplementary Fig. 16a–d). Inhibition of Jak kinase activity in fibroblasts overexpressing Cd109 also reduced pStat3 to a level similar to that in the parental cells, which express low levels of Cd109, suggesting that Cd109-induced Stat3 phosphorylation requires Jak kinase activity (Fig. 5f). The profile of drug sensitivity in lung cancer cells suggests that Jak1 and Jak3 likely function upstream of Stat3, and we confirmed reduced pStat3 levels in cells with *Jak1* expression knocked down (Supplementary Fig. 17a,b). Although multiple transmembrane receptors can scaffold the activation of Jak–Stat3, we found that knockdown of interleukin 6 signal transducer (*Il6st*; also known as *Gp130*) expression also reduced pStat3 levels in lung adenocarcinoma cells (Supplementary Fig. 18a,b). Finally, pyridone 6 treatment, *Jak1* knockdown, and *Gp130* knockdown each reduced the clonal-growth and migratory ability of lung cancer cells in culture, consistent with the phenotypes of cells with knockdown of *Cd109* and *Stat3* expression (Fig. 5k and Supplementary Figs. 16e–g, 17c,d and 18c,d).

Jak inhibition reduces metastatic ability

To determine whether Jak kinase activity was also required for metastatic ability, we initially inhibited these kinases before and during metastatic seeding, using the most potent pan-Jak-kinase inhibitor, pyridone 6 (refs. 45,46) (Fig. 5g and Supplementary Fig. 16a). Pyridone 6 treatment greatly reduced the number of micrometastases in the lung and liver after intravenous transplantation of 889 cells (Fig. 6a–d and Supplementary Fig. 19a–f). Pharmacologic Jak kinase inhibition with pyridone 6 also reduced liver metastases after intrasplenic transplantation of 889 cells into recipient mice (Fig. 6f,g and Supplementary Fig. 19g–i). Consistent with the effect of *CD109* knockdown on human H460 cells, pyridone 6 treatment also reduced metastatic seeding of H460 cells (Fig. 6e and Supplementary Fig. 19j). The Jak kinase inhibitors ruxolitinib[47] and filgotinib[48] also inhibited Stat3 phosphorylation and reduced metastatic seeding of 889 cells (Supplementary Fig. 19k–m). Finally, knockdown of Jak1 or Gp130 reduced metastatic ability after intravenous transplantation, consistent with Jak1 and Gp130 being integral components of a pro-metastatic Jak–Stat3 axis (Supplementary Figs. 17e–i and 18e–m).

We further investigated the role of Jak–Stat3 signaling in CD109-driven metastatic seeding and outgrowth by treating mice with subcutaneous tumors with pyridone 6. We tested the effect of Jak kinase inhibition on metastasis formation with and without surgical removal of the subcutaneous tumors. The development of metastases in each setting was reduced > 10-fold by pyridone 6 treatment (Fig. 6h–k). Jak kinase inhibition had a much less dramatic effect on subcutaneous tumor growth, and continued Jak kinase inhibition after subcutaneous tumor removal had little effect on the number or size of metastases, consistent with the dedicated function of the CD109–Jak–Stat3 axis to the metastatic process (Supplementary Fig. 19n–r).

The molecular mechanisms that drive metastasis are likely to be diverse across different types of cancer, but *CD109* expression has been documented in several other cancer types[49,50]. Notably, high *CD109* expression predicted poor patient outcome in several other diverse cancer types (Supplementary Fig. 20a–e). *CD109* knockdown also reduced STAT3 phosphorylation in several human renal cell carcinoma cell lines, suggesting that this axis may function in subsets of patients with other tumor types (Supplementary Fig. 20f).

DISCUSSION

The unbiased analysis of metastatic progression in genetically engineered mouse models affords a unique opportunity to study cancer progression in a controlled setting. Our tumor-barcoding approach allows primary tumor–metastasis and metastasis–metastasis relationships to be determined, thus enabling accurate comparisons throughout the spectrum of metastatic progression. The integration of molecular lineage-tracing and fluorescent-tagging approaches with additional low-input unbiased genomic methods[51] should enable an even greater understanding of the mechanisms that drive the progression of lung adenocarcinomas. Similar approaches can also be leveraged to investigate cancer progression and metastasis in mouse models of many other cancer types[9–11,52].

Our study provides key insights into the aberrant state of malignant cancer cells and the mechanisms that they use to overcome the hurdles that limit systemic spread. We defined a metastatic lung cancer cell state that is accompanied by a loss of differentiation and dramatic remodeling of many gene expression programs (Supplementary Figs. 4 and 5a). Notably, a gene expression signature of this metastatic state predicted the clinical outcome of patients with lung adenocarcinoma, suggesting that a fraction of patients' tumors likely undergo similar cell-state changes. This signature was an even better predictor of survival of patients with oncogenic-*KRAS*-driven lung adenocarcinoma; therefore, we speculate that the initiating oncogenic alteration influences the likelihood that a tumor will use a particular mechanism of metastatic progression. Our approach also allowed the unbiased analysis of metastases that arose naturally in different organs. Unexpectedly, we uncovered very few gene expression differences between macrometastases in different organs (Fig. 2b and Supplementary Fig. 3b–f). Although transient gene expression differences could promote early stages of organ tropism, our data suggest that the tissue environment at the metastatic site has a limited influence on global gene expression programs in established macrometastases.

We uncovered a pro-metastatic axis in which high levels of expression of the GPI-linked cell surface protein CD109 led to activation of Jak–Stat3 signaling in cancer cells (Fig. 4a,e). Across multiple mouse and human lung adenocarcinoma cell lines, CD109 was a major regulator of cellular phenotypes associated with malignancy and metastatic ability *in vivo* (Figs. 3 and 4). Our data suggested that in addition to the potential role of Jak–Stat signaling in stromal cell types, CD109 is a cell autonomous regulator of metastatic ability in lung adenocarcinoma. Several important questions regarding this pro-metastatic pathway remain. It remains unclear what leads to the dramatic upregulation of Cd109 expression in metastatic lung adenocarcinoma cells. The direct molecular link between Cd109 and Gp130–Jak–Stat3 signaling, the function of other pathways downstream of Cd109, and the full complement of Stat3 targets that drive metastatic ability also remain to be elucidated and will be important areas of future investigation (Fig. 61 and Supplementary Fig. 21).

Our studies suggest that direct targeting of CD109 or inhibition of the Jak family of kinases could provide therapeutic benefit in the neo-adjuvant or adjuvant setting. Unlike Stat3-deficient mice, Cd109-deficient mice are viable and show only minor defects[31,53], suggesting that Cd109-driven Jak–Stat3 signaling may be selectively important within metastatic tumors. There is a growing interest in the potential benefit of anti-metastatic therapies in cancers like lung adenocarcinoma, in which a large number of patients are initially diagnosed with localized disease[54]. Interrupting Cd109-induced Jak–Stat3 signaling could reduce metastasis from primary tumors, as well as the propagative seeding of additional metastases from pre-existing metastases, in a subset of patients.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ONLINE METHODS

Mouse strains and tumor induction. *Kras*$^{LSL\text{-}G12D}$, *Trp53*$^{flox/flox}$, and *Rosa26*$^{LSL\text{-}tdTomato}$ mice have been described[8,19,55]. The lentiviral-barcode-Cre (Lenti-BC-Cre) vector was generated as previously described[3]. Lenti-BC-Cre was co-transfected with packaging vectors (delta8.2 and VSV-G) into 293T cells using TransIT-LT1 (Mirus Bio). The supernatant was collected at 48 and 72 h, ultracentrifuged at 25,000 r.p.m. for 90 min, and resuspended in PBS. Tumors were initiated by intratracheal infection of mice with lentiviral vectors expressing Cre recombinase, as previously described[56]. The Stanford Institute of Medicine Animal Care and Use Committee approved all animal studies and procedures.

Tumor dissociation and cell sorting.

Primary tumors and metastases were dissociated using collagenase IV, dispase, and trypsin at 37 °C for 30 min. After dissociation, the samples remained continually on ice, were in contact with ice-cold solutions, and were in the presence of 2 mM EDTA and 1 U/ml DNase to prevent aggregation. All solutions contained the RNA polymerase inhibitors α-amanitin (Sigma-Aldrich, A2263) and actinomycin D (Sigma-Aldrich, A1410) to prevent changes in gene expression that might be induced during the dissociation and sorting steps. Cells within

the pleural cavity were collected immediately after euthanasia by making a small incision in the ventral aspect of the diaphragm followed by introduction of 1 ml of PBS. Cells were stained with antibodies to CD45 (30-F11), CD31 (390), F4/80 (BM8), and Ter119 (all from BioLegend) to exclude hematopoietic and endothelial cells (lineage-positive (Lin^+) cells). DAPI was used to exclude dead cells. BD LSR II analyzers and FACSAria sorters (BD Biosciences) were used for cell sorting.

Barcode PCR and analysis.

RNA and genomic DNA were extracted from sorted cancer cells using AllPrep DNA/RNA Micro Kit (Qiagen). Lentiviral barcode sequences were PCR-amplified from genomic DNA. PCR reactions were treated with ExoI and FastAP (New England BioLabs), to remove primers and nucleotides, followed by Sanger sequencing.

Mouse *ex vivo* RNA-seq preparation.

RNA quality of each tumor sample was assessed using the RNA6000 PicoAssay for the Bioanalyzer 2100 (Agilent) as per the manufacturer's recommendation. All of the RNA used for RNA-seq had a RNA integrity number (RIN) >8.0. 20 ng total RNA per sample was used for cDNA synthesis using the Ovation RNA-seq system (NuGEN Technologies, Inc.; San Carlos, CA, USA) as described in the manufacturer's instructions. Briefly, the mRNA was reverse-transcribed to synthesize the first-strand cDNA using a combination of random hexamers and poly(T) chimeric primer. Double-stranded DNA was generated by fragmentation of the mRNA template strand using RNA-dependent DNA polymerase. The dsDNA was purified using Agencourt RNAClean XP beads. The DNA was amplified linearly using a single-primer isothermal amplification (SPIA) process in which RNase H degrades RNA in a DNA–RNA heteroduplex at the 5′ end of the double-stranded cDNA, after which the SPIA primer binds to the cDNA and the polymerase starts replication at the 3′ end of the primer by displacement of the existing forward strand. Finally, random hexamers were used to amplify the second-strand cDNA linearly. The double-stranded cDNA was subjected to library preparation using the Illumina TruSeq DNA sample preparation kit (low-throughput protocol) according to manufacturer's protocol. Briefly, 5 μg of double-stranded cDNA was sheared to ~200-bp fragments and used for library preparation. cDNA was sheared by sonication (Covaris model S1) with a duty cycle of 10%, intensity of 5, and cycle/burst of 100 for 5 min, according to the manufacturer's instructions. 1 μg of cDNA fragments was then blunt-ended through an end-repair reaction and ligated to platform-specific double-stranded barcoded adapters using library preparation kits from the TruSeq DNA sample preparation kit. The purified cDNA library products were evaluated using the Agilent bioanalyzer and diluted to 10 nM for cluster generation *in situ* on HiSeq paired-end flow cells using the CBot automated cluster-generation system, followed by massively parallel sequencing (2 × 100 bp) on HiSeq 2000 instruments. We obtained 104-bp mate-paired reads from DNA fragments of an average length of 250 bp.

Detection of lentiviral integration sites from RNA-seq data.

To detect lentiviral integration sites in our samples, we used SAMtools to align all short reads from RNA-seq data against lentiviral LTR sequences (CAGTGTGGAAAATCTCTAGC)[57]. Adjacent chimeric reads were then aligned to the

mouse genome on the University of California, Santa Cruz (UCSC) genome browser to identify the lentiviral integration site.

Identification and confirmation of somatic mutations in RNA-seq data.

Mutations were called from the RNA-seq data using SNPiR[58]. First, a splice-aware reference genome was created by concatenating the mouse mm9 genome with a library of sequences spanning all of the splice junctions from three different gene annotations: UCSC Genes, RefSeq Genes, and Ensembl Genes. Next, reads were mapped to this reference genome using BWA[59], with the coordinates of reads mapping onto splice junctions adjusted to their corresponding position on mm9. Duplicate reads, unmapped reads and low-quality reads were removed. Insertion and deletion (indel) realignment was performed, and mutations were called using the Genome Analysis Toolkit (GATK)[60]. To filter false-positive mismatches caused by technical artifacts, mutations within the first 6 bp of reads, within the 4 bp of a splice junction, within homopolymer runs or repetitive regions, within reads that do not uniquely map to the reference genome, and those located at known sites of RNA editing were removed. Custom scripts further categorized these mutations by filtering out mutations that occurred universally in all of the samples, those that appeared in every tumor, those that appeared in any normal tissue samples, and those that appeared in any of our pooled samples (KT and KPT-E). Mutations were then categorized into families if they appeared in multiple samples from the same mouse. This method identified putative mutations that would relate primary metastatic tumors (T_{Met}) to metastases, as well as metastases to other metastases. When possible, we prioritized the validation of mutations to single-nucleotide polymorphisms (SNPs) with at least ten reads in each sample and allele frequencies of at least 0.1. To validate mutations at the DNA level, primers were designed to flank the genomic locus of interest, and PCR was performed using GoTaq DNA polymerase (Promega). Excess primer and dNTPs were removed using ExoI and FastAP (New England BioLabs) followed by Sanger sequencing. The traces were aligned using Sequencher (Gene Codes Corp.), and the traces were examined for presence of the mutated allele.

Mouse *ex vivo* RNA-seq data quantification, normalization, and clustering.

We used Sailfish[61] to quantify expression. We kept only transcripts with a maximum transcripts per million (TPM) value >1 in at least one sample. We performed the inverse hyperbolic sine transformation for variance stabilization and quantile-normalized the data. In Figure 2a, we performed hierarchical clustering on this normalized data using Euclidean distance and complete linkage for all KT, KPT-E, T_{nonMet}, T_{nonMet}, and Met samples. In Supplementary Figure 3a, we performed hierarchical clustering of the normalized data for all Met samples using Euclidean distance and complete linkage. In Figure 2b, we first estimated the effect of the clonal families on gene expression, as well as known batch effects, and discovered confounding factors in a linear-regression model using surrogate variable analysis (SVA)[62]. We then regressed out the estimated clonal effect and reclustered the adjusted data using the same hierarchical clustering procedure.

Genes driving metastasis.

To discover the genes that drive the transitions between different stages of cancer progression, we used the following procedure: (i) estimate batch effects and confounding

factors using SVA while protecting the states being compared, (ii) fit a linear model with the newly identified confounding factors, batch effects, and metastatic effect, and (iii) rank genes in the order of evidence for differential expression using an empirical Bayes method[63]. We compared T_{nonMet} versus Met samples, T_{nonMet} and T_{Met} versus KPT-E and KT samples, as well as liver versus lymph node metastases. Mouse genes were mapped to a direct human ortholog using the 'mmusculus_gene_ensembl' database in BioMart[64]. For GSEA[65], a rank-ordered list of differentially expressed genes was analyzed using GSEA v2.2.0 software available from the Broad Institute (http://www.broad.mit.edu/gsea). Only genes with a human homolog were used for this analysis.

Analysis of gene expression data from human lung adenocarcinoma.

Human expression and clinical data for lung adenocarcinomas was downloaded from the Cancer Genome Atlas (TCGA) Data Portal[66]. We used tumor samples with barcode '01' and kept only genes with nonzero expression. Gene expression data was *z*-score-normalized. For the clinical data, we annotated patients with an event as 'true' if the patient died and 'false' if the patient did not die. The accompanying time is either the time of death or the listed time of last visit. We converted those individuals who died of a cause other than cancer to 'alive at a time of death'. We converted individuals who died after 6 years to 'alive at 6 years'. We truncated all times >6 years to 6 years.

For each gene profiled in the TCGA expression data set, we chose seven *z*-score thresholds (−1.00, −0.50, −0.25, 0, 0.25, 0.50, and 1.00) and tested the difference in Kaplan–Meier survival curves for individuals with expression levels below and above the chosen *z*-score threshold using the 'G-rho' family of tests (survival package in R). We required at least 20 individuals to be in each group above and below the *z*-score threshold. We performed this procedure with tenfold cross-validation and selected the threshold that led to the minimum *P* value across the 10 folds. With this threshold we calculated both the *P* value and the hazard ratio between the two groups by fitting a Cox proportional-hazards regression model (https://cran.r-project.org/web/packages/survival/index.html). We performed multiple-hypothesis corrections by controlling the false discovery rate at a level of 0.05 (https://github.com/jdstorey/qvalue).

Metastasis signatures and survival model.

We developed a mouse metastasis signature in which we selected the highest-ranked 276 genes that differentiated Met and T_{nonMet} tumors (ranked by *q*-value). We combined the selected features into a metastasis-signature score. The score we used is the raw predicted value of each example when run through a Cox survival random forest model using the R package randomForestSRC[67,68] with the selected genes in our signature. We used tenfold cross-validation to find the metastasis score threshold that maximized survival differences between the corresponding two groups. In each fold, we trained a survival model based on the training examples and predicted human survival time on the testing examples. We then optimized over all possible score thresholds at increments of 0.1 while requiring at least 20 patients in each group, selected the threshold that had the most significant difference in Kaplan–Meier survival curve, and used that threshold to predict high or low rates of survival for the held out examples. We also tested the ability of our mouse metastasis gene signature

to differentiate survival times for patients with a known *KRAS* mutation using the same procedure. For evaluating our signature on the smaller set of patients with *KRAS* mutations, we reduced the minimum group size to 15.

We also used this method to determined the predictive power of several published signatures[23–25] (Supplementary Fig. 5c–j). It is important to note that because these signatures have different numbers of genes and were used in different predictive models, direct comparisons are not entirely fair. In addition, we chose predictive expression thresholds for each gene through cross-validation, which may not have been comparable to the original signature, to give each existing signature the best predictive potential with our method. Still, these results indicate that our signature was largely comparable to, or better than, these published signatures. As an alternate approach, we also assessed how well our signature performs relative to random signatures with the same number of genes (Supplementary Fig. 5k). This data suggests an empirical *P* value of ~0.05, but all signatures with strong predictive power achieved the same order of magnitude of *P* value (10^{-6}). Overall, this data supports our conclusion that aspects of this metastatic program are recapitulated in a subset of human lung adenocarcinomas.

Cell lines, lentiviral-shRNA vectors, gene expression, and quantification.

The 889 and 299DTC cell lines were generated from disseminated cancer cells within the pleural cavity of a lung-adenocarcinoma-bearing $Kras^{LSLG12D/+}$;$Trp53^{flox/flox}$;$R26^{LSL\text{-}Tomata}$ (KPT) mice. The 579ST cell line was generated from a soft-tissue metastasis. These cell lines were chosen because they are Tomato$^+$ and form metastases from subcutaneous tumors. Human lung cancer cells were originally purchased from the ATCC, and the H460 and H23 cell identities were validated by Genetica DNA Laboratories using short tandem repeat (STR) analysis. Human renal cell carcinoma cells (Caki2, ACHN, Caki1, and 786-C3) were kindly provided by Jonathan Pollack (Stanford Medical School). All cell lines were confirmed to be mycoplasma negative (MycoAlert Detection Kit, Lonza).

Lentivirus was generated using the Delta8.2 and VSV-G packaging plasmids. To generate stable pLKO-lentivirus-transduced cell lines, we infected cells with the pLKO virus and began puromycin (4 μg/ml) selection after 72 h. We tested between three and ten pLKO-puromycin vectors with shRNAs targeting each candidate gene from the RNAi Consortium (TRC) whole-genome library. After infection and puromycin selection, we screened for the best knockdown of the 23 candidate pro-metastatic-factor-encoding genes by qPCR (Supplementary Table 9). pLKO-shGFP (Target sequence: 5′GCAAGCTGACCCTGAAGTTCAT3′) was used as the shControl-expressing vector for most of the experiments.

To make retroviral vectors expressing TGFBR2 (TGF-βRII) or SMAD4 dominant-negative mutants, cDNAs were generated by PCR amplification from pCMV5-HA-TBRII-(delta Cyt) (Addgene# 14051) or pRK-DPC4-deltaC-Flag (Addgene# 12628) and then cloned into the NcoI and EcoRI sites of pMSCV-GFP-blasticidin. MSCV retroviral vectors were generated using pCL-Eco and pseudotyped with VSV-G. MSCV-infected cells were selected with blasticidin (2 μg/ml) for 10 d.

To express Cd109 in mouse fibroblasts, full-length *Cd109* cDNA was PCR-amplified and cloned into the Lenti-pGK-Cre vector to generate Lenti-CAAGS-Cd109-pGK-Cre. *loxP*-Stop-*loxP*-YFP mouse embryonic fibroblasts were infected with the Cd109–Cre lentivirus for at least 72 h and then FACS-purified based on yellow fluorescent protein (YFP) expression.

To express active Stat3 (Stat3C), lentivirus was generated using the EF.STAT3. Ubc.GFP (Addgene#24983) or FUGW control (GFP control, Addgene# 14883) vector. Stably transduced 889 cell lines were subsequently FACS-purified based on GFP expression.

Gene knockdown or expression was confirmed by qPCR using standard SYBR Green qPCR protocols and normalized to levels of mouse or human *Gapdh.* Primer sequences used in this study are listed in Supplementary Table 10.

Cd109-knockdown cells, RNA-seq library preparation, and analysis.

Total RNA was isolated from 1×10^6 cells of each line (889-shControl, 889-shCD109#1, and 889-shCd109#2) in duplicate using the Qiagen RNAeasy mini kit, resulting in six samples. For each sample, 1 μg of total RNA was used for library construction. The integrity and quality of RNA was assessed before library construction using an Agilent Bioanalyzer 2100. RNA-seq libraries were prepared using the Illumina TruSeq RNA v2 kit according to the manufacturer's instructions. High-throughput sequencing was performed on a single lane of an Illumina HiSeq 2000 for 100-bp paired-end reads. For analysis of the RNA-seq reads, we performed differential gene and transcript expression analysis using TopHat and Cufflinks[69]. RNA-seq reads were separately aligned to the mouse genome (mm10) using TopHat. The resulting alignment data from TopHat were then fed to the Cufflinks assembler (version 0.9.3) to assemble aligned RNA-seq reads into transcripts. Annotated transcripts were obtained from the UCSC genome browser (http://genome.ucsc.edu) and the Ensembl database; the category of transcripts was described at http://vega.sanger.ac.uk/info/about/gene_and_transcript_types.html. Transcript abundances were measured in fragments per kb of exon per million fragments mapped (FPKM). Finally, Cuffdiff was used to define differential expression.

Generation of *Cd109*-knockout lines.

A 20-bp guide sequence (5′-CCACCTTCTCTGCTTGTGCG-3′) targeting DNA within the first exon of *Cd109* was selected from the Broad Institute sgRNA designer (http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design-v1), which predicts high-specificity protospacer-adjacent motif (PAM) target sites in the mouse genome[70]. Two complementary oligos containing the *Cd109* guide sequence and BsmBI ligation adapters were annealed and ligated into the BsmBI-digested lentiCRISPRv2 vector[71]. To generate stable CRISPR-based knockout cell lines, cells were infected with lentiCRISPRv2 virus for 72 h and then selected with puromycin (4 μg/ml). Multiple FACS-sorted clonal cell lines were generated, and the target regions were PCR-amplified with primers (*Cd109* ex1 Fwd 5′-GCGGGGGAAATGGGCTCCCTATG′ and *Cd109* ini Rev 5′-CCGGTCACCTCAGTCACACTGA-3′) followed by Sanger sequencing to validate insertions and deletions (indels).

Cell line transplantation and analysis.

6- to 10-week-old NSG mice of similar weights were used for cell transplantation experiments. For the initial subcutaneous injections, 5×10^4 889 cells were injected into both shoulders and both flanks of NSG recipient mice. Mice were analyzed 3 weeks after transplantation. For the subcutaneous tumors that were injected with and without pyridone 6 treatment, 5×10^4 889 cells were injected into both flanks of NSG recipient mice. To model treatment before and after surgical resection, we removed the subcutaneous tumors 10 d after transplantation. The mice were anesthetized with isoflurane, adjusted to a prone position, and remained under surgical anesthesia throughout the duration of the procedure. The subcutaneous tissues were incised with a no. 11 blade scalpel and the tumors were removed with an ~2-mm circumferential resection margin. Hemostasis was ensured. The skin was reapproximated in a linear fashion and closed with surgical staples.

For intravenous transplantation, cell number was based on the time points for analysis: (i) 5×10^5 889 cells were used for the 10-min, 2-h, and 2-d experiments; (ii) 1×10^5 889 cells were used for the 5-d or 7-d experiments; and (iii) 5×10^4 889 cells or 5×10^5 H460 cells were used for the 12-d experiments or those of longer duration. Tumor cells were injected into the lateral tail veins (in 200 μl PBS) of NSG recipient mice. Cohorts of mice were analyzed at the indicated times after transplantation. For analyses, lungs and liver were digested, and cells were isolated as described above. Single-cell suspensions were analyzed using a LSRII analyzer (BD) for Tomato$^+$ cells or were stained with an antibody specific for human HLA-A, HLA-B, and HLA-C (W6/32; BioLegend). For proliferation analysis, mice were labeled for 24 h with one intraperitoneal injection of BrdU (50 mg/kg). After cell isolation, 50% of cells were fixed, stained, and analyzed for the cancer cells that were BrdU$^+$.

For intrasplenic injection to directly seed liver metastases, 5×10^4 889 cells were injected into the spleen of NSG recipient mice in 50 μl PBS using standard methods[72]. The spleen was removed 10 min after injection, and mice were analyzed 12 or 18 d after transplantation. In all experiments, surface-Tomato$^+$ metastases in the lung and liver were quantified using a fluorescence dissecting scope. Tumor area relative to total area was quantified using ImageJ on hematoxylin and eosin (H&E)-stained sections.

Western blotting.

Cells were lysed in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1mM EDTA, 1% deoxycorticosterone (sodium), 0.1% SDS, and 1% Triton X-100) containing Halt protease and phosphatase inhibitors (Thermo Scientific, 78440). Lysates were incubated on ice for 30 min and spun at 10,000g for 20 min at 4 °C. Supernatants were used as total lysates. Protein concentrations were determined with the BCA protein assay (Pierce, 23225). Supernatants were boiled in reducing SDS sample buffer, and 20 μg of protein lysate per lane was separated on a 4–12% Bis-Tris gel (NuPage) and transferred onto a PVDF membrane. Membranes were blocked for 1 h with 1% nonfat dried milk in TBS and immunoblotted using primary antibodies against Hsp90 (BD Transduction Laboratories, 610419), actin (Sigma-Aldrich; clone AC-15, A1978), Cd109 (Santa Cruz Biotechnology, sc-271085), phospho-Stat3 and Stat3 (Cell Signaling Technology, 8204), phospho-Smad2 and Smad2

(Cell Signaling Technology, 11958), Jak1 (Cell Signaling Technology, 3344), or Gp130 (Bethyl Laboratories, A304-929A). Incubations with primary antibodies were followed by incubations with horseradish peroxidase (HRP)-conjugated secondary anti-mouse-Ig (Santa Cruz Biotechnology, sc-2005) and anti-rabbit-Ig (Santa Cruz Biotechnology, sc-2004) antibodies, and membranes were developed with ECL 2 Western Blotting Substrate (P180196, ThermoScientific Pierce). The original blots are shown in Supplementary Figure 22.

Reagents.

The following inhibitors were used at the indicated concentrations: dasatinib (Selleck Chemicals, S1021), erlotinib (Selleck Chemicals, S1023), fedratinib (Selleck Chemicals, S2736), tofacitinib (Selleck Chemicals, S2789), ruxolitinib (Selleck Chemicals, S1378), pyridone 6 (EMD Millipore, 420099), cucurbitacin I (EMD Millipore, 238590), and filgotinib (Selleck Chemicals, S7605)

Histologic preparation and immunohistochemistry.

Samples were fixed in 4% formalin in PBS overnight and transferred to 70% ethanol before paraffin-embedding. Immunohistochemistry was performed on 4-μm sections with the ABC Vectastain kit (Vector Laboratories) with an antibody to RFP (Rockland; 600-401-379), BrdU (BD Biosciences, 3D4), or cleaved caspase-3 (Cell Signaling Technology, D3E9). Sections were developed with 3,3′-diaminobenzidine (DAB) and counterstained with H&E using standard methods. Percentage tumor area was calculated using ImageJ. Positive cells per optical field were counted using ImageJ on randomly chosen 20 × -fields per section.

Cell culture assays.

For the low-density plating assay, 2.5×10^3 cells were seeded in triplicate on a 10-cm plate. Mouse cell lines were allowed to grow for 1 week before quantification using ImageJ; human cell lines were grown for 2 weeks. To assess the proliferation rate, 1×10^5 cells were plated in each well of a 6-well plate. Eighteen hours later, the subconfluent cells were labeled with 10 μM BrdU for 2 h, followed by anti-BrdU staining using the BD APC flow kit according to the manufacturer's instructions. Proliferation was also assessed using PrestoBlue Cell Viability Reagent (A13261, Invitrogen), with 2.5×10^3 mouse cells or 5×10^3 human cells seeded in triplicate wells of a 96-well plate.

For anchorage-independent growth assays, 2×10^4 mouse cells or 4×10^4 human cells were plated in triplicate wells of a 6-well tissue culture plate in 0.4% agar in culture medium on top of a layer of 0.8% agar with culture medium. Cells were allowed to grow at 37 °C for 2–3 weeks. Colonies that were inherently tdTomato$^+$ (as the 889 cell line was derived from a KPT mouse) were quantified using the fluorescence microscope, otherwise colonies were stained with 0.2% crystal violet at room temperature for 30 min and subsequently destained with water for several days. Once the colonies were visible by eye, they were counted using a microscope and ImageJ.

For migration assays, 1×10^6 cells were plated in triplicate in 6-well plates per experiment overnight, such that confluency was reached the next day. A 200-μl pipette tip was used to

initiate the scratch. The gap was monitored over time, and gap measurement and analysis was quantified using Tscratch[73].

Statistical analysis.

Graphs and statistics were generated using the GraphPad Prism software. Significance, where indicated, was calculated using the unpaired Student's *t*-test. No statistical method was used to predetermine sample size.

Data availability.

Mouse *ex vivo* RNA-seq data has been deposited to the Gene Expression Omnibus under accession numbers GSE84447 and GSE84449. RNA-seq data for the 889-shControl and 889-shCd109 cell lines is under accession number GSE84448.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank P. Chu and S. Cheemalavagu for technical assistance, the Stanford Shared FACS Facility and Protein and Nucleic Acid Facility for expert assistance, J. Pollack and X. Gong for reagents, S. Dolan and A. Orantes for administrative support, and D. Feldser, C. Murray, and members of the Winslow lab for helpful comments. This work was supported by an American Lung Association Fellowship (C.-H.C.), the US National Institutes of Health (NIH) (grant no. F32-CA189659 (J.J.B.), T32-CA009302 (A.L.S.), R01-GM102484 (J.B.L.), R01-CA157510 (E.A.S.-C.), R01-CA175336 (M.M.W.), and R01-CA204620 (M.M.W.)), the Stanford Biomedical Informatics Training Grant from the National Library of Medicine (LM-07033; P.G.G.), a Bio-X Stanford Interdisciplinary Graduate Fellowship (P.G.G.), Stanford Graduate Fellowships (Z.N.R. and G.R.), National Science Foundation Graduate Research Fellowships (D.R.C. and Z.N.R.), the Spider Internship Funds (A.F.W.), an Alfred Sloan Fellowship (A.K.), a V Foundation for Cancer Research Martin D. Abeloff, M.D. V Scholar Award (M.M.W.), and, in part, by a Stanford Cancer Institute support grant (NIH P30-CA124435).

References

1. Nguyen DX, Bos PD & Massagué J Metastasis: from dissemination to organ-specific colonization. Nat. Rev. Cancer 9, 274–284 (2009). [PubMed: 19308067]
2. Massagué J & Obenauf AC Metastatic colonization by circulating tumor cells. Nature 529, 298–306 (2016). [PubMed: 26791720]
3. Winslow MM et al. Suppression of lung adenocarcinoma progression by Nkx2-1. Nature 473, 101–104 (2011). [PubMed: 21471965]
4. Bos PD et al. Genes that mediate breast cancer metastasis to the brain. Nature 459, 1005–1009 (2009). [PubMed: 19421193]
5. Nguyen DX et al. WNT–TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. Cell 138, 51–62 (2009). [PubMed: 19576624]
6. Yang J et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell 117, 927–939 (2004). [PubMed: 15210113]
7. Clark EA, Golub TR, Lander ES & Hynes RO Genomic analysis of metastasis reveals an essential role for RhoC. Nature 406, 532–535 (2000). [PubMed: 10952316]
8. Jackson EL et al. The differential effects of mutant *p53* alleles on advanced murine lung cancer. Cancer Res. 65, 10280–10288 (2005). [PubMed: 16288016]
9. Kirsch DG et al. A spatially and temporally restricted mouse model of soft-tissue sarcoma. Nat. Med 13, 992–997 (2007). [PubMed: 17676052]

10. Chiou SH et al. Pancreatic cancer modeling using retrograde viral vector delivery and *in vivo* CRISPR-Cas9-mediated somatic genome editing. Genes Dev. 29, 1576–1585 (2015). [PubMed: 26178787]
11. Meuwissen R et al. Induction of small-cell lung cancer by somatic inactivation of both *Trp53* and *Rb1* in a conditional mouse model. Cancer Cell 4, 181–189 (2003). [PubMed: 14522252]
12. Kim M et al. Comparative oncogenomics identifies *NEDD9* as a melanoma metastasis gene. Cell 125, 1269–1281 (2006). [PubMed: 16814714]
13. Ding L et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455, 1069–1075 (2008). [PubMed: 18948947]
14. Takahashi T et al. *p53:* a frequent target for genetic abnormalities in lung cancer. Science 246, 491–494 (1989). [PubMed: 2554494]
15. Rodenhuis S et al. Mutational activation of the *Kras* oncogene. A possible pathogenetic factor in adenocarcinoma of the lung. N. Engl. J. Med 317, 929–935 (1987). [PubMed: 3041218]
16. Rodenhuis S et al. Incidence and possible clinical significance of *Kras* oncogene activation in adenocarcinoma of the human lung. Cancer Res. 48, 5738–5741 (1988). [PubMed: 3048648]
17. Weir BA et al. Characterizing the cancer genome in lung adenocarcinoma. Nature 450, 893–898 (2007). [PubMed: 17982442]
18. Siegel RL, Miller KD & Jemal A Cancer statistics, 2016. CA Cancer J. Clin 66, 7–30 (2016). [PubMed: 26742998]
19. Jackson EL et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic *Kras*. Genes Dev. 15, 3243–3248 (2001). [PubMed: 11751630]
20. Caswell DR et al. Obligate progression precedes lung adenocarcinoma dissemination. Cancer Discov. 4, 781–789 (2014). [PubMed: 24740995]
21. Cheung WK et al. Control of alveolar differentiation by the lineage transcription factors GATA6 and HOPX inhibits lung adenocarcinoma metastasis. Cancer Cell 23, 725–738 (2013). [PubMed: 23707782]
22. Kim WY et al. HIF-2α cooperates with RAS to promote lung tumorigenesis in mice. J. Clin. Invest 119, 2160–2170 (2009). [PubMed: 19662677]
23. Shedden K et al. Gene-expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat. Med 14, 822–827 (2008). [PubMed: 18641660]
24. Der SD et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. J. Thorac. Oncol 9, 59–64 (2014). [PubMed: 24305008]
25. Van Laar RK Genomic signatures for predicting survival and adjuvant chemotherapy benefit in patients with non-small-cell lung cancer. BMC Med. Genomics 5, 30 (2012). [PubMed: 22748043]
26. Gao J et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal 6, pl1 (2013).
27. Győrffy B, Surowiak P, Budczies J & Lánczky A Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One 8, e82241 (2013). [PubMed: 24367507]
28. Brady JJ et al. An Arntl2-driven secretome enables lung adenocarcinoma metastatic self-sufficiency. Cancer Cell 29, 697–710 (2016). [PubMed: 27150038]
29. Finnson KW et al. Identification of CD109 as part of the TGF-β receptor system in human keratinocytes. FASEB J. 20, 1525–1527 (2006). [PubMed: 16754747]
30. Bizet AA et al. The TGF-β co-receptor, CD109, promotes internalization and degradation of TGF-β receptors. Biochim. Biophys. Acta 1813, 742–753 (2011). [PubMed: 21295082]
31. Mii S et al. Epidermal hyperplasia and appendage abnormalities in mice lacking CD109. Am. J. Pathol 181, 1180–1189 (2012). [PubMed: 22846721]
32. Litvinov IV et al. CD109 release from the cell surface in human keratinocytes regulates TGF-β receptor expression, TGF-β signaling and STAT3 activation: relevance to psoriasis. Exp. Dermatol 20, 627–632 (2011). [PubMed: 21539622]
33. Hagiwara S et al. Processing of CD109 by furin and its role in the regulation of TGF-β signaling. Oncogene 29, 2181–2191 (2010). [PubMed: 20101215]

34. Govindan R et al. Genomic landscape of non-small-cell lung cancer in smokers and never-smokers. Cell 150, 1121–1134 (2012). [PubMed: 22980976]
35. Lipson D et al. Identification of new *ALK* and *RET* gene fusions from colorectal and lung cancer biopsies. Nat. Med 18, 382–384 (2012). [PubMed: 22327622]
36. Xu YH & Lu S A meta-analysis of STAT3 and phospho-STAT3 expression and survival of patients with non-small-cell lung cancer. Eur. J. Surg. Oncol 40, 311–317 (2014). [PubMed: 24332948]
37. Jiang R et al. Correlation of activated STAT3 expression with clinicopathologic features in lung adenocarcinoma and squamous cell carcinoma. Mol. Diagn. Ther 15, 347–352 (2011). [PubMed: 22208386]
38. Bromberg JF et al. *Stat3* as an oncogene. Cell 98, 295–303 (1999). [PubMed: 10458605]
39. Zhang JM et al. CD109 attenuates TGF-β1 signaling and enhances EGF signaling in SK-MG-1 human glioblastoma cells. Biochem. Biophys. Res. Commun 459, 252–258 (2015). [PubMed: 25724945]
40. Ding Z et al. SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. Nature 470, 269–273 (2011). [PubMed: 21289624]
41. David CJ et al. TGF-β tumor suppression through a lethal EMT. Cell 164, 1015–1030 (2016). [PubMed: 26898331]
42. Zhong Z, Wen Z & Darnell JE Jr. Stat3: a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. Science 264, 95–98 (1994). [PubMed: 8140422]
43. Parganas E et al. Jak2 is essential for signaling through a variety of cytokine receptors. Cell 93, 385–395 (1998). [PubMed: 9590173]
44. Yu CL et al. Enhanced DNA-binding activity of a Stat3-related protein in cells transformed by the Src oncoprotein. Science 269, 81–83 (1995). [PubMed: 7541555]
45. Matsunaga Y et al. Effects of a Janus kinase inhibitor, pyridone 6, on airway responses in a murine model of asthma. Biochem. Biophys. Res. Commun 404, 261–267 (2011). [PubMed: 21111712]
46. Pedranzini L et al. Pyridone 6, a pan–Janus-activated kinase inhibitor, induces growth inhibition of multiple myeloma cells. Cancer Res. 66, 9714–9721 (2006). [PubMed: 17018630]
47. Quintás-Cardama A et al. Preclinical characterization of the selective JAK1/2 inhibitor INCB018424: therapeutic implications for the treatment of myeloproliferative neoplasms. Blood 115, 3109–3117 (2010). [PubMed: 20130243]
48. Van Rompaey L et al. Preclinical characterization of GLPG0634, a selective inhibitor of JAK1, for the treatment of inflammatory diseases. J. Immunol 191, 3568–3577 (2013). [PubMed: 24006460]
49. Tao J, Li H, Li Q & Yang Y CD109 is a potential target for triple-negative breast cancer. Tumour Biol. 35, 12083–12090 (2014). [PubMed: 25149155]
50. Zhang F et al. SWATH- and iTRAQ-based quantitative proteomic analyses reveal an overexpression and biological relevance of CD109 in advanced NSCLC. J. Proteomics 102, 125–136 (2014). [PubMed: 24667143]
51. Greenleaf WJ Assaying the epigenome in limited numbers of cells. Methods 72, 51–56 (2015). [PubMed: 25461774]
52. Denny SK et al. Nfib promotes metastasis through a widespread increase in chromatin accessibility. Cell 166, 328–342 (2016). [PubMed: 27374332]
53. Takeda K et al. Targeted disruption of the mouse *Stat3* gene leads to early embryonic lethality. Proc. Natl. Acad. Sci. USA 94, 3801–3804 (1997). [PubMed: 9108058]
54. Steeg PS Targeting metastasis. Nat. Rev. Cancer 16, 201–218 (2016). [PubMed: 27009393]
55. Tuveson DA et al. Endogenous oncogenic K-*ras*^{G12D} stimulates proliferation and widespread neoplastic and developmental defects. Cancer Cell 5, 375–387 (2004). [PubMed: 15093544]
56. DuPage M, Dooley AL & Jacks T Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat. Protoc 4, 1064–1072 (2009). [PubMed: 19561589]
57. Li H et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). [PubMed: 19505943]
58. Piskol R, Ramaswami G & Li JB Reliable identification of genomic variants from RNA–seq data. Am. J. Hum. Genet 93, 641–651 (2013). [PubMed: 24075185]

59. Li H & Durbin R Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]
60. McKenna A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010). [PubMed: 20644199]
61. Patro R, Mount SM & Kingsford C Sailfish enables alignment-free isoform quantification from RNA–seq reads using lightweight algorithms. Nat. Biotechnol 32, 462–464 (2014). [PubMed: 24752080]
62. Leek JT & Storey JD Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 3, 1724–1735 (2007). [PubMed: 17907809]
63. Smyth GK Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol 3, Article3 (2004).
64. Durinck S, Spellman PT, Birney E & Huber W Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc 4, 1184–1191 (2009). [PubMed: 19617889]
65. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550 (2005). [PubMed: 16199517]
66. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550 (2014). [PubMed: 25079552]
67. Ishwaran H & Kogalur UB Random survival forests for R. R News 7, 25–31 (2007).
68. Ishwaran H, Kogalur UB, Blackstone EH & Lauer MS Random survival forests. Ann. Appl. Stat 2, 841–860 (2008).
69. Trapnell C et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc 7, 562–578 (2012). [PubMed: 22383036]
70. Doench JG et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. 32, 1262–1267 (2014). [PubMed: 25184501]
71. Sanjana NE, Shalem 0 & Zhang F Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784 (2014). [PubMed: 25075903]
72. Morimoto-Tomita M, Ohashi Y, Matsubara A, Tsuiji M & Irimura T Mouse colon carcinoma cells established for high incidence of experimental hepatic metastasis exhibit accelerated and anchorage-independent growth. Clin. Exp. Metastasis 22, 513–521 (2005). [PubMed: 16320114]
73. Gebäck T, Schulz MM, Koumoutsakos P & Detmar M TScratch: a novel and simple software tool for automated analysis of monolayer wound healing assays. Biotechniques 46, 265–274 (2009). [PubMed: 19450233]

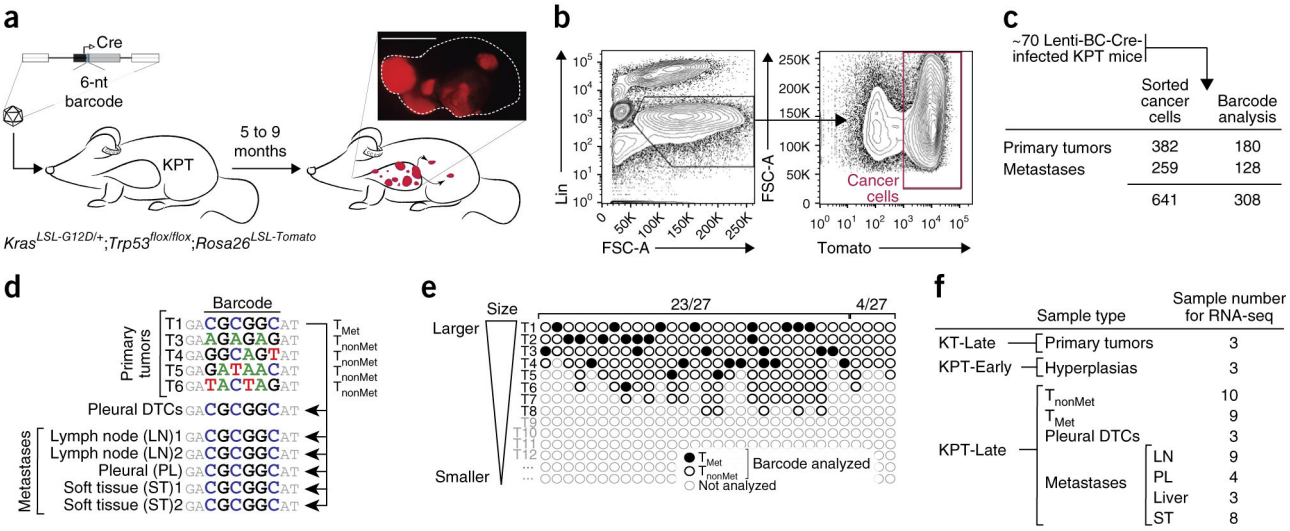


Figure 1.

Design, barcode analysis, and isolation of samples for gene expression profiling from different stages of metastatic progression. (**a**) Tumor initiation in *Kras*$^{LSL\text{-}G12D/+}$;*Trp53*$^{flox/flox}$;*Rosa26*$^{LSL\text{-}Tomato}$ (KPT) mice with a pool of barcoded Cre-expressing lentiviral vectors (Lenti-BC-Cre). Lung tumors were uniquely barcoded and stably expressed a fluorescent marker. Scale bar, 5 mm. (**b**) Representative plots showing sequential gating on dissociated tumor cells to allow FACS-based isolation of purified cancer cells. FSC-A, forward scatter. (**c**) Numbers of tumors from which cancer cells were isolated and on which barcode amplification and sequencing were performed. (**d**) Example of barcode analysis of primary tumors (T), disseminated tumor cells from the pleural cavity (DTCs), and metastases from a single mouse that identifies nonmetastatic primary tumors (T_{nonMet}) and a metastatic primary tumor (T_{Met}). (**e**) Depiction of 31 T_{Met} (black-filled circles) and 149 T_{nonMet} (black-open circles) tumors analyzed from 31 mice with metastatic lesions. Each mouse had many more tumors than we analyzed (light gray open circles). In 27 of 31 mice, we identified one or more T_{Met} tumors. In 23 of 27 mice with metastases, all of the metastases were from a single T_{Met} tumor. In four mice, we identified metastases with two unique barcodes and also found both T_{Met} tumors (mice with two black filled-in circles). (**f**) Samples for RNA-seq-based gene expression analysis. KT-Late samples are tumors from *Kras*$^{LSL\text{-}G12D/+}$;*Rosa26*$^{LSL\text{-}Tomato}$ (KT) mice 6–7 months after tumor initiation. KPT-Early (KPT-E) samples are neoplastic cells from KPT mice 10 weeks after tumor initiation. LN, lymph node; PL, pleura; ST, soft tissue.

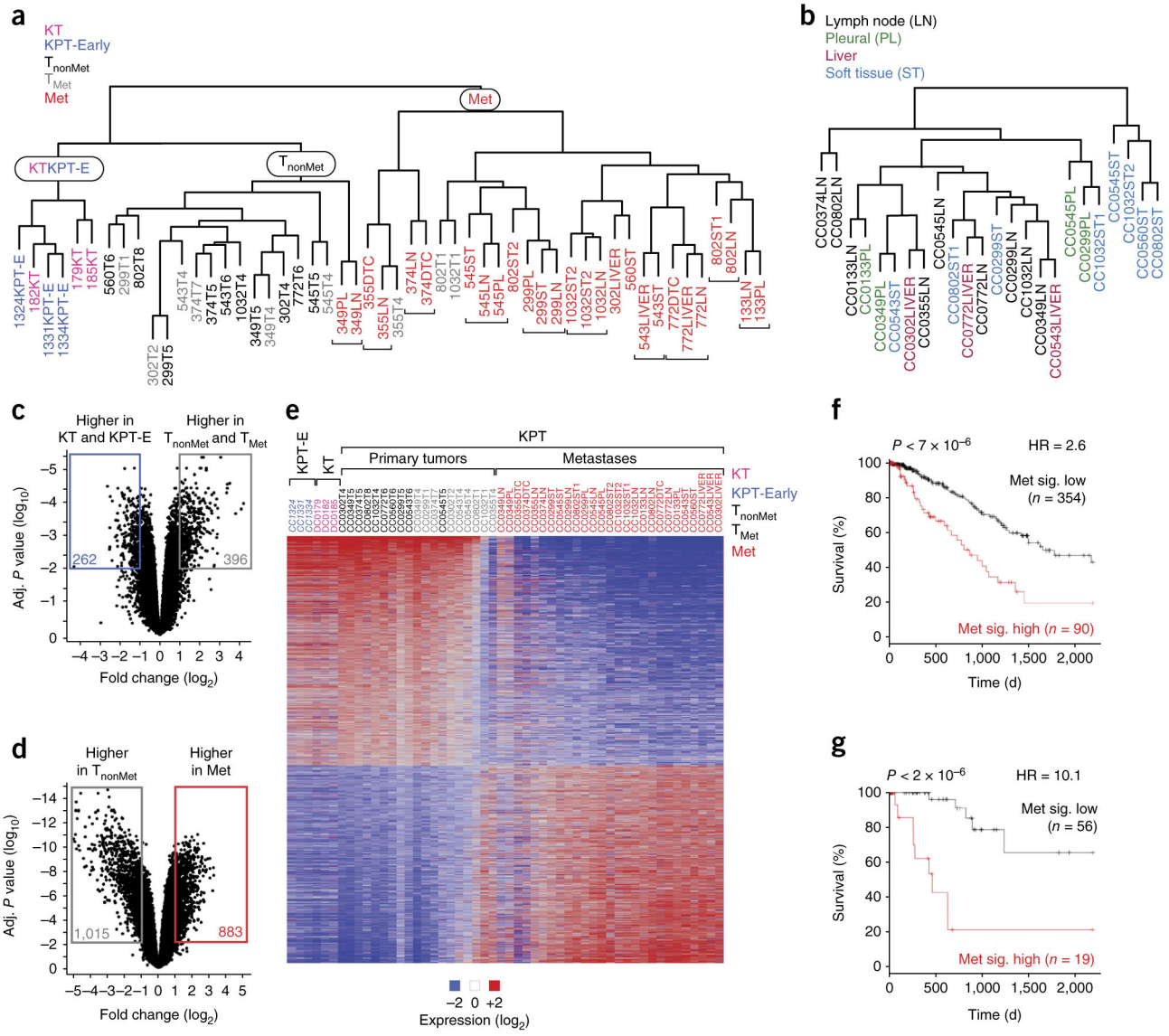


Figure 2.

Lung tumors undergo stepwise changes in gene expression programs during metastatic progression. (**a**) Clustering analysis of all samples identified three major groups: one with only KT and KPT-E samples, one with all T_{nonMet} samples, and one with almost all (>90%) metastases. Adjacent clonally related metastases are indicated with brackets. (**b**) Clustering of metastases samples after the gene expression differences that drive clustering of clonally related metastases are removed. (**c,d**) Robust and conserved changes in gene expression between KT and KPT-E versus primary tumor samples (T_{nonMet} and T_{Met}) (**c**) and between T_{nonMet} and Met samples (**d**). The number of genes with >2-fold difference and an adjusted $P < 0.01$ between groups is indicated. (**e**) Heat map of genes that are differentially expressed (fold change > 2, adjusted $P < 0.01$) between T_{nonMet} and Met samples. These samples were not clustered on the basis of their gene expression but rather were arranged by group. For all primary tumors, we dissociated the entire tumor before FACS-based purification of the Tomato$^+$ cancer cells (as opposed to only isolating cells from one part of the tumor).

Therefore, the gene expression profile is an average of the entire tumor. Two clonally related metastases 349LN and 349PL were unlike the other metastases. (**f,g**) Outcome of patients with lung adenocarcinoma after stratification based on a gene expression signature generated by comparing T_{nonMet} to Met samples (Met sig.). Analysis of 444 unselected patients (**f**) and 75 patients with KRAS-mutant tumors (**g**) from TCGA. *P* values, hazard ratios (HR), and number of patients in each group are indicated.

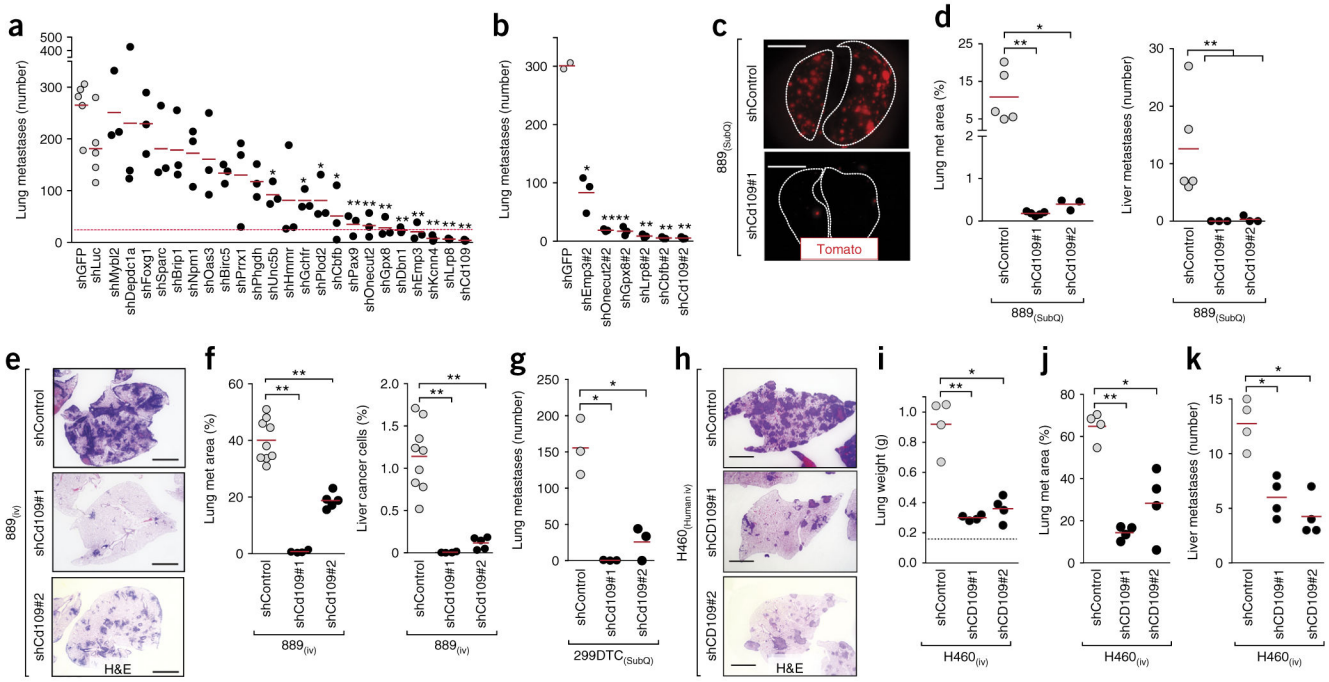
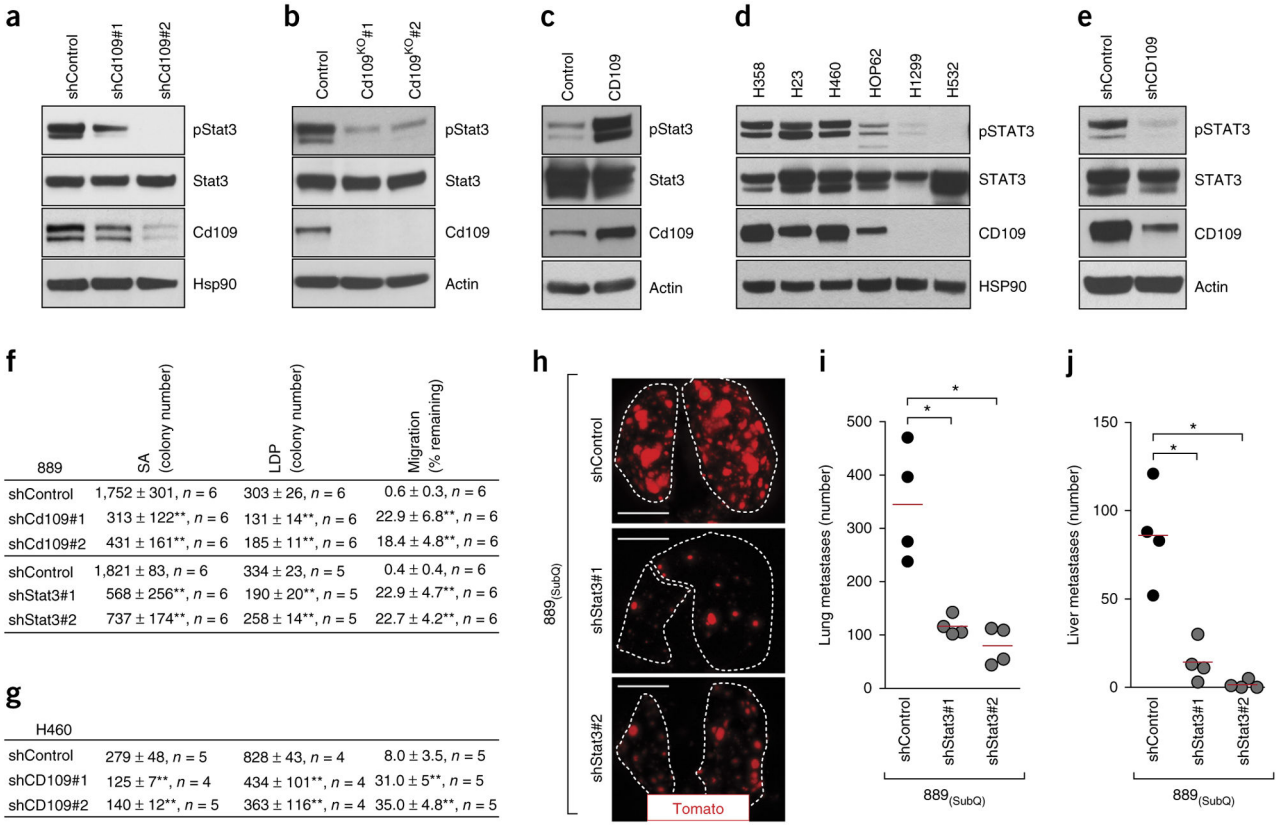


Figure 3.

In vivo functional screening identifies CD109 as a driver of metastatic ability. (**a,b**) Effect of stable knockdown of the expression of each candidate gene on metastatic ability (**a**) and confirmation with a second shRNA specific for six genes of interest (**b**). The dashed pink line represents a 90% reduction in metastatic ability relative to that of control shGFP-expressing and shLuciferase (shLuc)-expressing tumors. The number of metastases in mice with subcutaneous tumors of each genotype is shown. Each dot represents a mouse, and the bar denotes the mean. $*P < 0.05$ and $**P < 0.01$. (**c**) Representative images, taken on a fluorescence dissecting scope, of the lungs from mice with shControl- or shCd109-expressing subcutaneous (SubQ) tumors. Lung lobes are outlined with a white dashed lines. Scale bars, 5 mm. (**d**) Lung metastasis area (left) and liver metastases number (right) in mice with SubQ tumors expressing shControl or shCd109. Each dot represents a mouse, and the bar denotes the mean. $*P < 0.05$ and $**P < 0.01$. (**e**) Representative lung histology 12 d after intravenous (i.v.) transplantation of 889 cells expressing a control shRNA or either of two shRNAs specific for *Cd109*. H&E, hematoxylin and eosin. Scale bars, 2 mm. (**f**) Lung metastasis area quantified by histology (left) and quantification of micrometastatic cancers in the liver by flow cytometry (right). Each dot represents a mouse, and the bar denotes the mean. $**P < 0.001$. (**g**) Lung metastasis area in mice with shControl- or shCd109-expressing SubQ tumors generated from a second mouse lung adenocarcinoma cell line (299DTC). Each dot represents a mouse, and the bar denotes the mean. $*P < 0.01$. (**h**) Representative lung histology 28 d after i.v. transplantation of shControl- or shCD109-expressing H460 human lung cancer cells. Scale bars, 2 mm. (**i**–**k**) Lung weight (**i**), lung metastasis area (**j**), and number of liver metastases (**k**) after i.v. transplantation of shControl- or shCD109-expressing H460 cells. Normal lung weight is indicated by a dashed line (**i**). Each dot represents a mouse, and the bar denotes the mean. $*P < 0.01$ and $**P < 0.001$. All *P* values were calculated by an unpaired Student's *t*-test.



f

889	SA (colony number)	LDP (colony number)	Migration (% remaining)
shControl	1,752 ± 301, *n* = 6	303 ± 26, *n* = 6	0.6 ± 0.3, *n* = 6
shCd109#1	313 ± 122**, *n* = 6	131 ± 14**, *n* = 6	22.9 ± 6.8**, *n* = 6
shCd109#2	431 ± 161**, *n* = 6	185 ± 11**, *n* = 6	18.4 ± 4.8**, *n* = 6
shControl	1,821 ± 83, *n* = 6	334 ± 23, *n* = 5	0.4 ± 0.4, *n* = 6
shStat3#1	568 ± 256**, *n* = 6	190 ± 20**, *n* = 5	22.9 ± 4.7**, *n* = 6
shStat3#2	737 ± 174**, *n* = 6	258 ± 14**, *n* = 5	22.7 ± 4.2**, *n* = 6

g

H460			
shControl	279 ± 48, *n* = 5	828 ± 43, *n* = 4	8.0 ± 3.5, *n* = 5
shCD109#1	125 ± 7**, *n* = 4	434 ± 101**, *n* = 4	31.0 ± 5**, *n* = 5
shCD109#2	140 ± 12**, *n* = 5	363 ± 116**, *n* = 4	35.0 ± 4.8**, *n* = 5

Figure 4.

CD109 regulates Stat3 activity to drive malignant cellular phenotypes and metastatic ability. (**a,b**) Representative western blot analysis (of $n = 2$) for Tyr705 phosphorylation of Stat3 (pStat3) in shControl- or shCD109-expressing 889 cells (**a**) or in parental 889 cells (control) or 889 cells in which *Cd109* was knocked out (KO) (**b**). Hsp90 and actin were used as loading controls. (**c**) Representative western blot analysis (of $n = 2$) for pStat3 on a mouse embryonic fibroblast cell line stably expressing Cd109. Actin was used as a loading control. (**d**) Representative western blot analysis (of $n = 2$) for pSTAT3 and CD109 on a panel of human lung adenocarcinoma cell lines. HSP90 was used as a loading control. (**e**) Representative western blot analysis for (of $n = 2$) pSTAT3 and CD109 on human H460 cells without (control) or with *CD109* knockdown. Actin was used as a loading control. (**f**) Clonal growth and migration of 889 cells without (control) or with either *Cd109* knockdown or *Stat3* knockdown. Clonal growth was assessed in anchorage-independent conditions in soft agar (SA) and in tissue culture low-density-plating (LDP) conditions. Migration was assessed using a scratch assay, and the percentage of gap remaining after 16 h is indicated. Data from two independent experiments are presented as mean ± s.d. The n value per group is indicated. $^{**}P < 0.001$. (**g**) Clonal growth and migration of H460 cells without (control) or with *CD109* knockdown. Data from two independent experiments are presented as mean ± s.d. The n value per group is indicated. $^{**}P < 0.001$. (**h**) Representative fluorescence images showing the lung metastases from mice with SubQ tumors in which *Stat3* was knocked down or not (control). Lung lobes are outlined with a white dotted line. Scale bars, 5 mm. (**i,j**) Number of lung (**i**) and liver (**j**) metastases in mice without (control) or with

Stat3 knockdown in the SubQ tumors. Each dot represents a mouse, and the bar denotes the mean. $*P < 0.01$. All P values were calculated by unpaired Student's t-test.

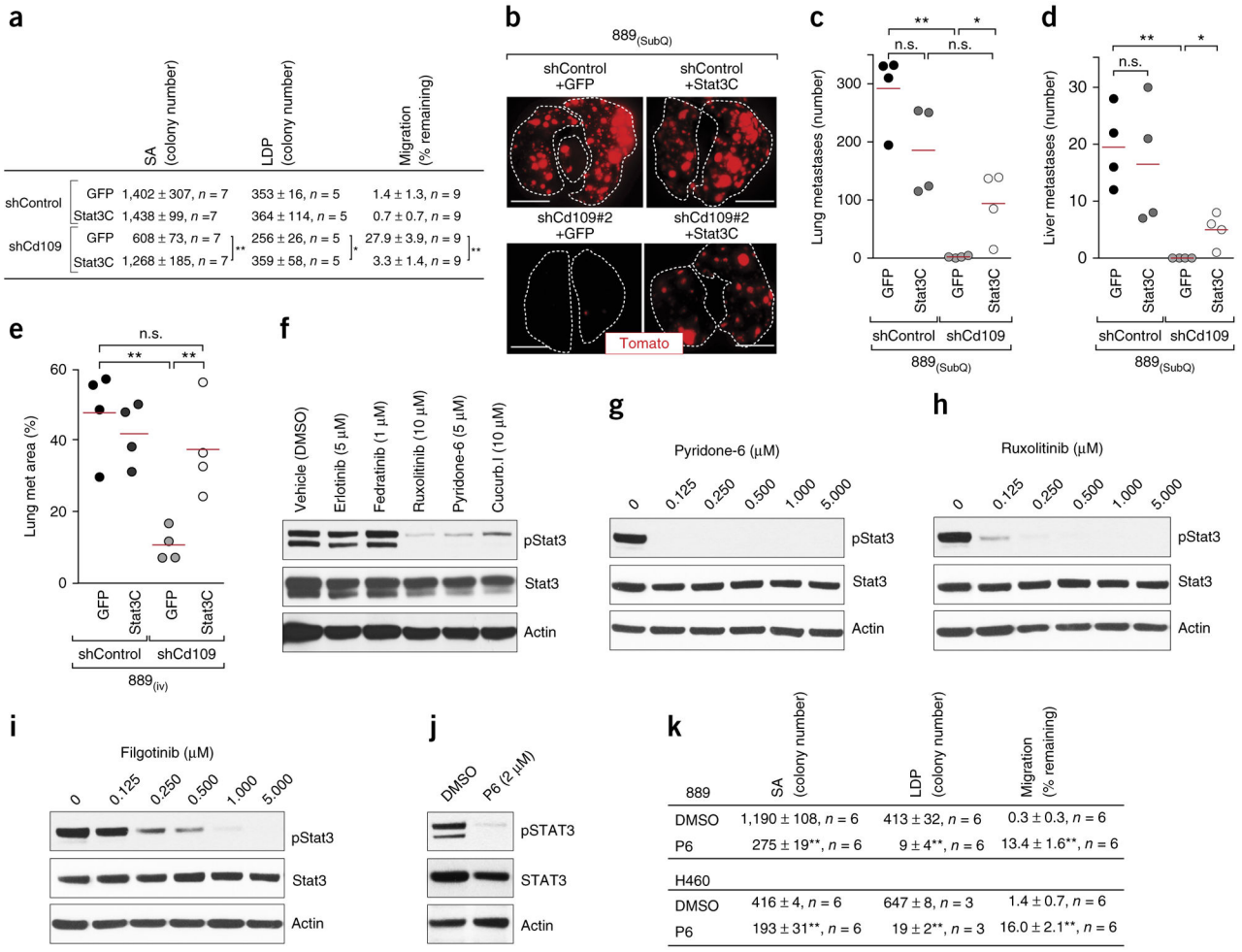


Figure 5.

Jak–Stat3 signaling is a critical pro-metastatic effector of Cd109. (**a**) Clonal growth and migration of shControl- or shCd109-expressing 889 cells without (GFP) or with expression of hyper-activatable Stat3 (Stat3C). Data from two independent experiments are presented as mean ± s.d. The *n* value per group is indicated. **P* < 0.01 and ***P* < 0.001. (**b**) Representative fluorescence images of the lungs from mice with subcutaneous (SubQ) tumors from the indicated cell lines. Scale bars, 5 mm. (**c,d**) Numbers of lung (**c**) and liver (**d**) metastases from SubQ tumors. Each dot represents a mouse, and the bar denotes the mean. **P* < 0.05 and ***P* < 0.01. (**e**) Lung metastasis area, as quantified by histology, in mice after intravenous (i.v.) transplantation of the indicated cell lines. Each dot represents a mouse, and the bar denotes the mean. **P* < 0.05 and ***P* < 0.01; n.s., not significant. (**f**) Representative western blot analysis (of *n* = 2) for pStat3 in 3T3-Cd109 cells that were treated with DMSO (vehicle control), an epidermal growth factor receptor inhibitor (erlotinib), a Jak2 inhibitor (fedratinib) or the broadly acting Jak kinase inhibitors ruxolitinib, pyridone 6 and cucurbitacin I. Actin was used as a loading control. (**g**–**i**) Western blot analyses for pStat3 in 889 cells that were treated with a titration of pyridone 6 (**g**), ruxolitinib (**h**) or filgotinib (**i**). Actin was used as a loading control. (**j**) Representative western blot analysis (of *n* = 2) for pSTAT3 in human H460 cells that were treated with 2

μM pyridone 6 (P6). Actin was used as a loading control. (**k**) Clonal growth and migration ability of 889 and H460 cells that were treated with 2 μM pyridone 6. Data from two independent experiments are presented as mean ± s.e.m. $^{**}P < 0.001$. All *P* values were calculated by an unpaired Student's *t*-test.

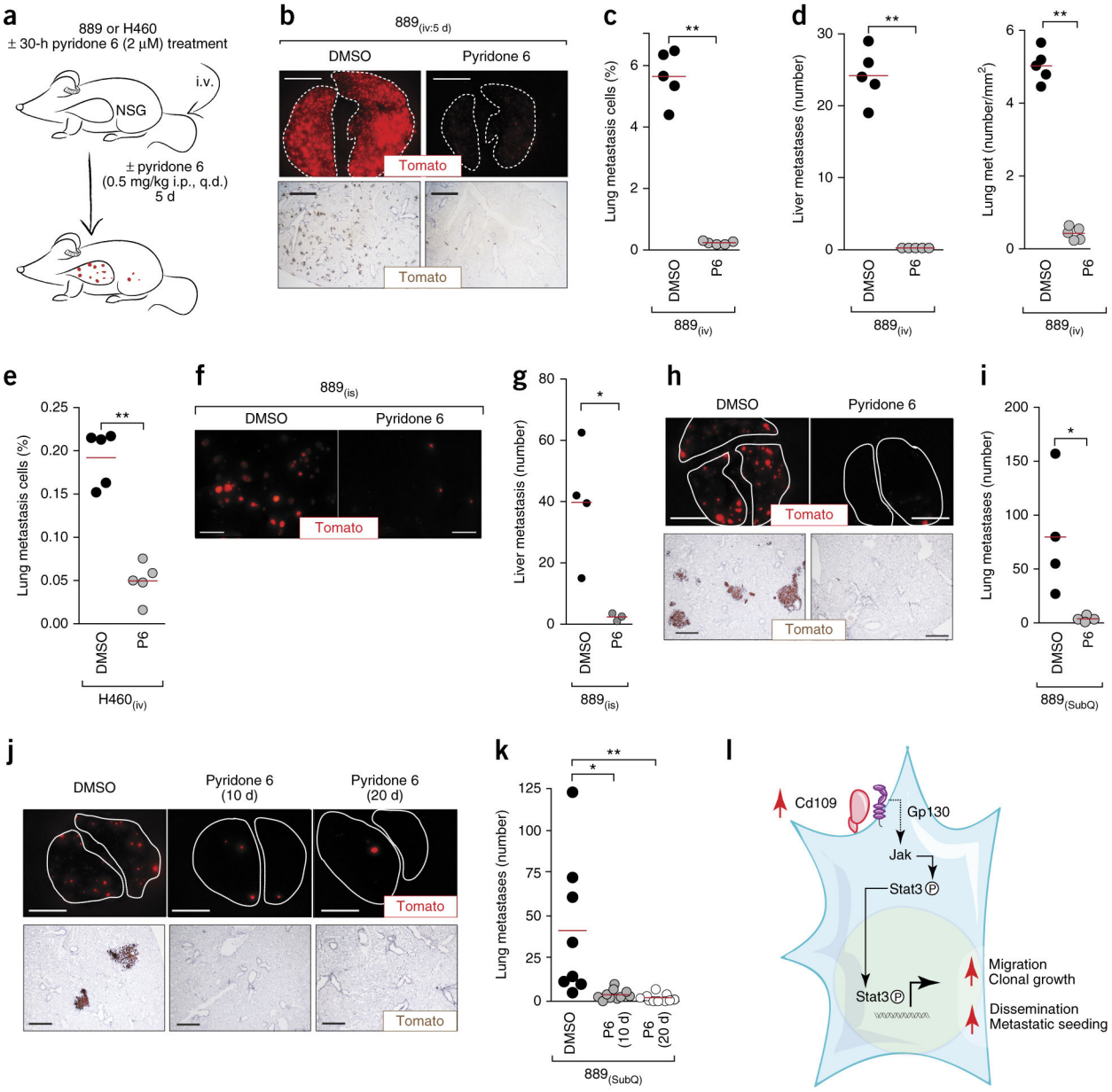


Figure 6.

Pharmacological inhibition of Jak–Stat signaling inhibits metastatic ability of lung adenocarcinoma cells. (**a**) Outline of *in vivo* Jak kinase inhibitor (pyridone 6) treatments. Immunodeficient NOD–Scid–*Il2rg*-deficient (NSG) mice were injected intravenously (i.v.) with mouse 889 or human H460 cells ± pyridone 6 pretreatment. Mice were treated daily with pyridone 6 or vehicle (DMSO), administered intraperitoneally (i.p.). (**b**) Representative fluorescence (top) and immunohistochemistry (bottom) images of Tomato expression in the lungs of NSG mice 5 d after i.v. transplantation of 889 cells. Scale bars, 5 mm (top) and 1 mm (bottom). (**c**) Quantification of cells in lung metastases by flow cytometry. $^{**}P < 0.001$. (**d**) Quantification of Tomato^{+} micro-metastases by direct fluorescence imaging and histology in liver (left) and lung (right) of recipient mice. $^{**}P < 0.001$. (**e**) Quantification of

cells in the lung 5 d after i.v. transplantation of human H460 cells by flow cytometry. (**f**) Representative fluorescence images of liver metastases 12 d after intrasplenic (i.s.) transplantation of 889 cells. Scale bars, 2 mm. (**g**) Quantification of liver metastases by direct counting. $*P < 0.01$. (**h**) Representative fluorescence (top) and histology (bottom) images of the lungs from mice with SubQ 889 tumors after treatment with or without pyridone 6. Lung lobes are outlined with a dashed line. Scale bars, 5 mm (top) and 500 μm (bottom). (**i**) Quantification of lung metastases by direct counting. $*P < 0.05$. (**j**) Representative fluorescence (top) and histology (bottom) images of the lungs from mice in which SubQ tumors were removed at day 10 and metastases were assessed at day 20, with Jak kinase inhibition for either the first 10 d or the entire 20 d. Lung lobes are outlined with a dashed line. Scale bars, 5 mm (top) and 500 μm (bottom). (**k**) Quantification of metastases by direct counting. $*P < 0.05$ and $**P < 0.01$. (**I**) Model of the Cd109-driven metastatic state. In all plots, each dot represents a mouse, and the bar denotes the mean. All *P* values were calculated by an unpaired Student's *t*-test.