Asparagine bioavailability governs metastasis in a model of breast cancer

Simon R. V. Knott[1,2,3,*], **Elvin Wagenblast**[2,4,5,*], **Showkhin Khan**[2,6], **Sun Y. Kim**[2], **Mar Soto**[2], **Michel Wagner**[7], **Marc-Olivier Turgeon**[7], **Lisa Fish**[8,9,10], **Nicolas Erard**[1], **Annika L. Gable**[2], **Ashley R. Maceli**[2], **Steffen Dickopf**[2], **Evangelia K. Papachristou**[1], **Clive S. D'Santos**[1], **Lisa A. Carey**[11], **John E. Wilkinson**[12], **J. Chuck Harrell**[13], **Charles M. Perou**[14], **Hani Goodarzi**[8,9,10], **George Poulogiannis**[7,15], and **Gregory J. Hannon**[1,2,6]

[1]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 ORE, UK

[2]Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

[3]Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048, USA

[4]Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, Ontario M5G 1L7, Canada

[5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5G 1L7, Canada

[6]New York Genome Center, 101 6th Avenue, New York, New York 10013, USA

[7]Division of Cancer Biology, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

[8]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94158, USA

[9]Department of Urology, University of California, San Francisco, San Francisco, California 94158, USA

Correspondence and requests for materials should be addressed to G.J.H. (Greg.Hannon@cruk.cam.ac.uk).
*These authors contributed equally to this work.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Supplementary Information is available in the online version of the paper.

Author Information The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer Information *Nature* thanks R. Agami and the other anonymous reviewer(s) for their contribution to the peer review of this work.

[10]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California 94158, USA

[11]Division of Hematology and Oncology, University of North Carolina at Chapel Hill, 170 Manning Drive, CB7305, Chapel Hill, North Carolina 27599, USA

[12]Department of Pathology, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA

[13]Department of Pathology, Virginia Commonwealth University, Richmond, Virginia 23284, USA

[14]Department of Genetics and Pathology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

[15]Division of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, London SW7 2AZ, UK

Abstract

Using a functional model of breast cancer heterogeneity, we previously showed that clonal sub-populations proficient at generating circulating tumour cells were not all equally capable of forming metastases at secondary sites[1]. A combination of differential expression and focused *in vitro* and *in vivo* RNA interference screens revealed candidate drivers of metastasis that discriminated metastatic clones. Among these, asparagine synthetase expression in a patient's primary tumour was most strongly correlated with later metastatic relapse. Here we show that asparagine bioavailability strongly influences metastatic potential. Limiting asparagine by knockdown of asparagine synthetase, treatment with L-asparaginase, or dietary asparagine restriction reduces metastasis without affecting growth of the primary tumour, whereas increased dietary asparagine or enforced asparagine synthetase expression promotes metastatic progression. Altering asparagine availability *in vitro* strongly influences invasive potential, which is correlated with an effect on proteins that promote the epithelial-to-mesenchymal transition. This provides at least one potential mechanism for how the bioavailability of a single amino acid could regulate metastatic progression.

Most women with breast cancer do not succumb to their primary tumour but instead to metastases that become apparent after the primary lesion has been removed. For cells to contribute to metastases, they must leave the primary site, enter the vasculature, survive in the blood, and then extravasate and colonize secondary sites. Our previous studies of a mouse model of breast tumour heterogeneity identified two clonal 4T1 sub-lines with a strong propensity to form circulating tumour cells (CTCS) through a non-invasive mechanism requiring vascular mimicry (4T1-E and -T)[1,2]. These two clones differed in their ability to form metastases, with 4T1-T preferentially colonizing brain, liver, and lungs. The distinction between the metastatic potential of the two CTC-forming clones offered the opportunity to identify drivers of metastasis, which exert their effects late in metastatic progression.

To validate the observation that 4T1-T had greater metastatic potential among CTC-proficient clones, we combined equal numbers of 4T1-E and -T cells and introduced these directly into the bloodstream of immune-compromised recipients (NOD-SCID-*Il2rg*[−/−]

(NSG) mice). Although the two clones were initially present in equal abundance (Fig. 1a), when cells were collected from the lung at day 7, clone T predominated, with its relative representation being inversely correlated with the total number of cells injected.

We identified 192 genes with higher expression in 4T1-T than 4T1-E cells (Supplementary Table 1). Their corresponding Gene Ontology terms were enriched for processes important for metastatic spread (Supplementary Table 2; for example, cell migration and locomotion). A retrospective analysis of patient data showed that genes within the set are more highly expressed in aggressive breast tumour subtypes[3] (Extended Data Fig. 1a). They were also more highly expressed in the primary tumours of patients with later relapse to the bone, brain, and lungs compared with primary tumours of relapse-free survivors (Extended Data Fig. 1b for lung).

To identify metastatic drivers, we performed an RNA interference (RNAi) screen, with two arms (Fig. 1b). In total, 26 pools of approximately 40 short hairpin RNAs (shRNAs), targeting protein-coding members of the 192-gene set, were introduced into 4T1-T cells[4]. These were collected onto Matrigel or introduced into NSG mice by tail vein injection. After 24 h, the cells that had invaded through the Matrigel were collected and, after 7 days, lungs were collected from the mice. Using high-throughput sequencing, we identified shRNAs that were depleted from the invaded cell populations or lung metastases, presumably because they targeted genes important for these processes. Strong overlap was observed when the *in vitro* and *in vivo* candidates were compared (Fig. 1c and Supplementary Table 3).

Of the 11 candidate genes that scored in both the *in vitro* and *in vivo* assays, asparagine synthetase (Asns) had the most robust clinical evidence supporting its relevance to disease progression (Extended Data Fig. 1c) Expression levels of the human orthologue, ASNS, were predictive of general and lung-specific relapse in two datasets of patients with breast cancer. Also, when a small collection of matched tumour and lung metastases were transcriptionally profiled, ASNS was found to be more highly expressed in secondary lesions. ASNS is more highly expressed in aggressive tumour subtypes (Extended Data Fig. 1d) and it is more highly expressed in patients with relapse to the lymph node, brain, liver, and lungs compared with relapse-free survivors (Extended Data Fig. 1e). Subsequent analyses identified ASNS as predictive of survival in three additional datasets of patients with breast cancer (Extended Data Fig. 1f). In addition to breast tumours, ASNS is negatively correlated with survival in four out of the ten other solid tumours in The Cancer Genome Atlas (TCGA) Pan-Cancer dataset (Extended Data Fig. 1g) and is a globally predictive biomarker for solid tumours (Extended Data Fig. 1h). Previous research has shown that asparagine is an important regulator of cellular metabolism and adaptation[5,6].

To validate Asns as a metastatic driver, we infected 4T1-T cells with two shRNAs targeting Asns or a control and introduced these cells intravenously into NSG mice (Supplementary Table 4). Asns-silenced cells produced significantly fewer lung metastases (Fig. 2a and Extended Data Fig. 2a). Asns-silenced cells also showed poor invasion into Matrigel (Fig. 2b and Extended Data Fig. 2b). Silencing Asns did affect proliferation *in vitro;* however, this defect was minor compared with that observed in the invasion assay (Extended Data Fig. 2c, d). Intracellular free asparagine was reduced by silencing of Asns in 4T1-T cells (Extended

Data Fig. 2e), and the abilities to invade and proliferate were increased in Asns-silenced cells when the culture medium was supplemented with asparagine (Extended Data Fig. 2f, g).

When Asns-silenced cells were injected into the mammary fat pad, no significant change in primary tumour formation was observed (Fig. 2c), yet CTCs and lung metastases were reduced (Fig. 2d, e). Although statistically insignificant, metastases initiated by silenced cells were noticeably smaller, hinting at a growth defect at the metastatic site (Fig. 2f). Similar results were obtained with Asns-silenced parental 4T1 cells, indicating that Asns dependency is not a peculiarity of a single clonal line (Extended Data Fig. 3a, b). Enforced Asns expression in parental 4T1 populations did not affect primary tumour growth but did increase metastases both in number and in size (Extended Data Fig. 3c–e). Similar outcomes were observed upon enforced ASNS expression in human MDA-MB-231 breast cancer cells (Extended Data Fig. 3f–i).

To determine whether the observed effects were unique to asparagine, we supplemented the medium separately with other nonessential amino acids (NEAAs) lacking in the culture medium, or with glycine, which is present and acted as a negative control, and assayed cells for invasiveness. 4T1 cells responded uniquely to asparagine supplementation, with an approximately twofold increase in invasiveness (Fig. 3a), although levels of uptake were similar for each of the amino acids with the exception of aspartic and glutamic acids (Extended Data Fig. 4a). More profound effects were observed with MDA-MB-231 cells (Extended Data Fig. 4b, c). Growth was not affected by asparagine supplementation for either cell line during the same period (Extended Data Fig. 4d, e).

Because invasiveness could be modulated either by altering asparagine biosynthetic capacity or by modifying extracellular pools, we investigated whether metastasis could be influenced by treatment with L-asparaginase. This enzyme is used to treat acute lymphoblastic leukaemia, which is generally highly dependent on extracellular asparagine[7,8]. L-Asparaginase has proved ineffective for treating solid tumours, consistent with Asns silencing not affecting growth at the primary site[9]. NSG mice harbouring orthotopic 4T1 tumours were treated with 60 U L-asparaginase 5 times per week for 19 days, reducing serum asparagine to undetectable levels (Supplementary Table 5). Although no significant difference was detected in primary tumours, compared with controls, a reduction in metastasis was observed (Extended Data Fig. 5a–c).

In TEL-AML1-negative patients with acute lymphoblastic leukaemia, resistance to L-asparaginase is thought to be achieved through increased biosynthetic production of asparagine[10]. When Asns-silenced cells were injected orthotopically into mice treated with L-asparaginase, metastases were nearly undetectable (Fig. 3b and Extended Data Fig. 5d). In this case, a reduction of primary tumour volume was also observed (Extended Data Fig. 5e). Similar results were obtained with ASNS-silenced MDA-MB-231 cells (Extended Data Fig. 5f, g).

The availability of extracellular asparagine can also be manipulated by altering asparagine levels in the diet. shRNA-infected 4T1-T cells were orthotopically injected into mice that

received a control, low-asparagine, or high-asparagine chow (0.6%, 0%, and 4%, respectively). High-performance liquid chromatography (HPLC) confirmed that serum asparagine levels were significantly altered in concordance with dietary intake (Extended Data Fig. 6a). Asparagine restriction did not affect primary tumour growth, regardless of Asns expression status (Extended Data Fig. 6b). By contrast, metastatic burden was decreased in animals that were fed low-asparagine diets and increased in animals given high-asparagine diets (Fig. 3c and Extended Data Fig. 6c). Metastases were nearly undetectable in mice that were injected with Asns-silenced cells and fed a low-asparagine diet. Similar results were obtained when parental 4T1 cells were orthotopically injected into animals fed these same asparagine-controlled diets (Extended Data Fig. 6d–f).

Analyses of the mammary gland, serum, and lungs of mice by mass spectrometry suggested that, under normal physiological conditions, asparagine levels are highest in mammary gland and lowest in serum (Fig. 3d). High asparagine availability in the mammary gland might blunt the effect of Asns silencing or changes in global asparagine levels on primary tumour growth, whereas low levels in the serum may make CTCs susceptible to these manipulations. Overall, asparagine abundance in tissues correlated with Asns expression (Extended Data Fig. 6g). ASNS expression levels follow a similar pattern across human tissues (Extended Data Fig. 6h), raising the possibility of similar effects if asparagine levels were altered in patients[11].

To understand the mechanism by which asparagine availability might affect invasion and metastasis, we examined expression changes induced by Asns silencing, at both the RNA and protein levels. RNA measurements were the strongest predictor of protein-level changes (Extended Data Fig. 7a). Consistent with a previous report of translational pausing at asparagine residues in L-asparaginase-treated cells, we also found asparagine content to be predictive of corresponding protein-level changes[12] (Extended Data Fig. 7a), irrespective of whether they were normalized for RNA levels (Extended Data Fig. 7b).

Among the asparagine-enriched proteins that were depleted on Asns silencing, we found genes whose human orthologues were upregulated after induction of the epithelial-to-mesenchymal transition (EMT)[13] (Fig. 4a, Extended Data Fig. 7c and Supplementary Table 6; EMT-up proteins). Overall, depleted proteins had 10% higher asparagine content than the analysed proteome as a whole, whereas EMT-up proteins had 19% higher asparagine content (Extended Data Fig. 7c). Human EMT-up proteins were also asparagine-enriched (Extended Data Fig. 7c). A re-analysis of existing ribosomal profiling data revealed high rates of pausing at asparagine residues within EMT-up genes in L-asparaginase-treated human prostate cancer cells (Extended Data Fig. 7d, e), and these same proteins increased in expression when 4T1-T cells were cultured in elevated asparagine[12] (Fig. 4b). Asparagine enrichment is a globally conserved property of EMT-up proteins, with enrichment being greatest in mammals (Extended Data Fig. 8).

EMT-up genes were also downregulated at the transcriptional level in Asns-silenced cells, and they were increased in their RNA levels when these cells were grown in asparagine-supplemented medium (Extended Data Fig. 9a, b). A re-analysis of existing data also showed reduced expression of EMT-up genes when ATF4, which regulates ASNS

transcription, was deleted in haploid cells (Extended Data Fig. 9c), and liver cells from L-asparaginase-treated ATF4-knockout mice were more perturbed in their EMT program than were similarly treated wild-type mice (Extended Data Fig. 9d)[14,15]. Considered together, these data suggest that asparagine bioavailability might affect metastasis, at least in part, through regulation of EMT.

To examine the role of EMT in metastasis in our model, we orthotopically injected 4T1-T cells in which we had silenced expression of transforming growth factor-β (Tgf-β, also known as Tgfb1), a key driver of EMT[16] (Supplementary Table 4). Primary tumour growth was unaffected by this manipulation (Extended Data Fig. 9e); however, the expression of two prototypical EMT markers (Twist1 and E-cadherin) were altered to indicate a perturbed EMT program (Extended Data Fig. 9f, g). Tgf-β-silenced cells produced fewer metastases from the primary tumour or when intravenously injected (Extended Data Fig. 9h, i).

Although no differences were detected in haematoxylin and eosin (H&E)-stained tumour sections, morphological distinctions were noticeable when shRNA-infected 4T1-T cells were isolated from primary tumours by 6-thioguanine selection, with most Asns-silenced cells displaying an epithelial morphology (Extended Data Fig. 10a, b). Most 6-thioguanine-isolated metastatic cells displayed a mesenchymal morphology, regardless of Asns expression status, and this was matched by an increase in the expression of EMT-up genes (Fig. 4c). Nevertheless, EMT-up genes were downregulated in Asns-silenced versus -expressing metastatic cells, indicating an increased representation of epithelial cells in the silenced populations. Similarly, EMT-down genes were upregulated, and Twist and E-cadherin expression measurements indicated a higher epithelial representation in Asns-silenced primary tumour cell populations (Fig. 4d). These results were validated by qPCR for Twist and E-cadherin in Asns-silenced and -expressing cells that were isolated by fluorescence-activated cell sorting from primary and secondary lesions (Extended Data Fig. 10c, d).

Staining for Twist and E-cadherin proteins confirmed that EMT is perturbed in Asns-silenced tumours, and this same pattern was observed in the corresponding metastases (Fig. 4e and Extended Data Fig. 11a–c). Similar patterns were observed in the primary tumours of mice that had been treated with L-asparaginase (Extended Data Fig. 11d, e) or subjected to dietary asparagine restriction (Extended Data Fig. 11f, g).

Our model of breast tumour heterogeneity has strongly implicated asparagine bioavailability as a regulator of metastatic progression. This also seems to be relevant in human cancers, as high ASNS expression is a marker of poor prognosis for many tumour types. One mechanism underlying our findings is a probable link between asparagine bioavailability and EMT, which can be observed *in vitro* and *in vivo*. In our breast cancer model, the gating step probably occurs in the circulation, where asparagine levels are low and are strongly affected by either L-asparaginase treatment or dietary restriction. Nonetheless, we do see effects on ratios of epithelial- to mesenchymal-like tumour cells at the primary and secondary sites, which could also affect both tumour progression and response to therapy.

METHODS

Cell culture

The mouse mammary tumour cell line 4T1 (American Type Culture Collection, ATCC) and any derived clonal cell line were cultured in DMEM high glucose supplemented with 5% fetal bovine serum, 5% fetal calf serum, MEM NEAA, and penicillin-streptomycin (Thermo Fisher Scientific). The human breast tumour cell line MDA-MB-231 (ATCC) was cultured in DMEM high glucose supplemented with 10% fetal bovine serum, NEAA and penicillin-streptomycin (Thermo Fisher Scientific). The 4T1 and MDA-MB-231 cell lines were tested and authenticated by ATCC. The Platinum-A (Cell BioLabs) and 293FT (Thermo Fisher Scientific) packaging cell lines for virus production were cultured in DMEM high glucose supplemented with 10% fetal bovine serum and penicillin-streptomycin. All cell lines were routinely tested for mycoplasma contamination.

Virus production

Retroviral vectors were packaged using the Platinum-A (Cell BioLabs) cell line and lentiviral vectors were packaged using the 293FT cell line (Thermo Fisher Scientific) as previously described[1].

Animal studies

All mouse experiments were approved by the Cold Spring Harbour Animal Care and Use Committee. The maximal permitted tumour size of 20 mm in any direction was never exceeded. All mouse injections were performed with 6- to 8-week-old female NOD-SCID-$Il2rg^{-/-}$ (NSG) mice (JAX). Balb/c mice were not used in this study as the different clonal cell lines have variable GFP levels due to the lentiviral barcode vector. Tail vein injections were performed using 5×10^5 mouse mammary tumour cells, which were re-suspended in 100 μl of PBS and injected via the tail vein. Orthotopic injections were performed using 1×10^5 mouse mammary tumour cells or 5×10^5 MDA-MB-231 cells. For this, cells were re-suspended in a 1:1 mix of PBS and growth-factor-reduced Matrigel (BD Biosciences). A volume of 20 μl was injected into mammary gland number 4 for mouse mammary tumour cells and a volume of 40 μl was injected for MDA-MB-231 cells. Primary tumour volume was measured using the formula $V = \frac{1}{2}(L \times W^2)$, in which L is length and W is width of the primary tumour. For L-asparaginase studies, mice were administered 200 μl of 60 U L-asparaginase (Abcam) five times per week through intraperitoneal injections. For L-asparagine-adjusted diets, mice were given a control amino-acid diet (0.6% asparagine), an asparagine-deficient diet (0% asparagine), or an asparagine-rich diet (4% asparagine). All diets were isonitrogenous and contained similar calorie densities. Sample size was chosen to give sufficient power for calling significance with standard statistical tests. Tumour and metastasis experiments were performed with ten animals per condition to account for the variability observed in such *in vivo* experiments. Animals were assigned to treatment groups through random cage selection. Mice were excluded from the analysis if the primary tumour engrafted into the peritoneum. All image quantifications were performed in a blinded manner. Source data are available for tumour volume measurements and lung metastases.

Barcode analysis

The barcodes of the 4T1-E and 4T1-T cells were amplified and sequenced as previously described[1].

In vivo shRNA lung screen and *in vitro* invasion screen

shRNAs were predicted on the basis of the Sherwood algorithm described in ref. 4. Pools of approximately 50 shRNAs were packaged in Platinum-A cells. For each pool, 10 million 4T1- T cells were infected at a multiplicity of infection of 0.3. The infected cells were selected with 500 μg ml^{-1} hygromycin for 5 days and each pool was injected into five mice each via the tail vein. Two pre-injection pools were collected at the time of injection. After 7 days, mice were euthanized and perfused with PBS to remove blood and non-extravasated cells from the lungs. Lungs were collected and genomic DNA was isolated using phenol chloroform extraction. Genomic DNA of the preinjection pools was isolated using a QIAamp DNA Blood Mini Kit (Qiagen).

The *in vitro* invasion assays were performed in parallel. Each pool was plated on two six-well BioCoat Matrigel invasion plates (Corning). Six hundred thousand cells were plated on top of each well in cell culture medium without serum. Cells were allowed to invade through the Matrigel into medium containing 5% fetal bovine serum and 5% fetal calf serum for 24 h. Invaded cells were scraped off, washed with PBS, and genomic DNA was isolated using a QIAamp DNA Blood Mini Kit.

The shRNAs were amplified using a two-step PCR protocol previously described in ref. 4 for next-generation sequencing.

First PCR forward primer 1: 5-CAG AAT CGT TGC CTG CAC ATC TTG GAA AC-3 and reverse primer 1: 5-CTG CTA AAG CGC ATG CTC CAG ACT GC-3.

Second PCR forward primer 2: 5-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT AGC CTG CGC ACG TAG TGA AGC CAC AGA TGT A-3 and reverse primer 2: 5-CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT GCT AAA GCG CAT GCT CCA GAC TGC-3. The reverse primer contained a barcode (NNNNNN) that enabled multiplexing.

Analysis of screening data

shRNA quantification was performed as previously described[4]. For each pool, shRNA quantities in each sample were normalized by their sum and log-fold changes were calculated using the initially infected cell population as a reference. Fold change values for each pool were then consolidated into a single table and an empirical Bayes-moderated t-test was applied to identify depleted molecules. Genes with two or more depleted shRNAs were considered hits in each screen.

Gene Ontology enrichment analysis

Gene Ontology enrichment analysis used the GOrilla web portal[17]. The Refseq identifiers of genes identified as overexpressed in 4T1-T cells, compared with 4T1-E cells, were used as foreground and the entire Refseq gene list was used as background.

Expression subtype and relapse analysis

For all clinical data analysis, we used the University of North Carolina '855 patient set', which is available as published data at https://genome.unc.edu. All data were derived from an initial matrix that was arranged with patients on the horizontal axis and genes on the vertical axis. Initial normalization involved quantile normalization to ensure that the global expression profile of each patient was similar. Following this, each gene was z-score normalized across patients. For Extended Data Fig. 1a, the average expression level of each gene was calculated for each patient subtype. For Extended Data Fig. 1b, the average expression level of each gene was calculated for each gene for patients with and without relapse to each secondary site. For Extended Data Fig. 1d, the level of ASNS in each patient, stratified by subtype, is plotted. For Extended Data Fig. 1e, each boxplot represents the level of ASNS in each patient with and without relapse to each secondary site.

Individual *in vitro* invasion assay

The *in vitro* invasive capacity of cells was measured using six-well BioCoat Matrigel invasion plates. For parental 4T1 cells, 1×10^6 cells were plated on individual wells; for 4T1-T cells, 8×10^5 cells were plated on individual wells; and for MDA-MB-231 cells, 5×10^5 cells were used per individual well. Cells were re-suspended in medium without serum or NEAA and cells invaded into medium with 5% fetal bovine serum and 5% fetal calf serum. For Fig. 3a and Extended Data Fig. 4b, 4T1 and MDA-MB-231 cells were cultured in medium containing $100 \times$ concentration of the specified amino acid (relative to the concentration achieved by supplementing the medium with $1 \times$ NEAA) for 2 or 3 days, respectively, before starting the invasion assay. After 24 h, non-invaded cells were removed and the invasion wells were washed in PBS, fixed in 2% glutaraldehyde for 2 min, and stained with 0.5% crystal violet for 10 min. The wells were washed in distilled H2O, air-dried, and scanned using an Odyssey infrared scanner. The signal was quantified using ImageJ (NIH).

Competition and proliferation assay

For Extended Data Fig. 2c, d, cells were grown in medium lacking NEAA. For the mCherry competition assay, shRNA-transduced mCherry-positive cells were admixed with untransduced cells. mCherry fluorescence was quantified on an LSR II flow cytometer (BD Biosciences). The proliferation assay was performed using a CellTrace Violet Cell Proliferation Kit (Thermo Fisher Scientific). For Extended Data Fig. 4d, e, 4T1 and MDA-MB-231 cells were cultured in medium containing $100 \times$ concentration of the specified amino acid (relative to the concentration that is achieved by supplementing the medium with $1 \times$ NEAA) for 2 or 3 days, respectively, before starting the proliferation assay. Cells were stained with CellTrace violet and then trypsinized and re-suspended in the medium. After 24

h, cells were collected to quantify violet fluorescence intensity using an SH800 flow cytometer (Sony).

Isolation of tumour and lung metastatic cells

Tumour and lung tissues were collected, minced, and digested into single cells as previously reported[1]. Cells were either grown in 4T1 cell culture medium containing 60 μM 6-thioguanine to deplete stromal cells or directly sorted on the basis of mCherry expression using the FACSAria III cell sorter (BD Biosciences).

RNA sequencing library preparation

RNA sequencing libraries from cultured 4T1-T cells were prepared as previously described[1]. Each sample was sequenced on an Illumina HiSeq sequencer generating 76-nucleotide single-end reads.

RNA sequencing analysis

Illumina sequencing reads were aligned to the mouse genome (mm10) using Bowtie2 with default parameters[18]. Genes were assigned a count using HTseq-count, and differential expression analysis was performed using DESeq[19,20].

shRNA knockdown and cDNA overexpression studies

Mouse and human cell lines were transduced with shRNA expressing retroviral or lentiviral constructs, respectively. After infection, 4T1-T cells were selected with 500 μg ml^{-1} hygromycin for 5 days and MDA-MB-231 cells were selected with 2 μg ml^{-1} puromycin for 4 days. Cell lines infected with cDNA overexpressing retroviral constructs were selected with G418 for 1 week. The parental 4T1 cell line was selected with 600 μg ml^{-1} G418 and MDA-MB-231 cells were selected with 1,500 μg ml^{-1} G418.

cDNA overexpression genes

Mouse Asns: NM_012055.1. Human ASNS: NM_001673.2.

shRNA knockdown sequences

Mouse shAsns-1:
TGCTGTTGACAGTGAGCGCCACTGCCAATAAGAAAGTATATAGTGAAGCCACAGA
TGTATATACTTTCTTATTGGCAGTGTTGCCTACTGCCTCGGA.

Mouse shAsns-2:
TGCTGTTGACAGTGAGCGCCACTATGAAGTTTTGGATTTATAGTGAAGCCACAGAT
GTATAAATCCAAAACTTCATAGTGTTGCCTACTGCCTCGGA.

Mouse shTgfbi-i:
TGCTGTTGACAGTGAGCGCCAGTATATAtatgttcttcaatagtgaagccacagatgtattgaagaacatatata
tactgttgcctactgcctcgga.

Mouse shTgfbi-2:
TGCTGTTGACAGTGAGCGAAGTATATATATGTTCTTCAAATAGTGAAGCCACAGAT
GTATTTGAAGAACATATATATACTGTGCCTACTGCCTCGGA.

Human shASNS-i:
TGCTGTTGACAGTGAGCGCCAGAAGCTAAAGGTCTTGTTATAGTGAAGCCACAGA
TGTATAACAAGACCTTTAGCTTCTGATGCCTACTGCCTCGGA.

Human shASNS-2:
TGCTGTTGACAGTGAGCGCAGCAATGACAGAAGATGGATATAGTGAAGCCACAGA
TGTATATCCATCTTCTGTCATTGCTTTGCCTACTGCCTCGGA.

Quantitative RT-PCR

Total RNA from cells was purified and DNase treated using an RNeasy Mini Kit (Qiagen).
For whole tissues, RNA was isolated using a TRIzol Plus RNA Purification Kit (Thermo
Fisher Scientific). The tissue lysate was homogenized using a Dounce homogenizer and
passed through a column homogenizer (Thermo Fisher Scientific) to reduce viscosity. RNA
integrity (RNA integrity score > 9) was measured on an Agilent Bioanalyzer (RNA nano
kit). cDNA was synthesized using SuperScript III Reverse Transcriptase (Sigma). qPCR was
performed on an Eppendorf Mastercycler ep realplex. All signals were quantified using the
ACt method and were normalized to the levels of Gapdh. For mCherry-positive flow
cytometer sorted tumour and lung metastatic cells, cDNA was produced directly from lysed
cells using a TaqMan Gene Expression Cells-to-Ct Kit (Thermo Fisher Scientific). qPCR
was performed on an CFX96 (Bio-Rad) using TaqMan primer/probe sets and all signals
were quantified as described above.

Quantitative RT–PCR primers

Mouse Asns (exon 1–2): 5′-CCTCTGCTCCACCTTCTCT-3′;5′-
GATCTTCATCGCACTCAGACA-3′.

Mouse Asns (exon 6–7): 5′-CCAAGTTCAGTATCCTCTCCAG-3′;5′-
CTTCATGATGCTCGCTTCCA-3′.

Mouse Tgfb1 (exon 1–2): 5′-CCGAATGTCTGACGTATTGAAGA-3′;5′-
GCGGACTACTATGCTAAAGAGG-3′.

Mouse Tgfb1 (exon 3–4): 5′-GTTATCTTTGCTGTCACAAGAGC-3′;5′-
CCCACTGATACGCCTGAG-3′.

Mouse Gapdh (exon 2–3): 5′-AATGGTGAAGGTCGGTGTG-3′;5′-
GTGGAGTCATACTGGAACATGTAG-3′.

Human ASNS (exon 8–9): 5′-GAGTCAGACCTTTGTTTAAAGCA-3′;5′-
GGAGTGCTTCAATGTAACAAGAC-3′.

Human ASNS (exon 12–13): 5′-CTGGATGAAGTCATATTTTCCTTGG-3′;5′-
CAGAGAAGATCACCACGCTATC-3′.

Human GAPDH (exon 2–3):5′-ACATCGCTCAGACACCATG-3′;5′-TGTAGTTGAGGTCAATGAAGGG-3′

TaqMan probes

Mouse Asns: Mm00803785_m1. Mouse E-cadherin (Cdh1): Mm01247357_m1. Mouse Twist1: Mm00442036_m1. Mouse Gapdh: Mm99999915_g1.

qPCR for CTCs

CTCs were quantified as previously described[1]. Genomic DNA was isolated from blood and CTC abundance was quantified using a qPCR assay against mCherry, which was expressed from the retroviral shRNA delivery vectors.

mCherry probes and primers were as follows: primer 1, 5′-GACTACTTGAAGCTGTCCTTCC-3′; primer 2, 5′-CGCAGCTTCACCTTGTAGAT-3′.

HEX probe: 5′-/56-FAM/TTCAAGTGG/ZEN/GAGCGCGTGATGAA/3IABkFQ//-3′.

Housekeeping probes and primers: primer 1, 5′-GACTTGTAACGGGCA GGCAGATTGTG-3′; primer 2, 5′-GAGGTGTGGGTCACCTCGACATC-3′.

HEX probe: 5′-/5HEX/CCGTGTCGC/ZEN/TCTGAAGGGCAATAT/3IABkFQ/-3′.

Quantification of lung metastasis

For each lung, 5 μm sections were prepared and stained with a standard H&E protocol. Lung metastatic burden was determined by counting individual lung nodules on one section.

E-cadherin and Twist1 analysis

For immunohistochemistry, primary tumours and lungs were processed as previously described[1]. E-Cadherin (24E10) Rabbit mAb (3195, Cell Signaling) was used in a 1:400 dilution and Twist1 (Twist2C1a) Mouse mAb (ab50887, Abcam) was used in a 1:100 dilution. E-cadherin and Twist1 diaminobenzidine-stained (DAB) and haematoxylin stainings were quantified using ImageJ (NIH). For this, images were colour deconvoluted according to ref. 21 and the percentage area of E-cadherin- and Twist1-positive staining was measured.

Free amino-acid quantification using HPLC

Free amino acids were quantified in cultured cells and blood serum. For Extended Data Fig. 2e, 4T1-T cells were cultured under normal medium conditions. For Extended Data Fig. 4a, c, 4T1 and MDA-MB-231 cells were cultured in medium containing 100× concentration of the specified amino acid (relative to the concentration achieved by supplementing the medium with 1× NEAA) for 2 or 3 days, respectively. All cultured cells were washed in ice-cold PBS, homogenized using a Dounce homogenizer, and the lysate subsequently filtered. Each biological sample was quantified in triplicates using HPLC and a fluorometric detector. For each replicate, nanomoles of each amino acid were measured.

Proteomic profiling using isobaric tags for relative and absolute protein quantification

All analyses were performed with three biological replicates per treatment. Cell pellets were lysed with 0.1 M triethylammonium bicarbonate, 0.1% SDS buffer, followed by tip sonication and boiling at 90 °C for 5 min. Protein concentration was estimated using a Bradford assay (Bio-Rad, Quick Start), and 90 μg of total protein per sample were reduced with tris-2-carboxymethyl phosphine (Sigma) for 1 h at 60 °C at a final concentration of 5 mM. Cysteines were blocked for 10 min at room temperature using methyl methanethiosulfonate (Sigma) at a final concentration of 10 mM. Samples were digested overnight at 37 °C with trypsin (Pierce) and peptides were labelled with the TMT10plex reagents according to the manufacturer's instructions (Thermo Fisher Scientific). The TMT mixture was fractionated on a Dionex Ultimate 3000 system at high pH using a XBridge C18 column (3.5 μm, 2.1 × 150 mm, Waters). Fractions were analysed on a Dionex Ultimate 3000 UHPLC system coupled with a nano Q-Exactive (Thermo Fisher Scientific) mass spectrometer. Samples were loaded on an Acclaim PepMap 100, 100 μm × 2 cm C18, 5 μm, 100 Å trapping column with the ulPickUp injection method at a loading flow rate of 4 μl min−1 for 10 min. For peptide separation, an EASY-Spray analytical column 75 μm × 25 cm, C18, 2 μm, 100 Å column was used for multi-step gradient elution. Mobile phase (A) was composed of 2% acetonitrile, 0.1% formic acid, 5% dimethyl sulfoxide (DMSO); mobile phase (B) was composed of 80% acetonitrile, 0.1% formic acid, 5% DMSO. The full scans were obtained at 70,000 resolution and the MS2 scans at 35,000 resolution, with collision energy 33% and isolation window 1.2 Th. Raw data were processed with the SequestHT search engine on Proteome Discoverer 1.4 software. The parameters for the SequestHT node were as follows: precursor mass tolerance 20 p.p.m., fragment mass tolerance 0.02 Da; dynamic modifications were oxidation of M (+ 15.995 Da), deamidation of N, Q (+ 0.984 Da); and static modifications were TMT6plex at any N terminus, K (+ 229.163 Da) and Methylthio at C (+ 45.988 Da).

For each sample, raw peptide intensities were normalized by their sum. Peptides that had a representation that fell within the bottom quartile based on summed intensities across samples were not analysed further. For the dataset represented in Fig. 4a and Extended Data Fig. 7a, b, proteins were removed from subsequent analysis if fewer than 5 corresponding proteins were identified in any sample. Protein-level differences between treatments were calculated based on the trimmed mean of corresponding peptide log-fold changes (trimming the top and bottom 25% for the experiment represented in Fig. 4a and Extended Data Fig. 7a, b, and the top and bottom 10% for the experiment represented in Fig. 4b). Relative protein abundances were calculated based on the trimmed mean of corresponding log-transformed peptide intensities (trimming the top and bottom 25% and 10% for the experiments representing in Fig. 4a and b, respectively).

Metabolite profiling using LC–MS/MS

Organ tissue samples were placed in 2 ml lysing tubes prefilled with 1.4-mm ceramic beads for mammary glands or 2.8-mm ceramic beads for lungs and 1 ml of pre-chilled (− 80 °C) 80% methanol. Samples were homogenized with a Precellys24 homogenizer (Bertin Instruments) programmed with three 30-s cycles at 6,500 Hz and 4-min pause times. At the end of each cycle, samples were snap-frozen in liquid nitrogen and placed on dry ice.

Metabolite extraction of blood serum samples (50 μl) was performed using 200 μl of 80% methanol at – 80 °C. After centrifugation for 10 min (13,200 revolutions per minute, 4 °C), supernatants were evaporated to dryness and stored at – 80 °C until liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis.

Dried-down extracts were re-suspended in 25 μl HPLC-grade water, and 1 μl was analysed using hydrophilic interaction chromatography coupled to tandem mass spectrometry analysis operated in the selected reaction monitoring mode (LC–SRM–MS). Analytical instrumentation consisted of a Nexera X2 (Shimadzu) liquid chromatography system with a XBridge BEH amide 2.1 × 100 mm, 2.5-μm column (Waters) and a QTRAP 6500 hybrid triple quadrupole/linear ion trap mass spectrometer (SCIEX) equipped with an electrospray ion source. Raw LC–SRM–MS data were acquired with Analyst 1.6.2 (SCIEX) and peak areas of LC–SRM–MS traces for each metabolite were integrated using MultiQuant 1.1 software (SCIEX).

Amino acid compositional analysis

For Extended Data Fig. 7b, the amino acid representations of genes that showed the greatest decrease in their protein or RNA levels (bottom 25% log-fold change) were compared to those of the analysed gene set, using a rank-sum test. For the amino acid enrichment analysis of protein minus RNA level expression changes, transcriptional- and protein-level log-fold changes were quantile normalized to the same distribution. RNA-level changes were then subtracted from protein-level changes. The amino acid representations of genes whose subsequent values fell within the lowest 25% were then compared to the analysed gene set using a rank-sum test, to identify amino acids whose abundance correlated with the protein-level changes that were not explained by transcriptional changes.

For Extended Data Fig. 7c, the same analysis was performed, this time comparing the proteins of genes that had been detected as upregulated during EMT compared with all other genes. For the mouse amino-acid enrichment, the EMT-up genes were mouse orthologues of the EMT-up human genes.

For Extended Data Fig. 8, each organism harbouring a minimum of ten genes that were orthologues of the pro-EMT human genes described in Extended Data Fig. 7c were analysed. For each organism, the asparagine percentage of each protein was calculated. Then the asparagine enrichment level for each organism was determined by calculating the ratio of the median asparagine percentage of pro-EMT proteins versus the remaining organism-specific proteins. The statistical significance of enrichment was calculated as described for Extended Data Fig. 7c.

Ribosome profiling analysis

For analysis, quality trimming and sequencing adapter removal was performed using Cutadapt22. Bowtie2 was used to remove reads that mapped to contaminating RNAs (for example, rRNA and tRNA sequences)18. STAR was subsequently used to map reads of length 29–33 to the human transcriptome23. The offset was corrected for each read on the basis of read length, and 12 and 15 nucleotides downstream were marked as P and A sites. For each gene, we then calculated the number of events at all positions and aggregated the

counts for each codon and subsequently amino acid. Genes with fewer than ten events in any sample were not considered in subsequent analyses.

Code availability

All custom code used during the current study is available from the corresponding author upon reasonable request.

Data availability

All raw and processed high-throughput sequencing data are available at the Gene Expression Omnibus under accession numbers GSE104968 and GSE107109. The mass spectrometry proteomics data have been deposited in the ProteomeXchange consortium via the PRIDE24 partner repository, under the dataset identifier PXD008805. Source Data are available for tumour volume measurements, lung metastases and proteomic measurements. All other datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Extended Data
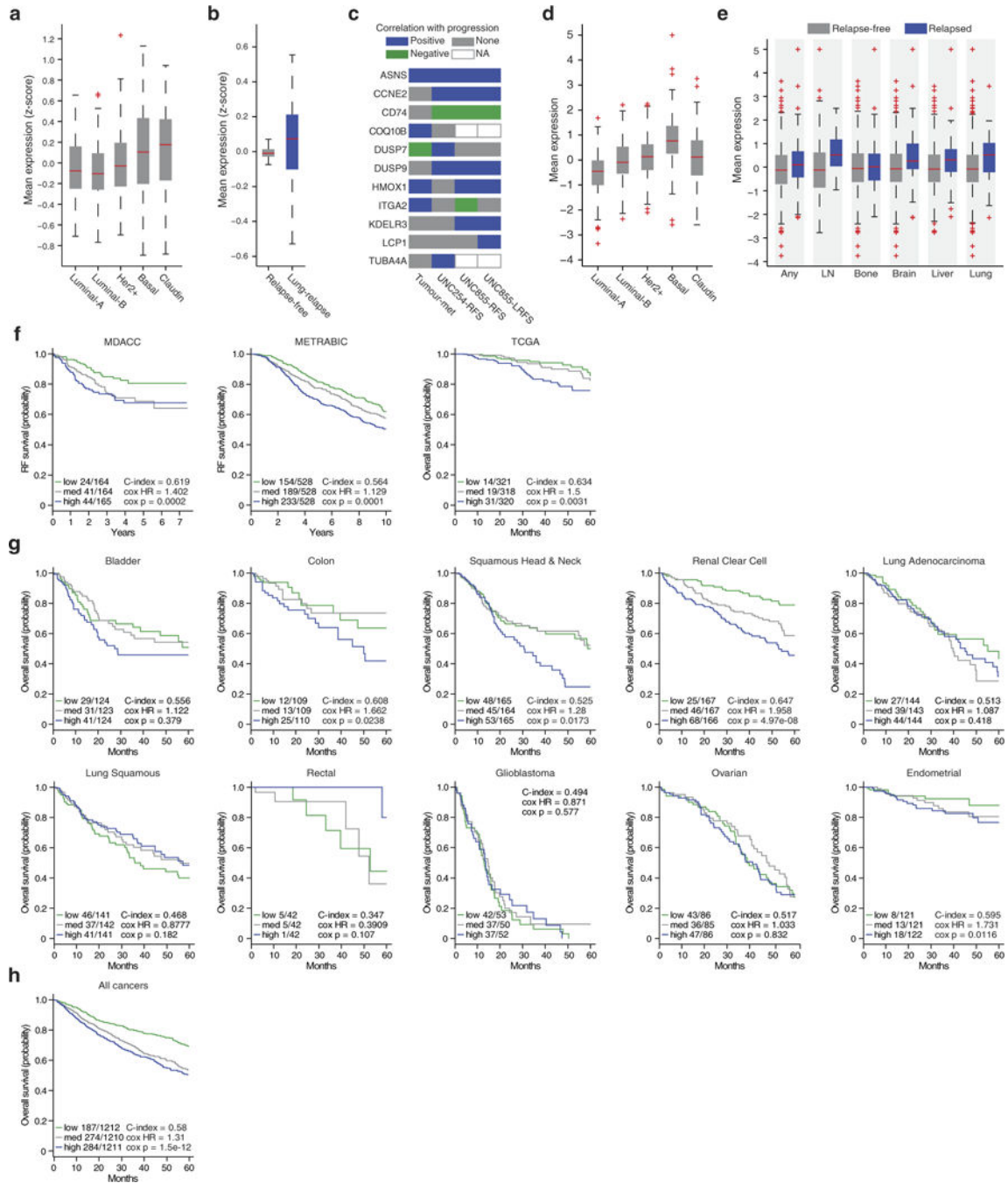
Extended Data Figure 1. Analysis of ASNS expression levels in patient data

a, Expression level of genes identified as overexpressed in 4T1-T compared with 4T1-E in the primary tumours of patients with different disease subtypes (edges of the box are the 25th and 75th percentiles and error bars extend to the values $q3 + w(q3 − q1)$ and $q1 − w(q3 − q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also true for **b**, **d**, and **e**, ANOVA $P < 0.0001$). **b**, Expression level of the same genes in disease-free

survivors and patients with relapse to the lung (rank-sum $P < 0.01$). **c**, For each gene that was identified in the screen, a prognostic value was calculated using three different datasets. One consisted of gene expression measurements in three patient-matched basal tumour and metastasis pairs (patients A1, A7, and A11). Here genes were classified as correlated with progression if expression was higher in each of the metastases and negatively correlated if expression was higher in each of the primaries. The other two datasets consisted of primary tumour gene-expression profiles with matched outcomes. For the UNC254 patient dataset, the site of relapse was not available and genes were deemed positively correlated with progression if they had significant (Cox $P < 0.05$) relapse-free survival hazard ratios greater than 1, and negatively correlated if these ratios were significant (Cox $P < 0.05$) and less than 1. As the UNC855 dataset also had site of relapse information, here both relapse-free and lung relapse-free survival (RFS and LRFS) hazard ratios were used to classify genes as positively or negatively correlated with progression based on the same criteria that were used for the UNC254 data. **d**, Expression level of ASNS in the primary tumours of patients with different disease subtypes (ANOVA $P < 0.0001$). **e**, Expression level of ASNS in the primary tumours of patients with non-specific relapse and relapse to the lymph node, bone, brain, liver, or lung compared with expression levels in patients without relapse to each corresponding site (rank-sum $P < 0.005$). **f**, Analysis of ASNS in three additional sets from patients with breast cancer (MDACC, METRABIC, and TCGA). Shown are survival plots and relevant statistics (Cox $P < 0.01$). **g**, Analysis of ASNS in the TCGA Pan-Cancer expression data. Shown are survival plots and relevant statistics for the ten non-breast solid tumours represented in the dataset (Cox $P < 0.05$ for colon, squamous head and neck, renal clear cell, and endometrial cancers). **h**, Analysis of ASNS across all tumours represented in the TCGA Pan-Cancer dataset (Cox $P = 1.5 \times 10^{-12}$).
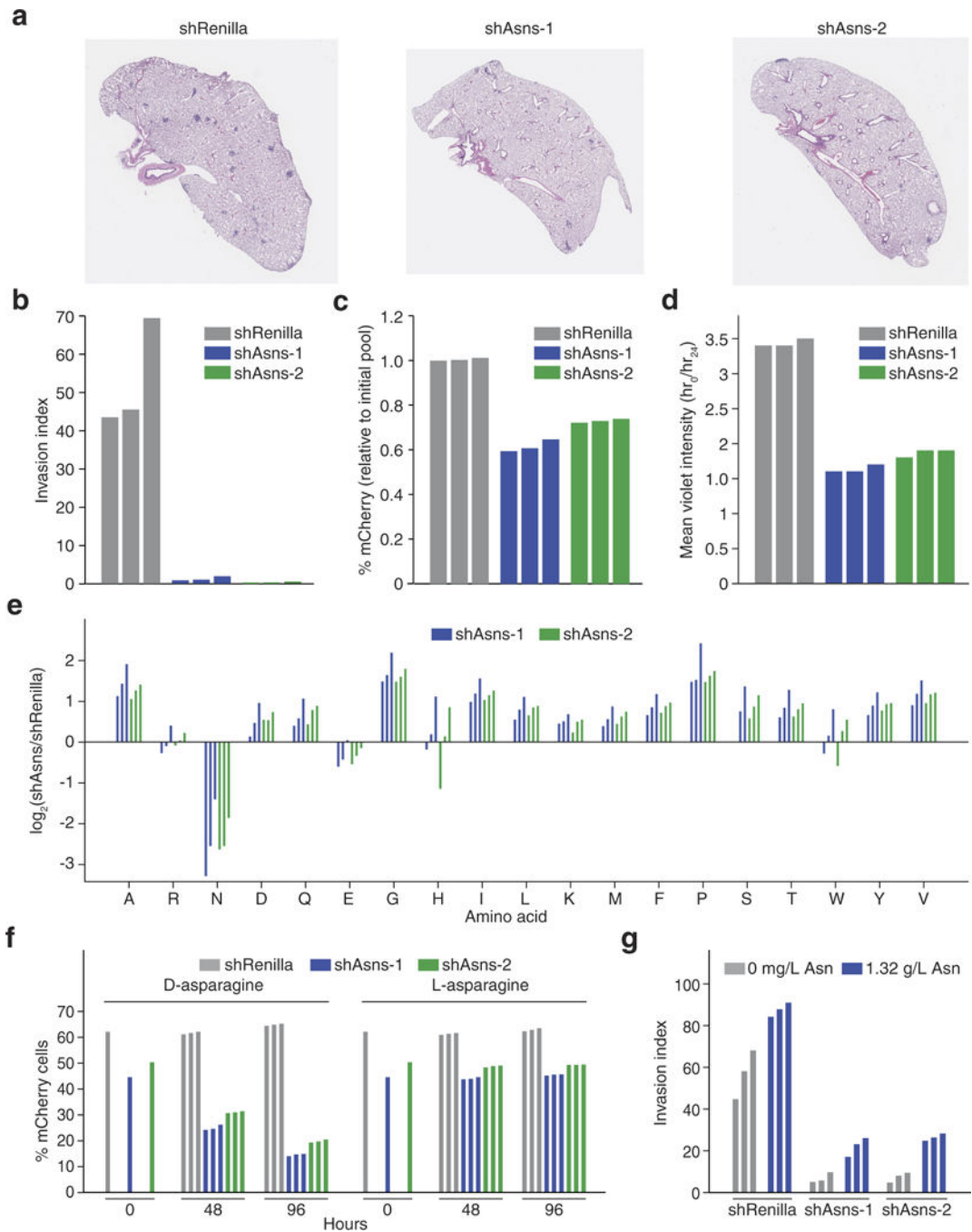
Extended Data Figure 2. Primary validation of Asns as a driver of invasion and metastasis
a, Representative images of the lungs of mice that were intravenously injected with Asns-silenced or -expressing 4T1-T cells as described in Fig. 2a. **b**, Quantification of Matrigel invasion capacity for Asns-silenced and -expressing 4T1-T cells ($n = 3$ replicates per cell line). **c**, Quantification of mCherry-positive 4T1-T cells after roughly 50% of cells were infected with mCherry-expressing constructs harbouring shRNAs targeting *Renilla* luciferase and Asns. Cells were grown during the 24-h period that the Matrigel invasion assay described in Fig. 2b was being performed ($n = 3$ replicates per cell line). **d**, Violet cell-

labelling intensity of Asns-silenced and -expressing 4T1-T cells, relative to the initial population. Cells were grown during the 24-h period that the Matrigel invasion assay described in Fig. 2b was being performed ($n = 3$ replicates per cell line). **e**, Free amino-acid quantification by HPLC for each amino acid in Asns-expressing and -silenced cells. Shown are the log-fold changes for each amino acid ($n = 3$ replicates per cell line). **f**, Quantification of mCherry-positive 4T1-T cells after roughly 50% of cells were infected with mCherry-expressing constructs harbouring shRNAs targeting *Renilla* luciferase and Asns. After infection, cells were grown in medium supplemented with l-asparagine or d-asparagine and mCherry percentages were measured at 48 and 96 h ($n = 3$ replicates per cell line). **g**, Quantification of Matrigel invasion for Asns-silenced and -expressing cells when assayed in medium supplemented with and without l-asparagine ($n = 3$ invasion chambers per cell line).
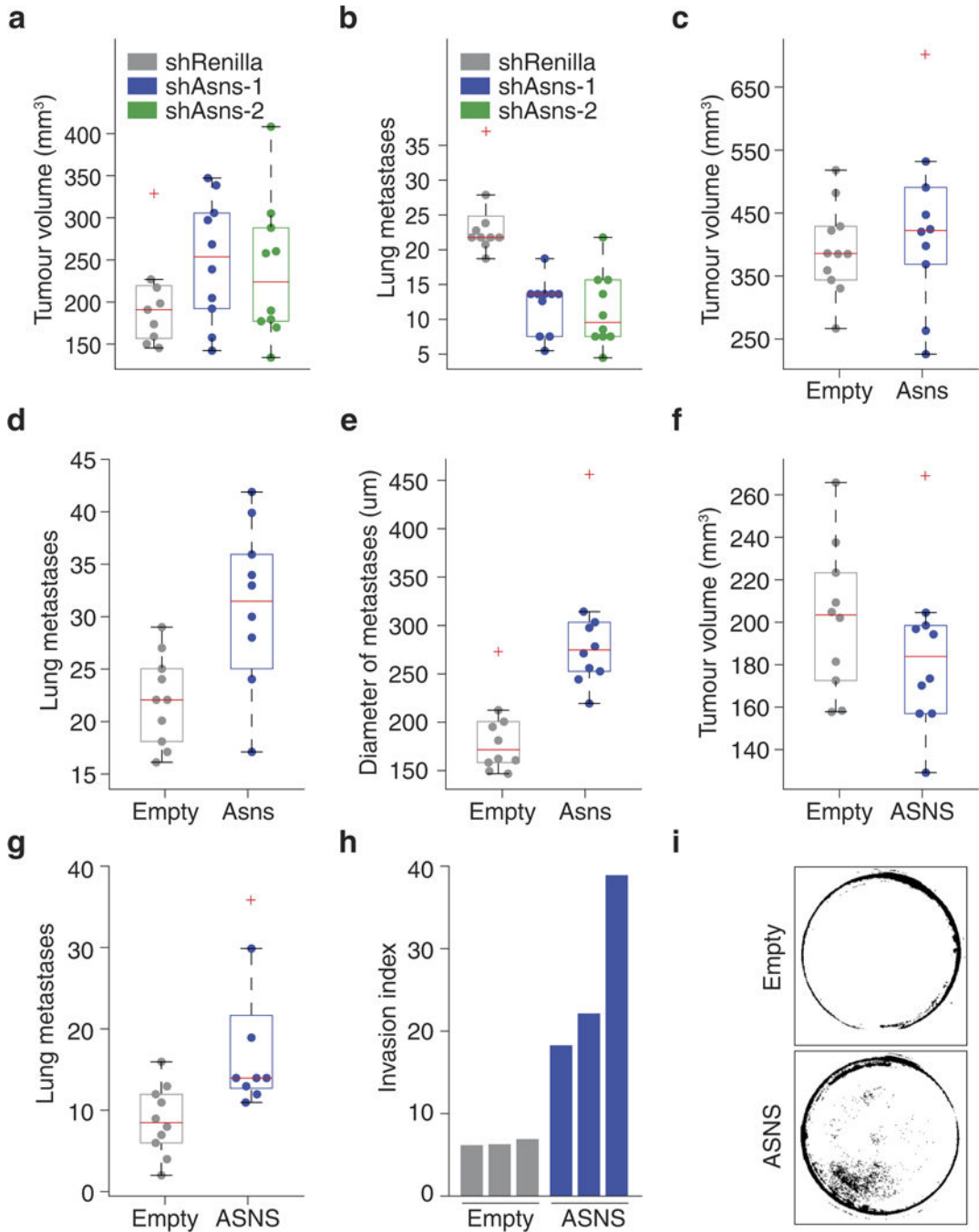
Extended Data Figure 3. Secondary validation of Asns as a driver of invasion and metastasis
a, Volume measurements of tumours resulting from orthotopic injection of Asns-silenced and -expressing parental 4T1 cells ($n = 10$ mice per cell line, edges of the box are the 25th and 75th percentiles and error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also the case for **b**–**g**). **b**, Quantification of lung metastases corresponding to the tumours described in **a** (rank-sum $P < 0.002$). **c**, Volume measurements of tumours resulting from orthotopic injection of parental 4T1 cells with basal (Empty) or enforced expression of Asns ($n = 10$

mice per cell line). **d**, Quantification of lung metastases corresponding to the tumours described in **c** (rank-sum $P < 5.0 \times 10-5$). **e**, Average diameters of the metastases of each mouse described in **d** (rank-sum $P < 0.001$). **f**, Volume measurements for tumours resulting from orthotopic injection of MDA-MB-231 cells with basal (Empty) or enforced expression of ASNS ($n = 10$ mice per cell line). **g**, Quantification of lung metastases corresponding to the tumours described in **f** (rank-sum $P < 0.005$). **h**, Quantification of Matrigel invasion for the MDA-MB-231- derived cell lines described in **f** ($n = 3$ invasion chambers per cell line). **i**, Representative images of the collection wells for the invasion assays described in **h**. See Source Data.
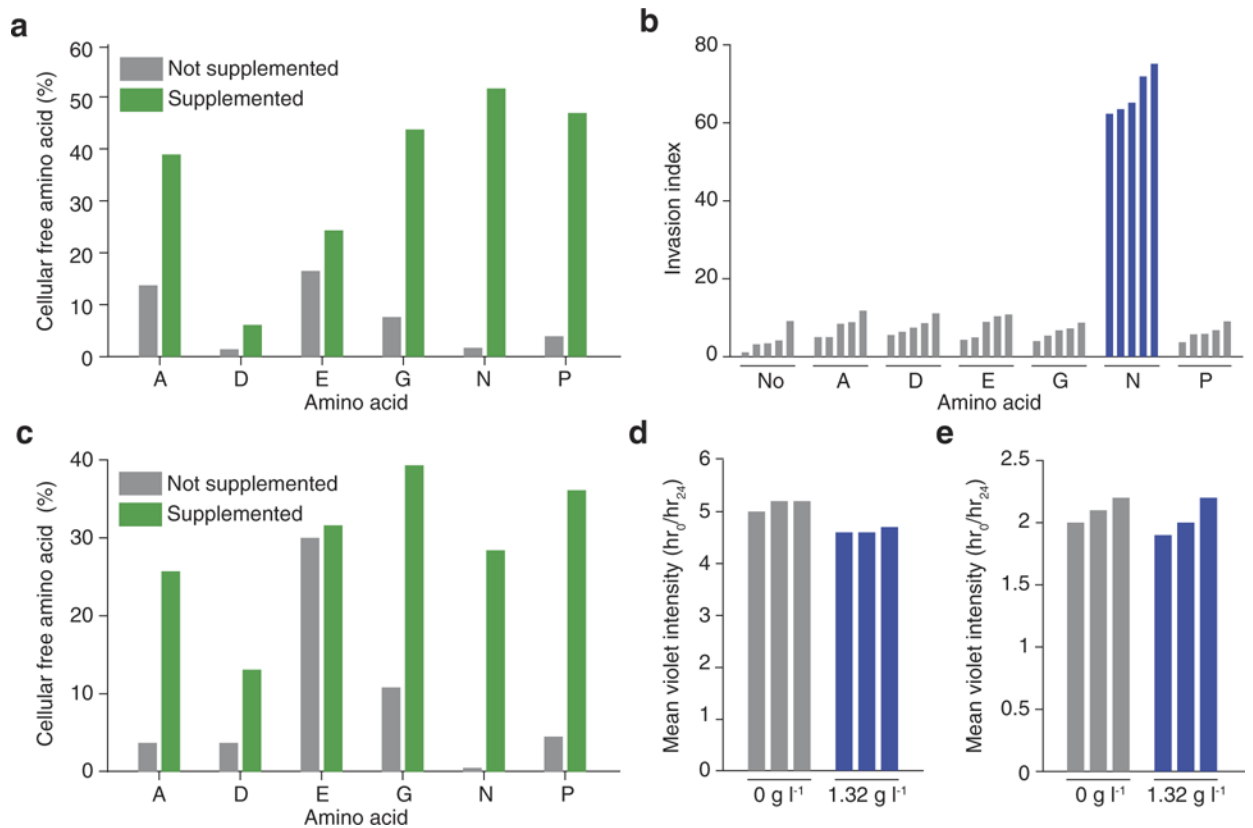
Extended Data Figure 4. Primary validation that extracellular asparagine availability affects invasion and metastasis

a, HPLC quantification of cellular free amino-acid percentages for parental 4T1 cells when the medium is supplemented with each of the NEAAs lacking in the DMEM culture medium ($n = 3$ replicates per cell line). **b**, Quantification of MDA-MB-231 Matrigel invasion rates under the same conditions as described in Fig. 3a (n = 5 invasion chambers per condition, rank-sum $P < 0.001$). **c**, HPLC quantification of cellular free aminoacid percentages for MDA-MB-231 cells when cultured in the medium conditions described in **a** ($n = 3$ replicates per cell line). **d**, Violet celllabelling intensity of parental 4T1 cells when grown in asparagine-lacking or -supplemented medium for the same period that the Matrigel invasion assay described in Fig. 3a was being performed ($n = 3$ replicates per cell line). **e**, Violet cell-labelling intensity of MDA-MB-231 cells when grown in asparagine-lacking or -

supplemented medium for the same period that the Matrigel invasion assay described in **b** was being performed ($n = 3$ replicates per cell line).
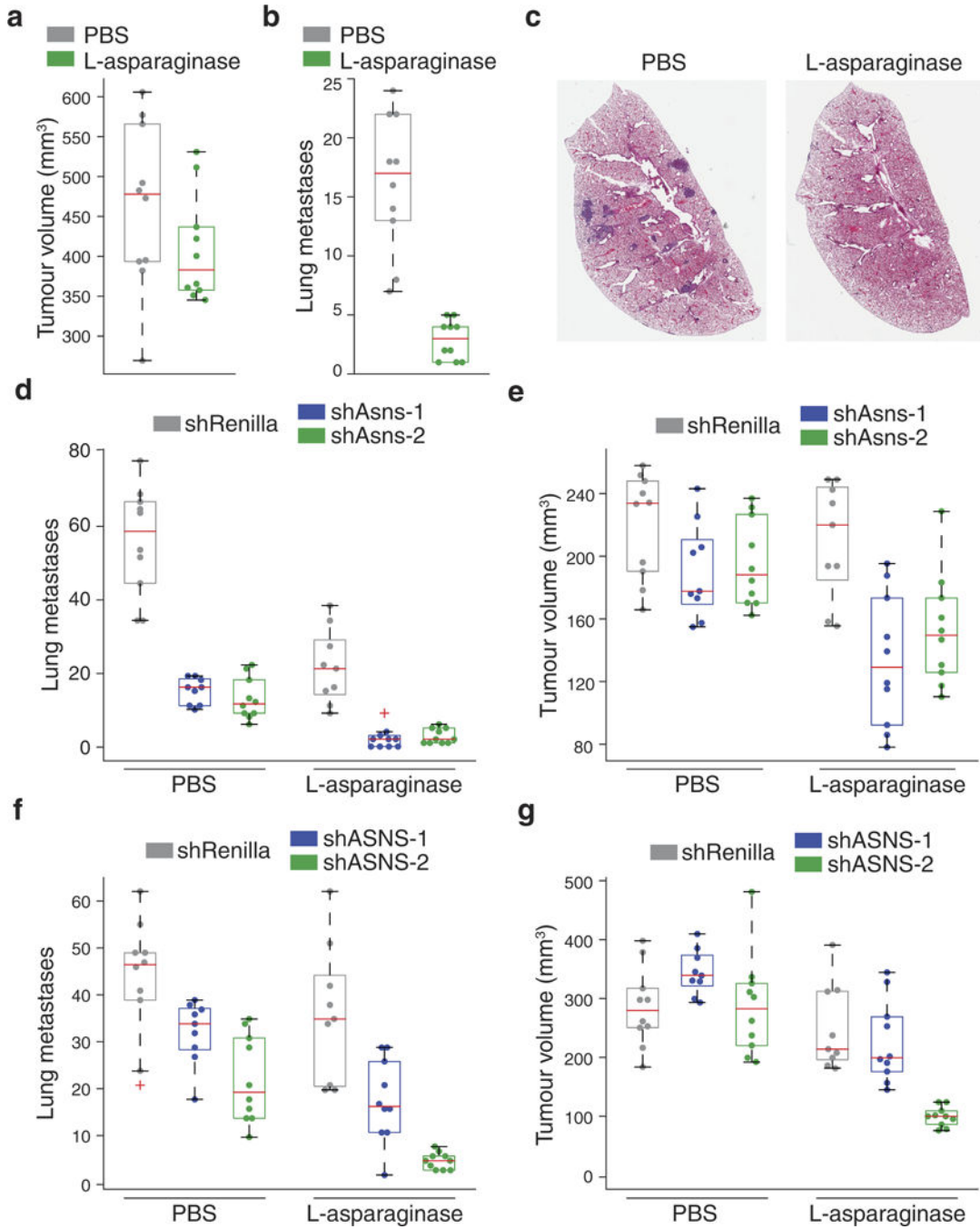
Extended Data Figure 5. Secondary validation that extracellular asparagine availability affects invasion and metastasis

a, Tumour volumes resulting from the orthotopic injection of parental 4T1 cells. Half of the mice received l-asparaginase and the other half received an equivalent volume of PBS at the same injection rate ($n = 10$ mice per condition, edges of the box are the 25th and 75th percentiles and error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in

which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also the case for **b–g**). **b**, Quantification of lung metastases detected in the animals described in **a** (rank-sum $P < 0.001$). **c**, Representative H&E-stained lung sections as described in **b**. **d**, Quantification of the lung metastases described in Fig. 3b, in which Asns-silenced and -expressing 4T1-T cells were injected into mice. Half of the mice received l-asparaginase and the other half received an equivalent volume of PBS at the same injection rate ($n = 10$ mice per condition, rank-sum $P < 0.0005$ for l-asparaginase versus control for each line and for Asns-silenced versus -unsilenced cells in each drug condition). **e**, Tumour volumes corresponding to the lung metastases described in **d** (rank-sum $P < 0.005$ for Asns-silenced versus -expressing cells in l-asparaginase-treated mice). **f**, Lung metastases resulting from the orthotopic injection of ASNS-silenced and -expressing MDA-MB-231 cells and subsequent treatment of the injected animals with l-asparaginase or PBS ($n = 10$ mice per cell line, rank-sum $P < 0.05$ for ASNS-silenced versus -expressing cells in both conditions and for silenced cells in treated versus untreated mice). **g**, Tumour volumes corresponding to the mice described in **f** (rank-sum $P < 0.05$ for Asns-silenced versus -expressing cells under both treatments and for PBS versus l-asparaginase-treated animals for each cell line). See Source Data.
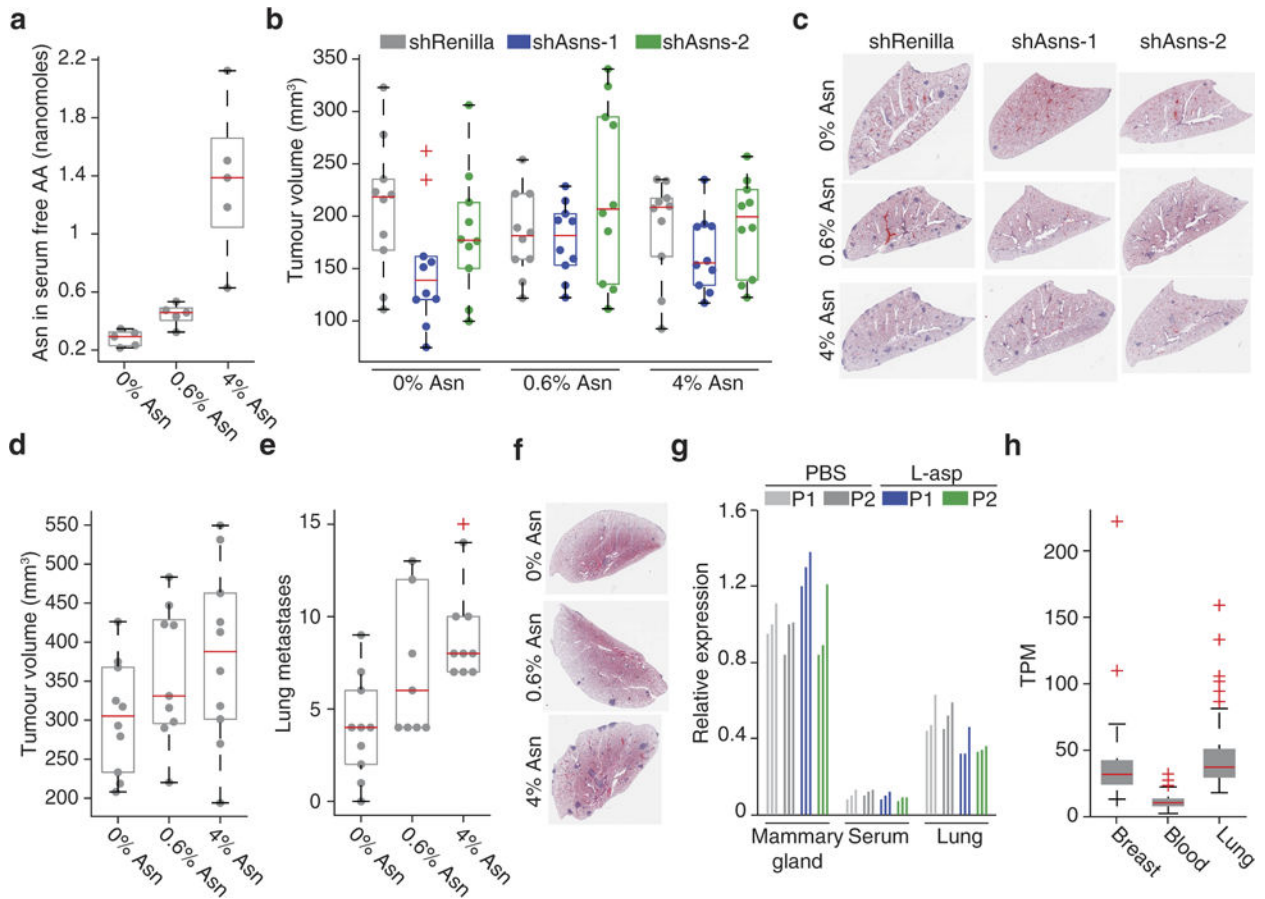
Extended Data Figure 6. Tertiary validation that extracellular asparagine availability affects invasion and metastasis

a, Asparagine content in the serum free amino-acid pool, for mice fed 0%, 0.6%, or 4% asparagine diets ($n = 5$ mice per diet, edges of the box are the 25th and 75th percentiles and
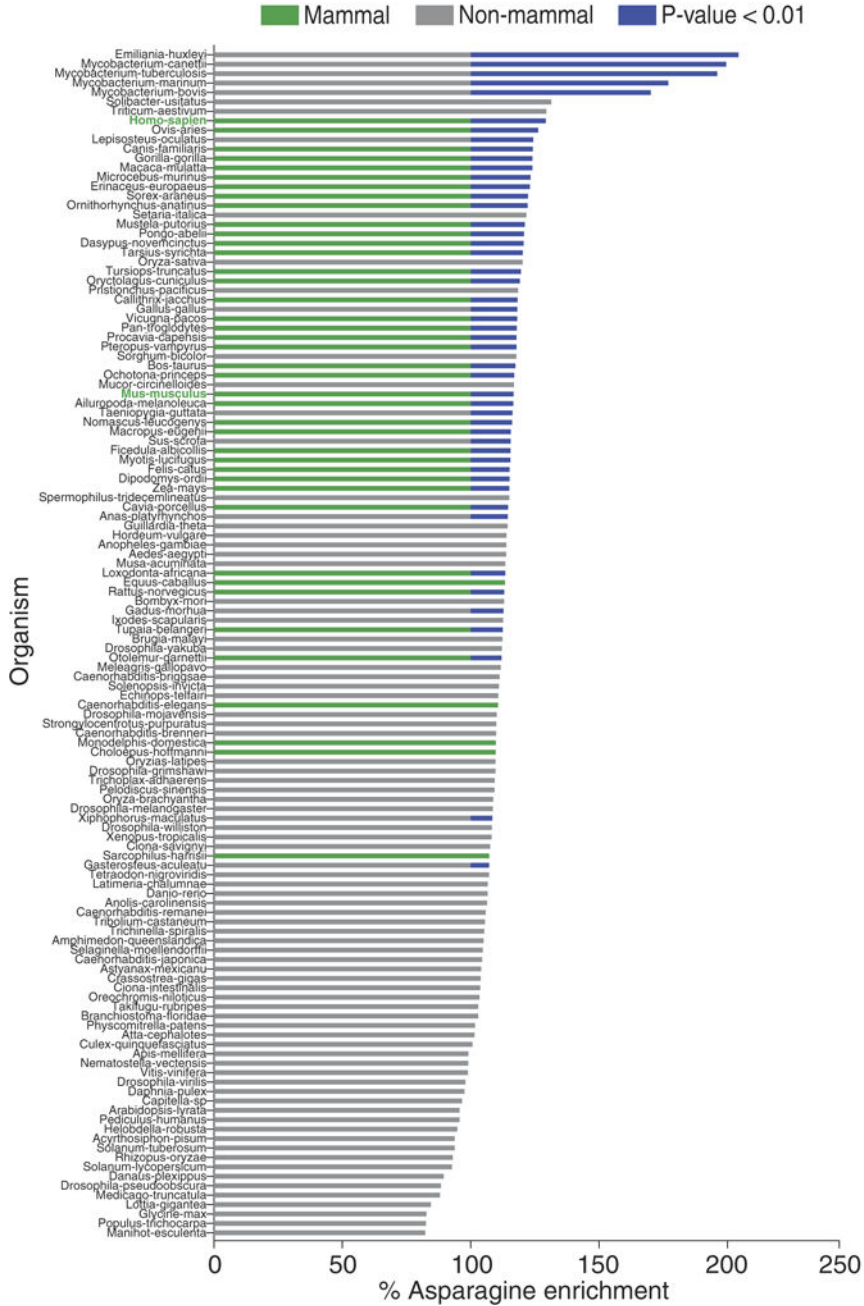
error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also the case for **b**, **d**, **e** and **h**, rank-sum $P < 0.05$ between each diet). **b**, Volumes of orthotropic tumours corresponding to the lung metastases described in Fig. 3c, in which Asns-silenced and -expressing 4T1-T cells were orthotopic injected into mice fed with 0%, 0.6%, and 4% asparagine diets ($n = 10$ mice per condition). **c**, Representative images of the lung metastases described for Fig. 3c, which also correspond to the mice described in **b**. **d**, Volumes of tumours resulting from the orthotopic injection of parental 4T1 cells into mice fed with 0%, 0.6%, or 4% asparagine diets ($n = 10$ mice per diet). **e**, Quantification of metastases in the lungs of the animals described in **d** (rank-sum $P < 0.05$ for mice receiving 0% versus 0.6% and 0% versus 4% diets). **f**, Representative images of H&E-stained sections of the lungs described in **e**. **g**, Relative expression of Asns in the mammary gland, serum, and lungs of mice treated with l-asparaginase or PBS, as measured by qPCR with two primer pairs P1 and P2 ($n = 3$ per condition). **h**, Transcripts per million (TPM) expression measurements for ASNS in human breast, lung, and whole-blood samples ($n > 114$ for each tissue, rank-sum $P < 2.8 \times 10^{-37}$ for blood versus breast and blood versus lung). See Source Data.
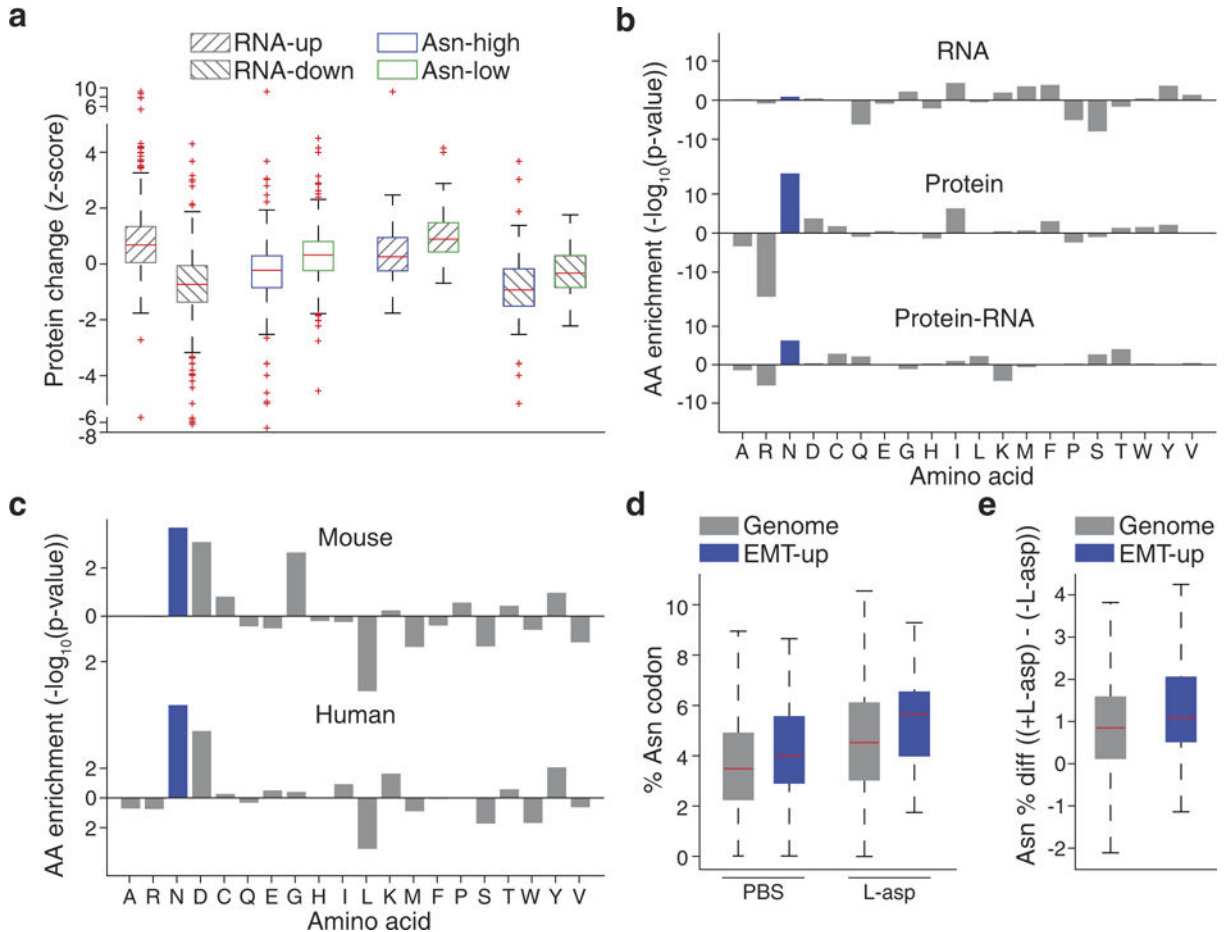
Extended Data Figure 7. Primary validation that asparagine availability regulates EMT
a, Protein-level changes between Asnssilenced and -expressing cells when genes are stratified by transcriptionlevel changes (top and bottom 10% of genes based on log-fold

change in Asns-silenced cells, gene-up and -down, respectively) and asparagine content (top and bottom 10% of genes based on asparagine content, Asp-high and -low, respectively), edges of the box are the 25th and 75th percentiles and error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also the case for **d** and **e**, rank-sum $P < 5.0 \times 10{-}24$ for both individual variables, and rank-sum $P < 0.005$ for interacting variables). **b**, Amino acid enrichment analysis of downregulated genes (bottom 25% based on log-fold change) on the basis of RNA and protein levels in Asns-expressing versus -silenced 4T1-T cells. Negative correlations indicate the amino acid is depleted in the downregulated genes, whereas positive correlations indicate the amino acid is enriched. For protein minus RNA level expression changes, amino acids with positive correlations are enriched in proteins in which depletion levels exceed what is predicted by corresponding RNA changes. Negative correlations indicate the amino acid is enriched in proteins in which depletion levels are less than what is predicted by corresponding RNA changes (rank-sum $P < 1.0 \times 10{-}5$ for asparagine in protein and protein– RNA). **c**, Amino-acid enrichment in mouse and human EMT-up proteins (rank-sum $P < 0.01$ for both human and mouse). **d**, Position 15 asparagine codon enrichment in ribosome protected fragments from PC-3 cells grown with and without l-asparaginase, when all genes or only EMT-up genes are analysed (outliers were not plotted to improve interpretability, which is also the case for **e**, rank-sum $P < 0.05$ for EMT-up versus all genes in both untreated and l-asparaginase-treated cells). **e**, Increase in asparagine codon representation in ribosome protected fragments, when PC-3 cells are grown in l-asparaginase (relative to without), and all genes or EMT-up genes are analysed (rank-sum $P < 0.05$). See Source Data.
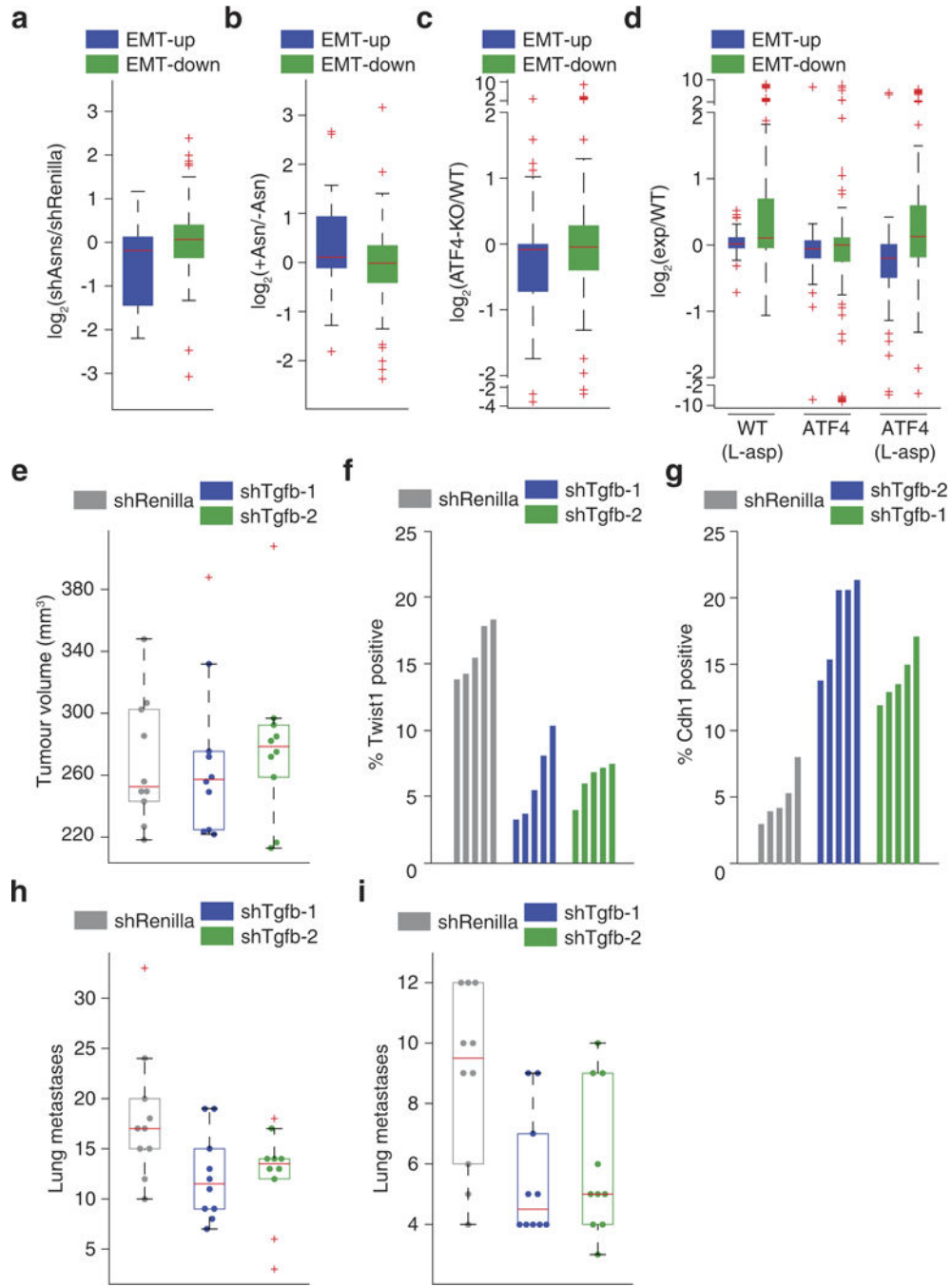
Extended Data Figure 8. Conservation of asparagine enrichment in EMT promoting proteins
Asparagine enrichment analysis of EMTpromoting protein orthologues in the 126 species listed in the Orthologous MAtrix database that harbour at least 10 orthologues (sign-rank $P < 1.0 \times 10{-}13$ for all species and rank-sum $P < 9.0 \times 10{-}9$ for mammals versus other species).

Extended Data Figure 9. Secondary validation that asparagine availability regulates EMT
a, Transcription-level changes in EMT-up and -down genes that occur in response to Asns silencing in 4T1-T cells ($n = 2$ replicates per condition, edges of the box are the 25th and 75th percentiles and error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also the case for **b**–**e**, **h** and **i**, rank-sum $P < 0.001$ for EMT-up genes, DESeq false discovery rate < 0.05 for Twist1 and Cdh1). **b**, Transcription-level changes in EMT-up and -down genes that occur in response to the medium of Asns-silenced 4T1-T cells being supplemented with l-asparagine

(n = 2 replicates per condition, rank-sum $P < 0.005$ for EMT-up genes). **c**, Gene expression changes in EMT-up and -down genes that result from ATF4 knockout in near haploid KBM-7 chronic myelogenous leukaemia (HAP-1 cells, rank-sum $P < 0.05$ for EMT-up genes). **d**, Gene expression changes in EMT-up and -down genes, which result in the liver cells of homozygous ATF4-deleted mice when treated with l-asparaginase (rank-sum $P < 0.05$ for EMT-down genes in wild-type mice treated with l-asparaginase (WT+ L-asp) mice, and both EMT-up and -down genes in ATF4 mice treated with l-asparaginase (ATF4+ L-asp) mice). **e**, Volumes of tumours resulting from orthotopic injection of Tgf-β-silenced and -expressing 4T1-T cells (n = 10 mice per cell line). **f**, Percentage of Twist1-positive regions based on IHC staining of sections from tumours described in **e** (n = 5 tumour sections per cell line, rank-sum $P < 0.01$). **g**, Percentage of Cdh1-positive regions based on IHC staining of sections from tumours described in **e** (n = 5 tumour sections per cell line, rank-sum $P < 0.01$). **h**, Quantification of metastases resulting from the tumours described in **e** (rank-sum $P < 0.05$). **i**, Quantification of metastases resulting from intravenous injection of Tgf-β -silenced and -expressing cells (n = 10 mice per cell line, rank-sum $P < 0.05$). See Source Data.
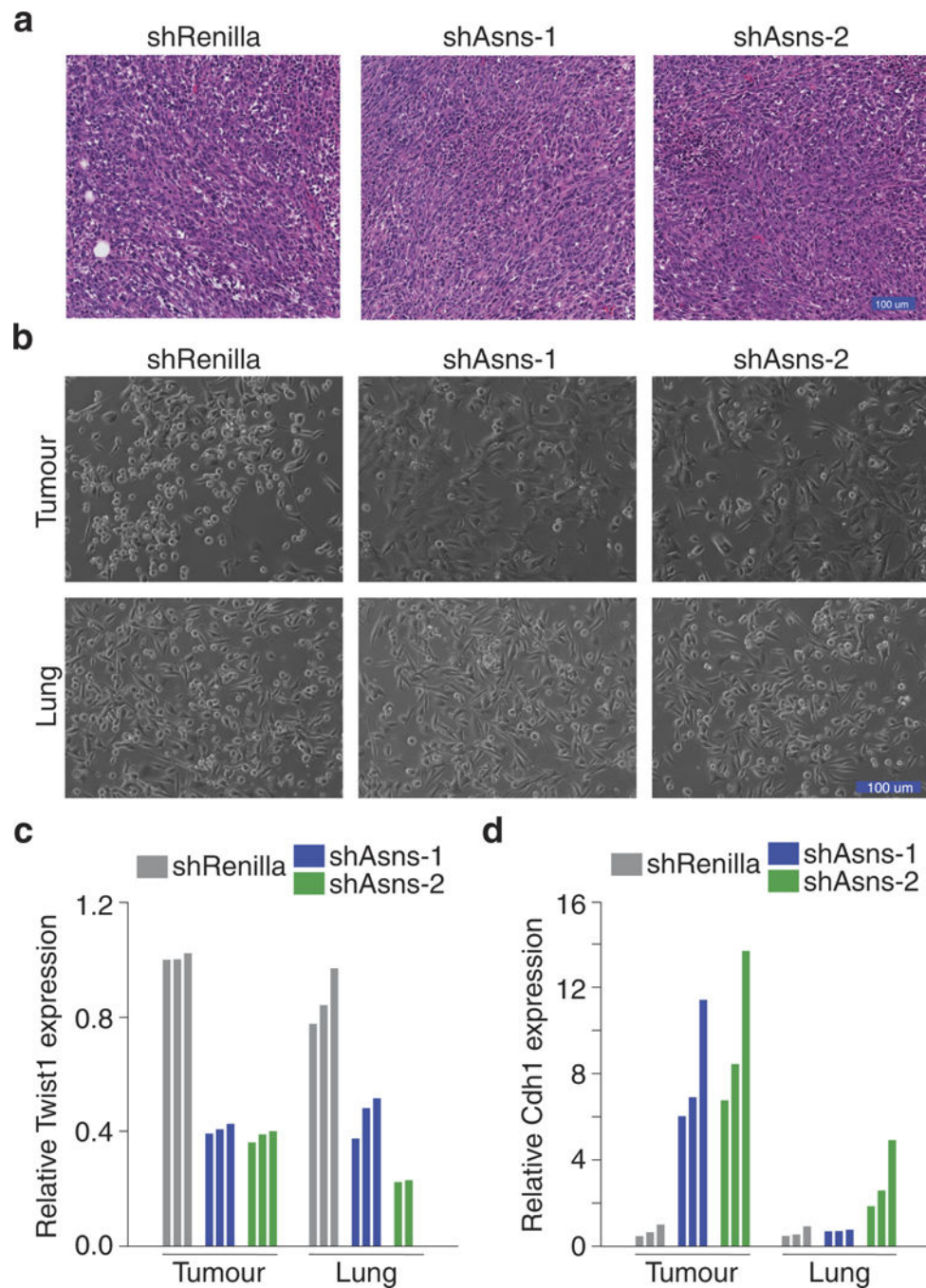
Extended Data Figure 10. Tertiary validation that asparagine availability regulates EMT
a, Representative H&E-stained sections of the tumours described in Fig. 2c, in which Asns-silenced and -expressing 4T1-T cells were orthotopically injected into NSG mice. **b**, Images of cultured cells after they were isolated from the tumours and lungs of mice injected orthotopically with Asns-silenced and -expressing 4T1-T cells. **c**, Relative Twist1 expression, as measured by qPCR, which were sorted from the tumours and lungs of mice injected orthotopically with Asnssilenced and -expressing 4T1-T cells ($n = 3$ tumours and

lungs per cell line). **d**, Relative Cdh1 expression, as measured by qPCR, in the tumours and lungs described in **c** ($n = 3$ tumours and lungs per cell line).
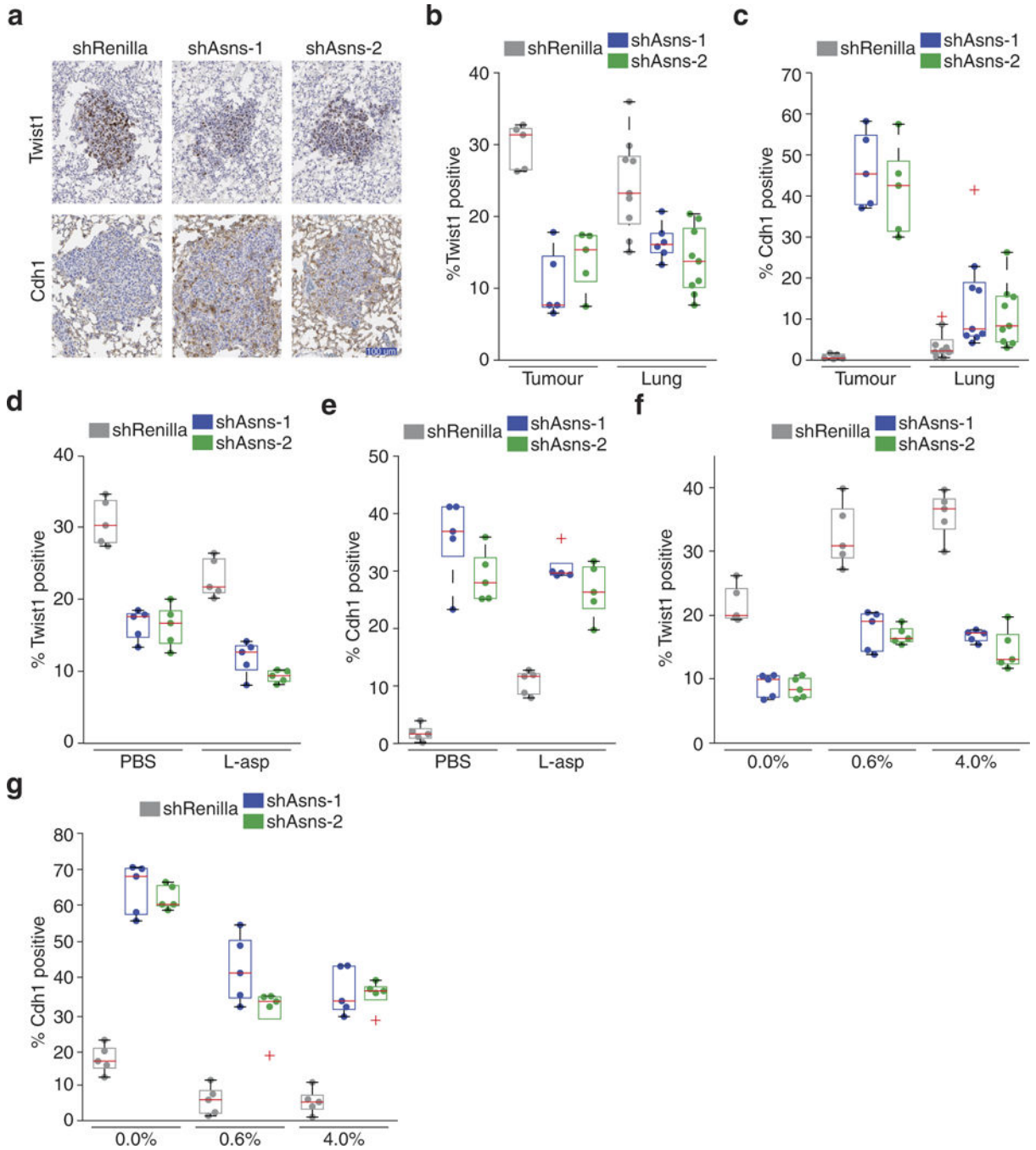
Extended Data Figure 11. Quaternary validation that asparagine availability regulates EMT

a, Representative images of IHC staining for Twist1 and Cdh1 on sections from lungs described in Fig. 4e, in which mice were injected orthotopically with Asns-silenced and -expressing 4T1-T cells. **b**, Quantification of all Twist1 stainings, described in Fig. 4e and **a** ($n = 5$ tumour sections and $n > 5$ lung metastases, edges of the box are the 25th and 75th

percentiles and error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also the case for **c**–**g**, rank-sum $P < 0.01$ and $P < 0.05$ for Asns-silenced versus -expressing tumours and metastases, respectively). **c**, Quantification of all Cdh1 stainings, described in **b** ($n = 5$ tumour sections and $n = 9$ lung metastases, rank-sum $P < 0.01$ and $P < 0.05$ for Asns-silenced versus -expressing tumours and metastases, respectively). **d**, Quantification of Twist1-positive regions in the tumours resulting from orthotopic injection of Asns-expressing and -silenced 4T1-T cells into animals treated with PBS or l-asparaginase ($n = 5$ tumour sections per condition, rank-sum $P < 0.01$ for Asns-silenced versus -unsilenced cells and rank-sum $P < 0.05$ for each cell line in treated versus untreated mice). **e**, Quantification of Cdh1-positive regions in the tumours described in **d** ($n = 5$ tumour sections per condition, rank-sum $P < 0.01$ for Asns-silenced versus -unsilenced cells and rank-sum $P < 0.05$ for each cell line in treated versus untreated mice). **f**, Quantification of Twist1- positive regions in tumours resulting from orthotopic injection of Asnsexpressing and -silenced cells into mice fed a 0%, 0.6%, or 4% asparagine diet ($n = 5$ tumour sections per condition, rank-sum $P < 0.01$ between Asns-silenced and -expressing cells and between diets). **g**, Quantification of Cdh1-positive regions in the tumours described in **f** ($n = 5$ tumour sections per condition, rank-sum $P < 0.01$ between Asns-silenced and -expressing cells and between diets).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Wagenblast E, et al. A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. Nature. 2015; 520:358–362. [PubMed: 25855289]

2. Miller FR, Miller BE, Heppner GH. Characterization of metastatic heterogeneity among subpopulations of a single mouse mammary tumor: heterogeneity in phenotypic stability. Invasion Metastasis. 1983; 3:22–31. [PubMed: 6677618]

3. Harrell JC, et al. Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse. Breast Cancer Res Treat. 2012; 132:523–535. [PubMed: 21671017]

4. Knott SRV, et al. A computational algorithm to predict shRNA potency. Mol Cell. 2014; 56:796–807. [PubMed: 25435137]

5. Zhang J, et al. Asparagine plays a critical role in regulating cellular adaptation to glutamine depletion. Mol Cell. 2014; 56:205–218. [PubMed: 25242145]

6. Krall AS, Xu S, Graeber TG, Braas D, Christofk HR. Asparagine promotes cancer cell proliferation through use as an amino acid exchange factor. Nat Commun. 2016; 7:11457. [PubMed: 27126896]

7. Richards NG, Schuster SM. Mechanistic issues in asparagine synthetase catalysis. Adv Enzymol. 1998; 72:145–198. [PubMed: 9559053]

8. Richards NGJ, Kilberg MS. Asparagine synthetase chemotherapy. Annu Rev Biochem. 2006; 75:629–654. [PubMed: 16756505]

9. Tallal L, et al. *E. coli* l-asparaginase in the treatment of leukemia and solid tumors in 131 children. Cancer. 1970; 25:306–320. [PubMed: 4905155]

10. Stams WAG, et al. Asparagine synthetase expression is linked with l-asparaginase resistance in TEL-AML1-negative but not TEL-AML1-positive pediatric acute lymphoblastic leukemia. Blood. 2005; 105:4223–4225. [PubMed: 15718422]

11. GTEx Consortium. et al. Genetic effects on gene expression across human tissues. Nature. 2017; 550:204–213. [PubMed: 29022597]

12. Loayza-Puch F, et al. Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. Nature. 2016; 530:490–494. [PubMed: 26878238]

13. Taube JH, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci USA. 2010; 107:15449–15454. [PubMed: 20713713]

14. Gowen BG, et al. A forward genetic screen reveals novel independent regulators of ULBP1, an activating ligand for natural killer cells. eLife. 2015; 4:1876.

15. Al-Baghdadi RJT, et al. Role of activating transcription factor 4 in the hepatic response to amino acid depletion by asparaginase. Sci Rep. 2017; 7:1272. [PubMed: 28455513]

16. McEarchern JA, et al. Invasion and metastasis of a mammary tumor involves TGF-β signaling. Int J Cancer. 2001; 91:76–82. [PubMed: 11149423]

17. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009; 10:48. [PubMed: 19192299]

18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]

19. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–169. [PubMed: 25260700]

20. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]

21. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol. 2001; 23:291–299. [PubMed: 11531144]

22. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011; 17:10.

23. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. [PubMed: 23104886]

24. Vizcaíno. et al. 2016 update of the PRIDE database and related tools. Nucleic Acids Res. 2016; 44:D447–D456. [PubMed: 26527722]
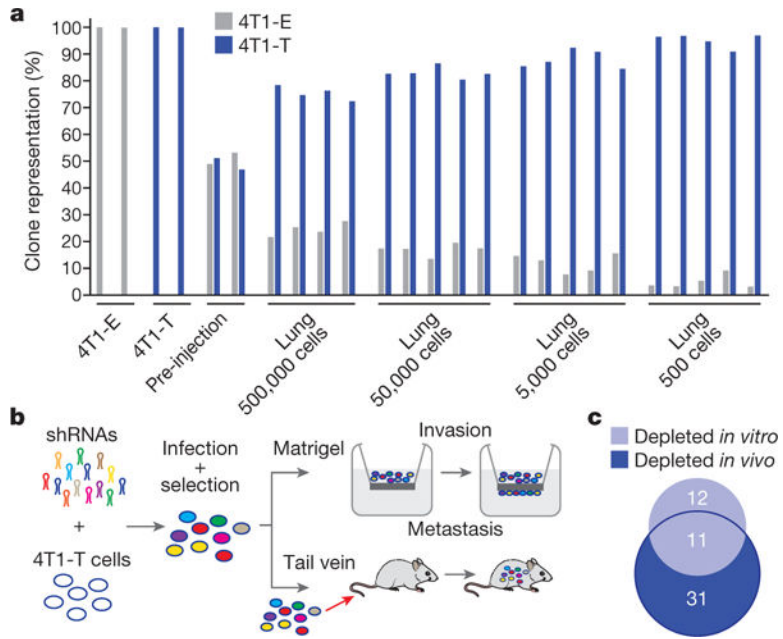
Figure 1. Identification of metastatic drivers
a, Relative proportions of 4T1-E and -T cells extracted from the lungs of NSG mice, into which mixtures of cells were introduced via tail vein at different concentrations. Each bar represents a sample or independent mouse. **b**, RNAi screening scheme to identify drivers of invasion *in vitro* and extravasation and colonization *in vivo* ($n = 5$ mice or $n = 2$ Matrigel six-well invasion chambers per approximately 50-construct shRNA pool, gene-level hit calls with empirical Bayes-moderated *t*-test false discovery rate < 0.05 and 0.1 for *in vivo* and *in vitro* screens, respectively). **c**, Overlap between genes identified in each arm of the RNAi screen depicted in **b** (hypergeometric test $P < 0.01$).
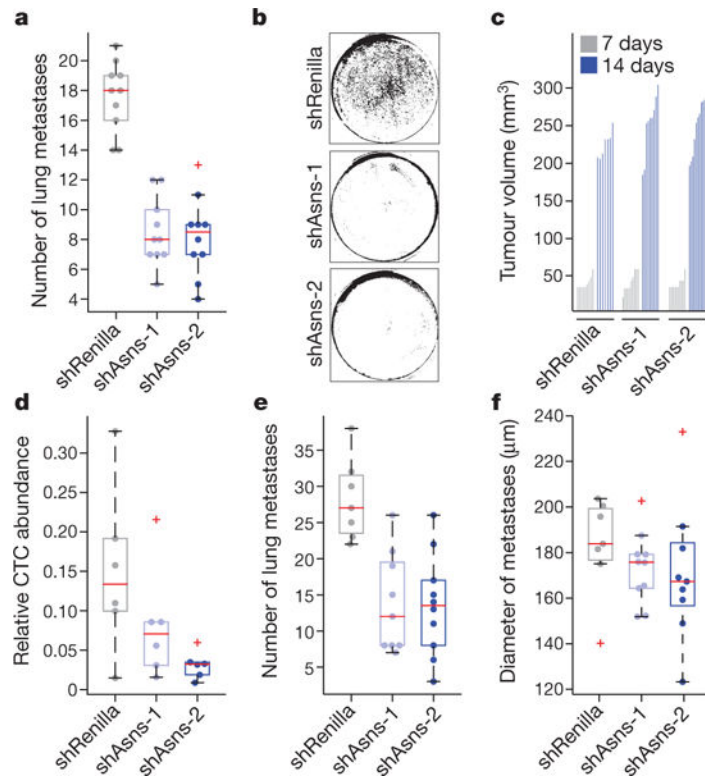
Figure 2. Validation of Asns as a driver of invasion and metastasis

a, Quantification of metastases in the lungs of mice intravenously injected with Asns-silenced or -expressing 4T1-T cells ($n = 10$ mice per cell line, edges of the box are the 25th and 75th percentiles, and error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also true for **d**–**f**, rank-sum test $P < 0.001$). **b**, Representative images of collection wells of the Matrigel assay after Asns-silenced and -expressing cells were applied 24 h previously ($n = 3$ invasion chambers per cell line). **c**, Tumour volumes resulting from the orthotopic injection of the cells described in **a** ($n = 10$ mice per cell line). **d**, Relative abundance of CTCs in animals corresponding to the tumours described in **c** ($n = 6$ mice per cell line, rank-sum $P < 0.05$ for shAsns-2). **e**, Quantification of metastases in H&E-stained lung sections, from mice described in **d** (rank-sum $P < 0.0002$). **f**, Average diameters of the metastases of each mouse described in **e**. See Source Data.
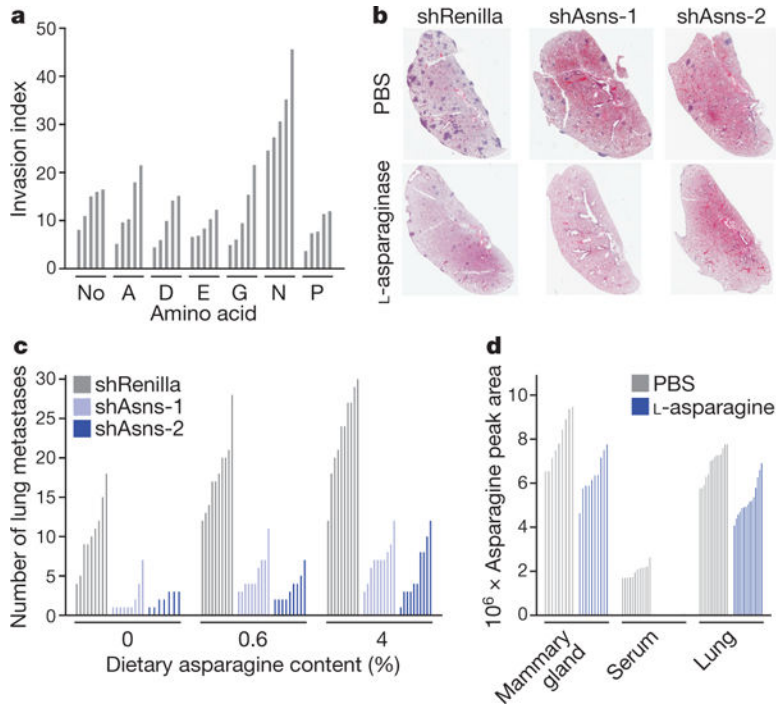
Figure 3. Extracellular asparagine availability drives invasion and metastasis

a, Quantification of parental 4T1 cell invasion rates, as measured by the Matrigel invasion assay, in culture medium supplemented with the indicated NEAAs ($n = 5$ invasion chambers, rank-sum $P < 0.01$). **b**, Representative H&E-stained lung sections from animals injected with Asns-silenced or -expressing 4T1-T cells. Animals were administered l-asparaginase or PBS ($n = 10$ mice per condition). **c**, Quantification of lung metastases in animals injected with Asns-silenced or -expressing 4T1-T cells. Animals were administered a diet with either 0%, 0.6%, or 4% asparagine content for the duration of the experiment ($n = 10$ mice per condition, ranksum $P < 0.05$ for Asns-silenced versus -expressing cells across all diets, for each cell line with 0% versus 4% diets, for shRenilla and shAsns-1 infected cells with 0% versus 0.6% diet, and for unsilenced cells with 0.6% versus 4% diet). **d**, Mass-spectrometric quantification of the asparagine levels in the mammary gland, blood serum, and lungs of animals administered l-asparaginase or PBS (relative abundance normalized by total metabolite peak area, $n > 8$ tissue sections per condition, rank-sum $P < 0.005$ for PBS versus l-asparaginase across all tissues, rank-sum $P < 0.05$ for mammary gland versus lung, and rank-sum $P < 0.0005$ for serum versus lung and serum versus mammary gland). See Source Data.
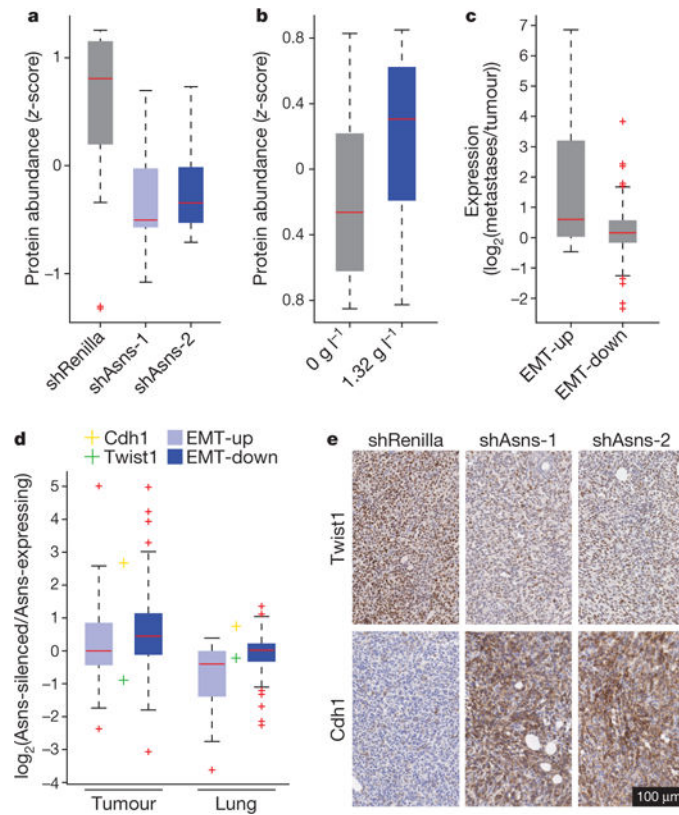
Figure 4. Asparagine availability regulates EMT

a, Relative protein abundances of EMT-up proteins in Asns-silenced and -expressing 4T1- T cells ($n = 3$ replicates per cell line, edges of the box are the 25th and 75th percentiles, and error bars extend to the values $q3 + w(q3 - q1)$ and $q1 - w(q3 - q1)$, in which w is 1.5 and $q1$ and $q3$ are the 25th and 75th percentiles, which is also true for **b**–**d**, rank-sum $P < 0.01$).
b, Relative protein abundances of EMT-up proteins in 4T1-T cells when grown in normal and asparagine-supplemented medium ($n = 3$ replicates per cell line, rank-sum $P < 0.05$). **c**, Relative expression levels of EMT-up and -down genes in lung metastases versus primary tumours derived from orthotopic injection of 4T1-T cells ($n = 4$ mice, rank-sum $P < 5.0 \times 10^{-9}$ for EMT-up genes). **d**, Relative expression levels of Twist1, Cdh1, and EMT-up and -down genes in cells isolated from tumours and lungs derived from Asns-silenced versus -expressing 4T1-T cells ($n = 4$ replicates per condition, rank-sum $P < 5.0 \times 10^{-11}$ for EMT-down genes in the tumour and rank-sum $P < 5.0 \times 10^{-8}$ for EMT-up genes in the lung; Cdh1 was differentially expressed in the tumours and lungs, and Twist1 was differentially expressed in tumours; DESeq false discovery rate < 0.05). **e**, Representative images of immunohistochemistry (IHC) stainings for Twist1 and Cdh1 in orthotopic tumours derived from Asns-silenced and -expressing 4T1-T cells. See Source Data.