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], 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] CRUK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK

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

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

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

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

[6] New York Genome Center, 101 6th Ave, New York, NY, 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, CA, USA

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

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

[11] Division of Hematology and Oncology, University of North Carolina at Chapel Hill, 170 Manning Dr, CB7305, Chapell Hill, NC, USA

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

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

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

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

*these authors contributed equally to this work
[#]to whom correspondence should be addressed: Greg.Hannon@cruk.cam.ac.uk

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* RNAi screens revealed candidate drivers of metastasis that discriminated metastatic clones. Among these, Asparagine Synthetase (ASNS) expression in a patient's primary tumour was most strongly correlated with later metastatic relapse. Here, we have shown that asparagine bioavailability strongly influences metastatic potential. Limiting asparagine by Asns knockdown, treatment with L-asparaginase, or dietary asparagine restriction reduced metastasis without impacting growth of the primary tumour, whereas increased dietary asparagine or enforced Asns expression promoted metastatic progression. Altering asparagine availability *in vitro* strongly influenced invasive potential, and this was correlated with an impact 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.

The majority of 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. In order 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 prior studies of a murine 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). Though the two clones were initially present in equal abundance (Fig. 1a), when cells were harvested 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, e.g. 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 (Extended Data Fig. 1a)[3]. They were also more highly expressed in the primary tumours of patients with later relapse to the bone, brain, and lungs as compared to primary tumours of relapse-free survivors (Extended Data Fig. 1b for lung).

To identify metastatic drivers, we carried out an 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 placed onto matrigel or introduced into NSG mice by tail vein injection. After 24 hours, the cells that had invaded through matrigel were collected and, after 7 days, lungs were harvested 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, 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 breast cancer patient datasets. 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 as compared to relapse-free survivors (Extended Data Fig. 1e). Subsequent analyses identified ASNS as predictive of survival in three additional breast cancer patient datasets (Extended Data Fig. 1f). In addition to breast, ASNS is negatively correlated with survival in 4 out of the 10 other solid tumours in the 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 impact proliferation *in vitro*; however, this defect was minor compared to 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 media 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 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 media separately with other non-essential amino acids lacking in the culture media, 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 2-fold increase in invasiveness (Fig. 3a), though levels of uptake were similar for each of the amino acids with the exception of aspartic and glutamic acid (Extended Data Fig. 4a). More profound impacts 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).

Since invasiveness could be modulated either by altering asparagine biosynthetic capacity or by modifying extracellular pools, we asked whether metastasis could be influenced by treatment with L-asparaginase. This enzyme is used to treat acute lymphoblastic leukaemia (ALL), which is generally highly dependent on extracellular asparagine[7,8]. L-

asparaginase has proven ineffective for treating solid tumours, in accord with Asns silencing not impacting 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). While no significant difference, compared to controls, was detected in primary tumours, a reduction in metastasis was observed (Extended Data Fig. 5a-c).

In TEL-AML1 negative ALL patients, resistance to L-asparaginase is thought to be achieved through increased biosynthetic production of asparagine[10]. When Asns-silenced cells were injected orthotopically into L-asparaginase treated mice, 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 either a control, low-asparagine, or high-asparagine chow (0.6%, 0%, and 4%, respectively). HPLC confirmed that serum asparagine levels were significantly altered in concordance with dietary intake (Extended Data Fig. 6a). Asparagine restriction did not impact primary tumour growth, regardless of Asns-expression status (Extended Data Fig. 6b). In 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 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 suggest 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 impact of Asns silencing or changes in global asparagine levels on primary tumour growth, while 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 impacts if asparagine levels were altered in patients[11].

To understand the mechanism by which asparagine availability might impact invasion and metastasis, we examined expression changes induced by Asns silencing, both at the RNA and protein level. RNA measurements were the strongest predictor of protein-level changes (Extended Data Fig. 7a). In accord 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 (Extended Data Fig. 7a)[12], irrespective of whether they were normalized for RNA levels (Extended Data Fig. 7b).

Amongst the asparagine-enriched proteins that were depleted upon Asns silencing, we found genes whose human orthologues were up-regulated after induction of the epithelial-to-mesenchymal transition (EMT) (Fig. 4a and Extended Data Fig. 7c, Supplementary Table 6, EMT-up proteins)[13]. Overall, depleted proteins had 10% higher asparagine content than the analysed proteome as a whole, while EMT-up proteins had 19% higher asparagine content (Extended Data Fig. 7c). Human EMT-up proteins are also asparagine-enriched (Extended Data Fig. 7c). A reanalysis 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 parental 4T1 cells were cultured in elevated asparagine (Fig. 4b)[12]. 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 down-regulated at the transcriptional level in Asns-silenced cells (Extended Data Fig. 9a). EMT-up genes were also increased in their mRNA levels when parental 4T1 cells were grown in asparagine supplemented media (Extended Data Fig. 9b). A reanalysis 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 WT mice (Extended Data Fig. 9d)[14,15]. Considered together, these data suggested that asparagine bioavailability might impact 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 Tgf-ß, a key driver of EMT (Supplementary Table 4)[16]. 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-ß-silenced cells produced fewer metastases from the primary tumour or when intravenously injected (Extended Data Fig. 9h, i).

Although no differences were detected in H&E stained tumour sections, morphological distinctions were noticeable when shRNA infected 4T1-T cells were isolated from primary tumours by 6-TG selection, with most Asns-silenced cells displaying an epithelial morphology (Extended Data Fig. 10a, b). The majority of 6-TG 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 down-regulated in Asns-silenced vs. –expressing metastatic cells, indicating an increased representation of epithelial cells in the silenced populations. Similarly, EMT-down genes were up-regulated, 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 FACS isolated 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 is also likely relevant in human cancers, as high ASNS expression is a marker of poor prognosis for many tumour types. One mechanism underlying our findings is likely a 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 site, which could also impact both tumour progression and response to therapy.

Acknowledgements
This work was performed with assistance from CSHL Shared Resources, which are funded, in part, by the Cancer Center Support Grant 5P30CA045508. We thank M. Mosquera, M. Cahn, J. Coblentz, L. Bianco for support with mouse work and K. Cheng, J. Bourbonniere, D. Hoppe, A. Nourjanova and R. Puzis for histology support. We thank E. Hodges and E. Lee for support with next-generation sequencing. We thank J. Johnson for assistance with HPLC. This work was also performed with the assistance of the CRUK, Cambridge Institute Proteomics Core Facility. We thank E. Papachristou and C. D'Santos for their help with proteomic data. S.R.V.K. is supported by a fellowship from The Hope Funds For Cancer Research. E.W. is supported by a long-term fellowship from the Human Frontier Science Program. L.A.C is supported by the Susan G. Komen Foundation (SAC110006) and the NCI Breast SPORE program (P50-CA58223-09A1). J.C.H. and C.M.P. are supported by funds from the NCI Breast SPORE program (P50-CA58223-09A1), the Breast Cancer Research Foundation and the Triple Negative Breast Cancer Foundation. H.G. is supported by a grant from the NIH (NCI R00 CA194077). Work in the G.P. lab is supported by the ICR and a CRUK grand challenge award (C59824/A25044). G.J.H. is the Royal Society Wolfson Research Professor and is supported by core funding from Cancer Research UK, by a Program Project grant from the NIH (5 P01 CA013106-44) and by a grant from the DOD BCRP (W81XWH-12-1-0300).

Author Contributions
S.R.V.K., E.W., and G.J.H. conceived the project, supervised research and wrote the paper. S.R.V.K. and E.W. analysed experiments. S.R.V.K., E.W., S.K., S.Y.K. and M.S. performed *in vitro* and *in vivo* experiments. N.E., A.L.G., A.R.M. and S.D. assisted with virus production, shRNA cloning and library preparation. L.A.C., J.C.H. and C.M.P. assisted with human expression data. J.E.W. performed histological analysis. L.F. and H.G. assisted with proteomic analysis. M.W., M.T. and G.P. carried out metabolite profiling experiments.

Competing financial interests

C.M.P. is an equity stock holder of BioClassifier LLC and University Genomics, and has filed a patent on the PAM50 subtyping assay. S.R.V.K, E.W., and G.J.H. are seeking patent protection on the manipulation of asparagine availability for patient benefit in the metastatic setting. The remaining authors declare no competing financial interests.

References

1. Wagenblast, E. *et al.* A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* **520,** 358–362 (2015).
2. Miller, F. R., Miller, B. E. & Heppner, G. H. Characterization of metastatic heterogeneity among subpopulations of a single mouse mammary tumor: heterogeneity in phenotypic stability. *Invasion Metastasis* **3,** 22–31 (1983).
3. Harrell, J. C. *et al.* Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse. *Breast Cancer Res. Treat.* **132,** 523–535 (2012).
4. Knott, S. R. V. *et al.* A computational algorithm to predict shRNA potency. *Mol. Cell* **56,** 796–807 (2014).
5. Zhang, J. *et al.* Asparagine plays a critical role in regulating cellular adaptation to glutamine depletion. *Mol. Cell* **56,** 205–218 (2014).
6. Krall, A. S., Xu, S., Graeber, T. G., Braas, D. & Christofk, H. R. Asparagine promotes cancer cell proliferation through use as an amino acid exchange factor. *Nature Communications* **7,** 11457 (2016).
7. Richards, N. G. & Schuster, S. M. Mechanistic issues in asparagine synthetase catalysis. *Adv. Enzymol. Relat. Areas Mol. Biol.* **72,** 145–198 (1998).
8. Richards, N. G. J. & Kilberg, M. S. Asparagine synthetase chemotherapy. *Annu. Rev. Biochem.* **75,** 629–654 (2006).
9. Tallal, L. *et al.* E. coli L-asparaginase in the treatment of leukemia and solid tumors in 131 children. *Cancer* **25,** 306–320 (1970).
10. Stams, W. A. G. *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* **105,** 4223–4225 (2005).
11. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550,** 204–213 (2017).
12. Loayza-Puch, F. *et al.* Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. *Nature* **530,** 490–494 (2016).
13. Taube, J. H. *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. U.S.A.* **107,** 15449–15454 (2010).
14. Gowen, B. G. *et al.* A forward genetic screen reveals novel independent regulators of ULBP1, an activating ligand for natural killer cells. *Elife* **4,** 1876 (2015).
15. Al-Baghdadi, R. J. T. *et al.* Role of activating transcription factor 4 in the hepatic response to amino acid depletion by asparaginase. *Sci Rep* **7,** 1272 (2017).
16. McEarchern, J. A. *et al.* Invasion and metastasis of a mammary tumor involves TGF-beta signaling. *Int. J. Cancer* **91,** 76–82 (2001).

Methods

Cell culture
The mouse mammary tumour cell line 4T1 (ATCC) and any derived clonal cell line were cultured in DMEM high glucose supplemented with 5% fetal bovine serum, 5% fetal calf serum, MEM non-essential amino acids (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 have routinely been 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 in Wagenblast *et al*. 2015[1].

Animal studies
All mouse experiments were approved by the Cold Spring Harbor Animal Care and Use Committee. The maximal permitted tumour size of 20 mm in any direction was never exceeded. All mouse injections were carried out with 6 - 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 carried out using 5 x 10^5 mouse mammary tumour cells, which were re-suspended in 100 ul of PBS and injected via tail vein. Orthotopic injections were performed using 1 x 10^5 mouse mammary tumour cells or 5 x 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 ul was injected into the mammary gland number 4 for mouse mammary tumour cells and a volume of 40 ul was injected for MDA-MB-231 cells. Primary tumour volume was measured using the formula $V = 1/2(LxW^2)$, where L is length and W is width of the primary tumour. For L-asparaginase studies, mice were administered 200 ul of 60 U of L-asparaginase (Abcam) 5 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 10 animals per condition to account for the variability that is 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 is 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 in Wagenblast *et al*. 2015[1].

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

shRNAs were predicted based on the Sherwood algorithm described in Knott *et al*. 2014[4]. Pools of ~50 shRNAs were packaged in Platinum-A cells. For each pool, 10 million 4T1-T cells were infected at a multiplicity of infection (MOI) of 0.3. The infected cells were selected with 500 ug/ml hygromycin for 5 days and each pool was injected into 5 mice each via the tail vein. A pre-injection pool was collected at the time of injection to validate equal representation of each shRNA. After 7 days, mice were sacrificed and perfused with PBS to remove blood and non-extravasated cells from the lungs. Lungs were harvested and genomic DNA was isolated using phenol chloroform extraction. Genomic DNA of the pre-injection pools was isolated using the QIAamp DNA Blood Mini Kit (Qiagen).

The *in vitro* invasion assays were carried out in parallel. Each pool was plated on two 6-well BioCoat Matrigel invasion plates (Corning). 6×10^5 cells were plated on top of each well in cell culture media without serum. Cells were allowed to invade through the matrigel into media 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 the QIAamp DNA Blood Mini Kit.

The shRNAs were amplified using a two-step PCR protocol previously described in Knott et al. 2014[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 described in ref 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 was performed using the GOrilla web portal. The Refseq identifiers of genes identified as over-expressed in 4T1-T cells, as compared to

4T1-E cells, were used as foreground and the entire Refseq gene list was used as background.

Expression subtype and relapse analysis

All clinical data analysis was performed using the University of North Carolina "855 patient set", which is available as published data at https://genome.unc.edu. All data was 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 6-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 media without serum and cells invaded into media 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 media containing 100x concentration of the specified amino acid (relative to the concentration that is achieved by supplementing media with 1x NEAA) for 2 days 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% glutaaldehyde 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 the Odyssey infrared scanner. The signal was quantified using ImageJ (NIH).

Competition and proliferation assay

For the mCherry competition assay, shRNA-transduced mCherry-positive cells were admixed with untransduced cells. mCherry fluorescence was quantified on the LSR II flow cytometer (BD Biosciences). The proliferation assay was carried out using the CellTrace Violet Cell Proliferation Kit (Thermo Fisher Scientific). For Extended Data Fig. 4d, e, 4T1 and MDA-MB-231 cells were cultured in media containing 100x concentration of the specified amino acid (relative to the concentration that is achieved by supplementing media with 1x NEAA) for 2 days or 3 days, respectively, before starting the proliferation assay. Cells were stained with CellTrace violet and then trypsinized and re-suspended in media. After 24 h, cells were collected in order to quantify violet fluorescence intensity using the SH800 flow cytometer (Sony).

Isolation of tumour and lung metastatic cells

Tumour and lung tissue were harvested, minced and digested into single cells as previously reported in Wagenblast *et al.* 2015[1]. Cells were either grown in 4T1 cell culture media containing 60 μM 6-thioguanine (6-TG) to deplete stromal cells or directly sorted based on mCherry expression using the FACSAria III cell sorter (BD Biosciences).

RNAseq library preparation
RNAseq libraries from cultured 4T1-T cells were prepared in duplicates as previously described in Wagenblast *et al.* 2015[1]. Each sample was sequenced on the Illumina HiSeq sequencer generating 76 nt single-end (SE) reads.

RNAseq analysis
Illumina sequencing reads were aligned to the mouse genome (mm10) using bowtie2 with default parameters[17]. Genes were assigned a count using HTseq-count[18]. Differential expression analysis was performed using DESeq[19].

shRNA knock-down 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 ug/ml hygromycin for 5 days and MDA-MB-231 cells were selected with 2 ug/ml puromycin for 4 days. Cell lines infected with cDNA overexpressing retroviral constructs were selected with G418 for one week. The parental 4T1 cell line was selected with 600 ug/ml G418 and MDA-MB-231 cells were selected with 1500 ug/ml G418.

cDNA overexpression genes:
mouse ASNS: NM_012055.1
human ASNS: NM_001673.2

shRNA knock-down sequences:
mouse shAsns-1:
TGCTGTTGACAGTGAGCGCCACTGCCAATAAGAAAGTATATAGTGAAGCCAC
AGATGTATATACTTTCTTATTGGCAGTGTTGCCTACTGCCTCGGA
mouse shAsns-2:
TGCTGTTGACAGTGAGCGCCACTATGAAGTTTTGGATTTATAGTGAAGCCACA
GATGTATAAATCCAAAACTTCATAGTGTTGCCTACTGCCTCGGA
mouse shTgfb1-1:
TGCTGTTGACAGTGAGCGCCAGTATATATATGTTCTTCAATAGTGAAGCCACA
GATGTATTGAAGAACATATATATACTGTTGCCTACTGCCTCGGA
mouse shTgfb1-2:
TGCTGTTGACAGTGAGCGAAGTATATATATGTTCTTCAAATAGTGAAGCCAC
AGATGTATTTGAAGAACATATATATACTGTGCCTACTGCCTCGGA
human shASNS-1:
TGCTGTTGACAGTGAGCGCCAGAAGCTAAAGGTCTTGTTATAGTGAAGCCAC
AGATGTATAACAAGACCTTTAGCTTCTGATGCCTACTGCCTCGGA
human shASNS-2:

TGCTGTTGACAGTGAGCGCAGCAATGACAGAAGATGGATATAGTGAAGCCAC
AGATGTATATCCATCTTCTGTCATTGCTTTGCCTACTGCCTCGGA

qRT-PCR

Total RNA from cells was purified and DNAse treated using the RNeasy Mini Kit (Qiagen). For whole tissues, RNA was isolated using the TRIzol Plus RNA Purification Kit (Thermo 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 the Agilent Bioanalyzer (RNA nano kit). cDNA was synthesized using SuperScript III Reverse Transcriptase (Sigma). Quantitative PCR analysis was performed on the Eppendorf Mastercycler ep realplex. All signals were quantified using the ΔCt 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 the TaqMan Gene Expression Cells-to-Ct Kit (Thermo Fisher Scientific). Quantitative PCR analysis was performed on the CFX96 (Bio-Rad) using TaqMan primer/probe sets and all signals were quantified as described above.

qRT-PCR primers

mouse Asns (Exon 1 – 2):
5'-CCT CTG CTC CAC CTT CTC T-3'
5'-GAT CTT CAT CGC ACT CAG ACA-3'
mouse Asns (Exon 6 – 7):
5'-CCA AGT TCA GTATCC TCT CCA G-3'
5'-CTT CAT GAT GCT CGCTTC CA-3'
mouse Tgfb1 (Exon 1 – 2):
5'-CCG AAT GTC TGA CGT ATT GAA GA-3'
5'-GCG GAC TAC TAT GCT AAA GAG G-3'
mouse Tgfb1 (Exon 3 – 4):
5'-GTT ATC TTT GCT GTC ACA AGA GC-3'
5'-CCC ACT GAT ACG CCT GAG-3'
mouse Gapdh (Exon 2 – 3):
5'-AAT GGT GAA GGT CGGTGT G-3'
5'-GTG GAGTCA TACTGG AAC ATG TAG-3'
human ASNS (Exon 8 – 9):
5'-GAGTCA GAC CTT TGT TTA AAG CA-3'
5'-GGA GTG CTT CAATGT AAC AAG AC-3'
human ASNS (Exon 12 – 13):
5'-CTG GAT GAA GTC ATATTT TCC TTG G-3'
5'-CAG AGA AGATCA CCA CGCTAT C-3'
human GAPDH (Exon 2 – 3):
5'-ACA TCG CTC AGA CAC CAT G-3'
5'-TGT AGT TGA GGT CAA TGA AGG G-3'

TaqMan probes:

Mouse Asns: Mm00803785_m1

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

qPCR for circulating tumour cells (CTCs):
CTCs were quantified as previously described in Wagenblast *et al.* 2015[1]. Genomic DNA was isolated from blood and CTC abundance was quantified using a qPCR assay against mCherry, which is expressed from the retroviral shRNA delivery vectors.

mCherry probes and primers:
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'-GACTTGTAACGGGCAGGCAGATTGTG-3'
primer 2: 5'-GAGGTGTGGGTCACCTCGACATC-3'
HEX probe: 5'-/5HEX/CCGTGTCGC/ZEN/TCTGAAGGGCAATAT/3IABkFQ/-3'

Quantification of lung metastasis
For each lung, five-micron 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 in Wagenblast *et al.* 2015[1]. E-Cadherin (24E10) Rabbit mAb (3195, Cell Signaling) was utilized 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 hematoxylin stainings were quantified using ImageJ (NIH). For this, images were color deconvoluted according to Ruifrok *et al.* 2001[20] and the percent 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 media conditions. For Extended Data Fig. 4a, c, 4T1 and MDA-MB-231 cells were cultured in media containing 100x concentration of the specified amino acid (relative to the concentration that is achieved by supplementing media with 1x NEAA) for 2 or 3 days, respectively. All cultured cells were washed in ice cold PBS, homogenized using a Dounce homogenizer and the lysate was subsequently filtered. Each biological sample was quantified in triplicates using High Performance Liquid Chromatography (HPLC) and a flourometric detector. For each replicate, nanomoles of each amino acid was measured.

Proteomic profiling using isobaric tags for relative and absolute protein quantification (iTRAQ)

For each sample, raw peptide intensities were normalized by their sum. Peptides whose representation fell within the bottom quartile based on summed intensities across samples were not analyzed 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 peptides 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 represented in Fig. 4a and 4b, respectively).

Metabolite profiling using liquid-chromatography tandem mass spectrometry (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 6500 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 ul) was performed using 200 ul of 80% methanol at -80 °C. Following centrifugation for 10 min (13.2 kRPM, 4 °C), supernatants were evaporated to dryness and stored at -80 °C until LC-MS/MS analysis.

Dried-down extracts were re-suspended in 25 ul HPLC-grade water, and 1 ul was analysed using hydrophilic interaction chromatography (HILIC) 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 x 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 was acquired with Analyst 1.6.2 (SCIEX) and peak areas of LC-SRM-MS traces for each metabolite were integrated using the MultiQuant 1.1 software (SCIEX). Metabolite differences were analysed by normalizing samples to total peak area and comparing replicates for each group using one-way ANOVA with multiple comparisons.

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 of protein- minus RNA-level expression changes, RNA- 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 protein-level differences 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 up-regulated during EMT as compared to all other genes. For the murine 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 10 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 vs. 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 linker removal was performed using cutadapt[22]. bowtie2 was used to remove reads that map to contaminating RNAs (e.g. rRNA and tRNA sequences)[17]. STAR was subsequently used to map reads of length 29-33 to the human transcriptome[23]. The offset was corrected for each read based on 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.

Code Availability

All custom code used during the current study will be made available by the corresponding author on reasonable request.

Data Availability

All raw and processed high-throughput sequencing data is available through the Gene Expression Omnibus (GEO) under the accession numbers GSE104968 and GSE107109. Source data is available for tumour volume measurements and lung metastases. All other datasets generated during and/or analysed during the current study will be made available by the corresponding author on reasonable request.

Methods References

17. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
18. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).
19. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
20. Ruifrok, A. C. & Johnston, D. A. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23,** 291–299 (2001).
21. Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* **3,** 1154–1169 (2004).

22. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).
23. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

Figure Legends

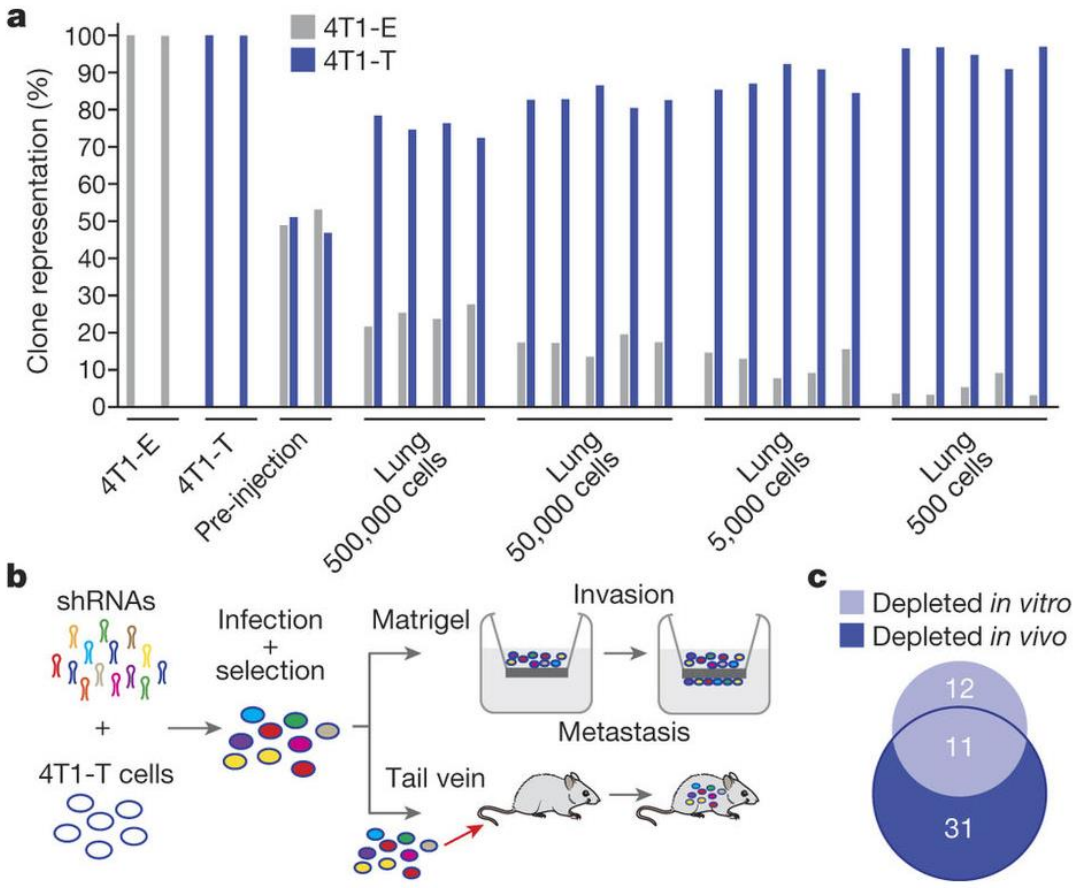


Figure 1: Identification of metastatic drivers
a, Relative proportions of 4T1-E and –T cells extracted from the lungs of NSG mice, into which mixtures of cells were introduced via tail vein at different concentrations. Each bar represents a sample or independent mouse. **b,** RNAi screening scheme to identify drivers of invasion *in vitro* and extravasation and colonization *in vivo* (n = 5 mice or n = 2 matrigel 6-well invasion chambers per ~50 construct shRNA pool, gene-level hit calls with empirical bayes moderated t-test FDR < 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-value < 0.01).

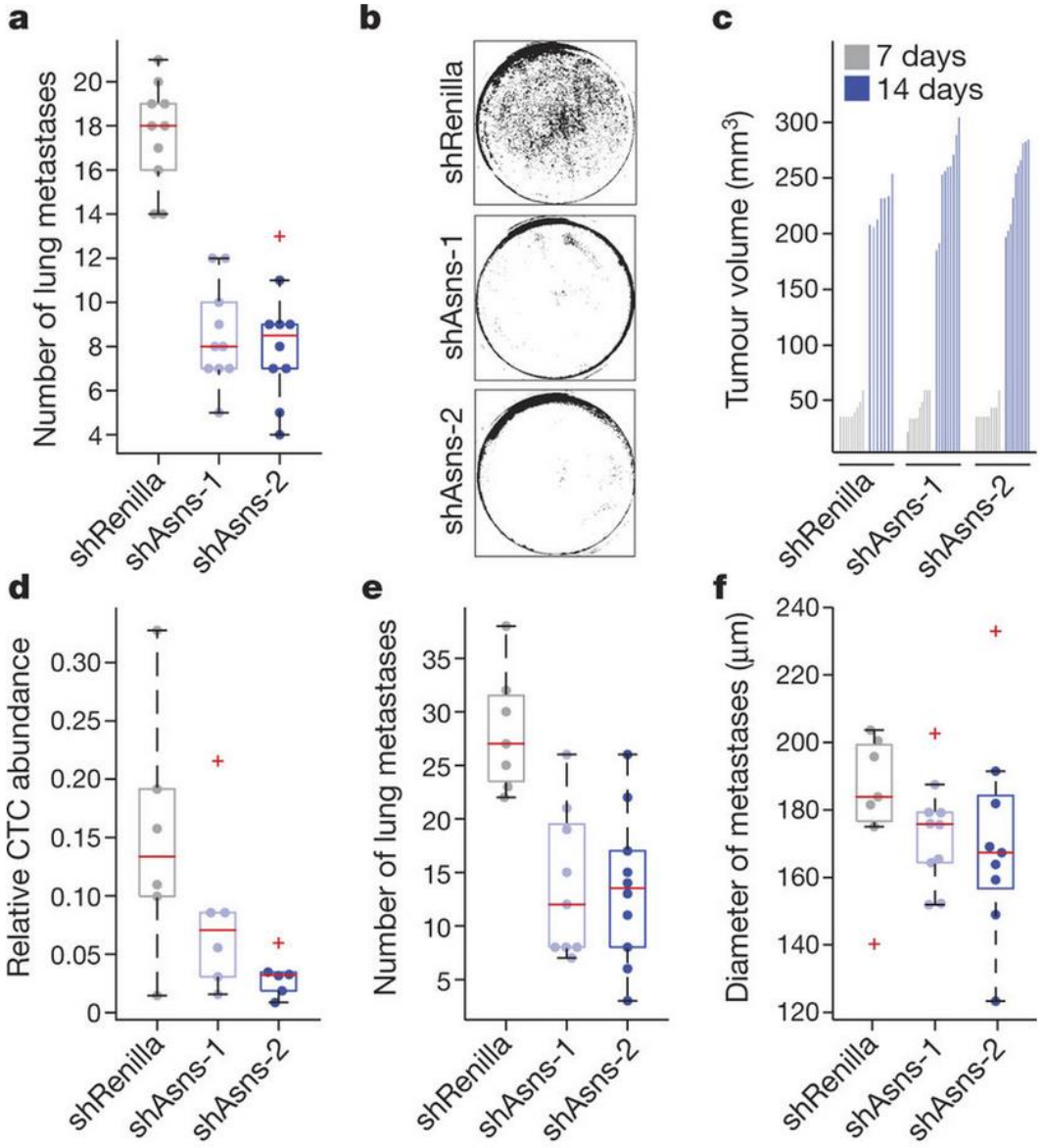


Figure 2: Validation of Asparagine Synthetase as a driver of invasion and metastasis **a,** Quantification of metastases in the lungs of mice that were intravenously injected with Asns-silenced or –expressing 4T1-T cells (n = 10 mice per cell line, edges of the box are the 25th and 75th percentiles and error bars extend to the values q3 + w(q3 − q1) and q1 − w(q3 − q1), where w is 1.5 and q1 and q3 are the 25th and 75th percentiles and this is also true for **d**, **e** and **f**, rank-sum test p-value < 0.001). **b,** Representative images of collection wells of the matrigel assay after Asns-silenced and –expressing cells were applied 24 h prior (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-value < 0.05 for shAsns-2). **e,** Quantification of metastases in H&E stained lung sections, from mice described in **d** (rank-sum p-value < 0.0002). **f, Average** diameters of the metastases of each mouse described in **e**. See source data.

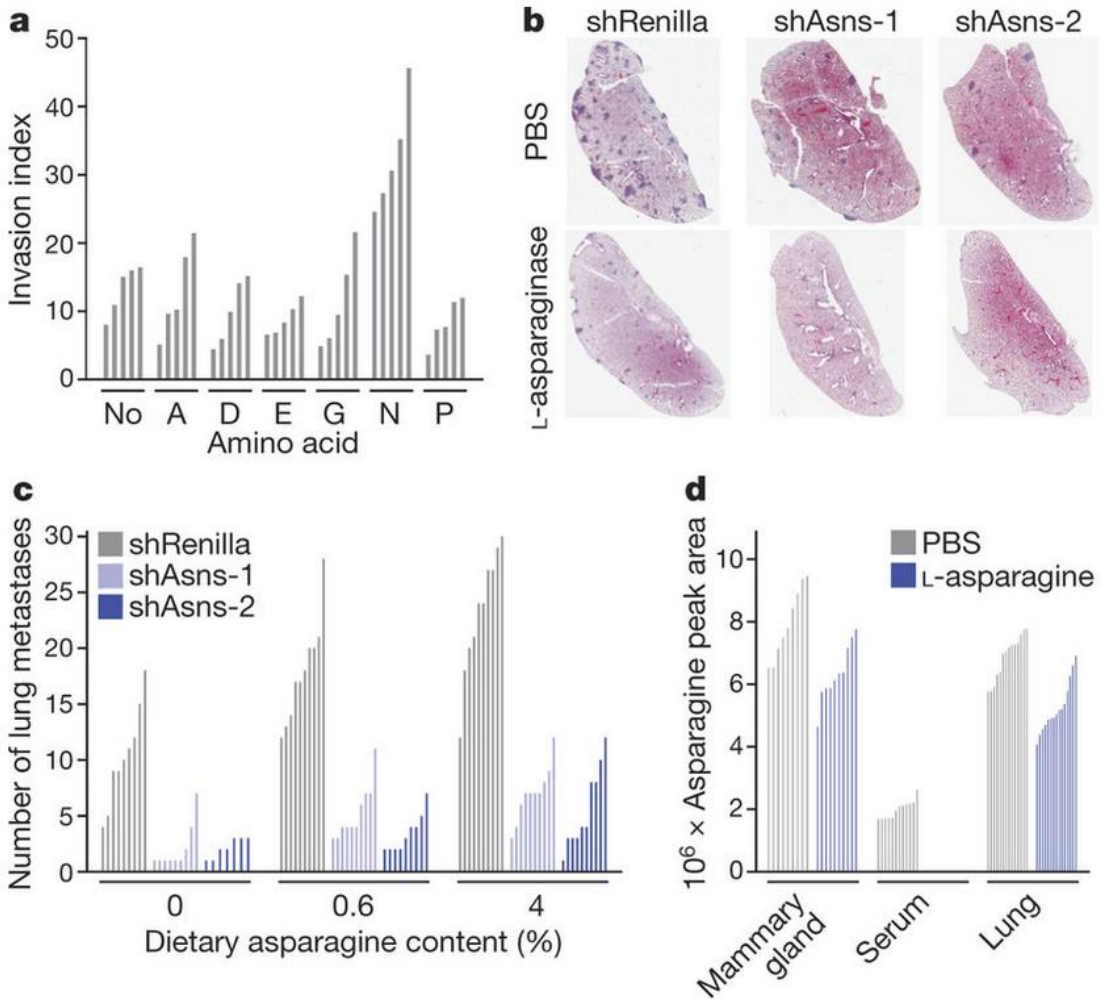


Figure 3: Extracellular asparagine availability drives invasion and metastasis
a, Quantification of parental 4T1 cell invasion rates, as measured by the matrigel invasion assay, in culture media supplemented with the indicated non-essential amino acids (n = 5 invasion chambers, rank-sum p-value < 0.01). **b,** Representative H&E stained lung sections from animals that had been 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 that had been 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, rank-sum p-value < 0.05 for Asns-silenced vs. –expressing cells across all diets, for each cell line with 0% vs. 4% diets, for shRenila and shAsns-1 infected cells with 0% vs. 0.6% diet and for unsilenced cells with 0.6% vs. 4% diet). **d,** Mass-spectrometric quantification of the asparagine levels in the mammary gland, blood serum and lungs of animals that were administered L-asparaginase or PBS (relative abundance normalized by total metabolite peak area, n > 8 tissue sections per condition, rank-sum p-value < 0.005 for PBS vs. L-asparaginase across all tissues, rank-sum p-value < 0.05 for mammary gland vs. lung and rank-sum p-value < 0.0005 for serum vs. lung and serum vs. mammary gland). See source data.

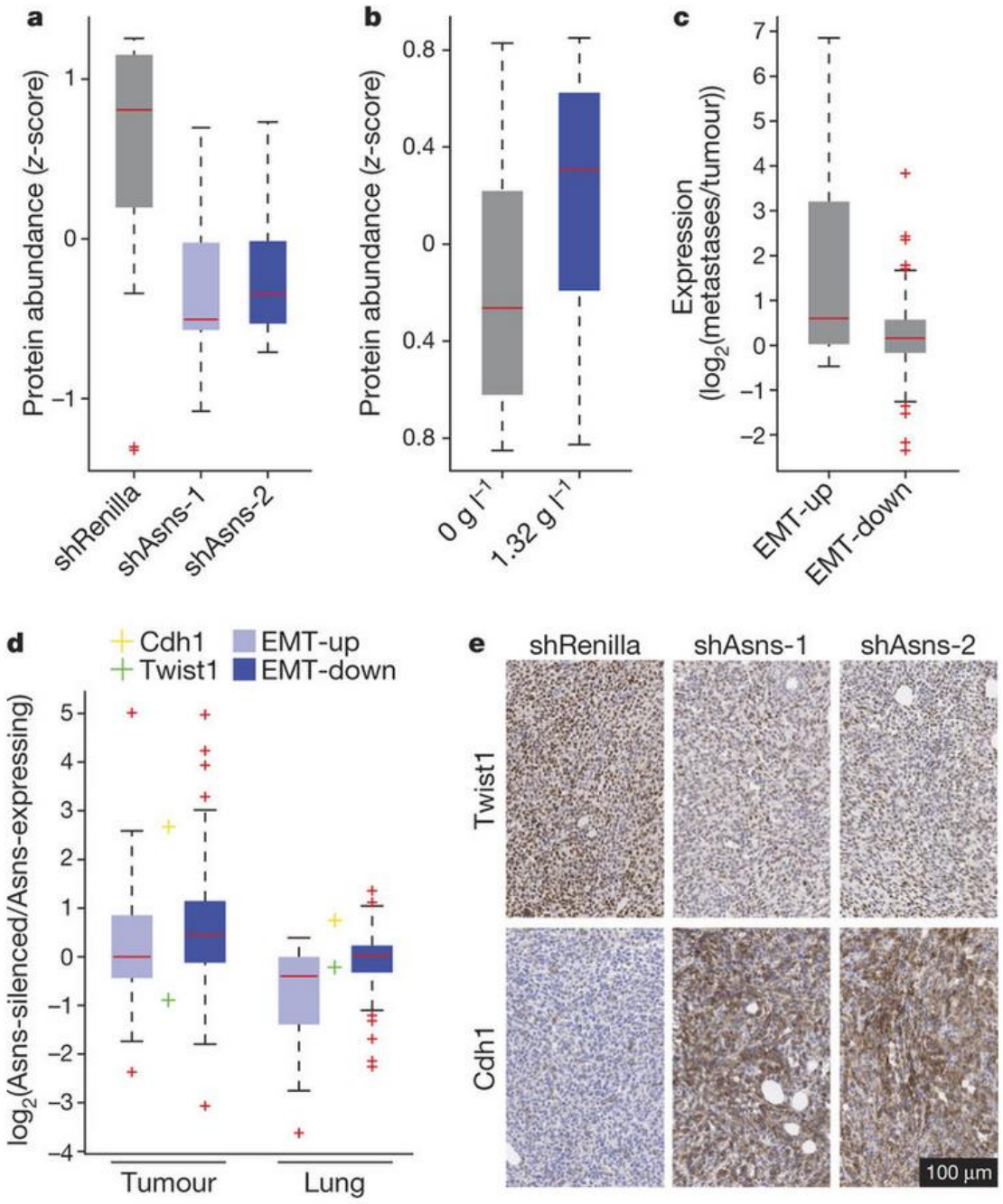
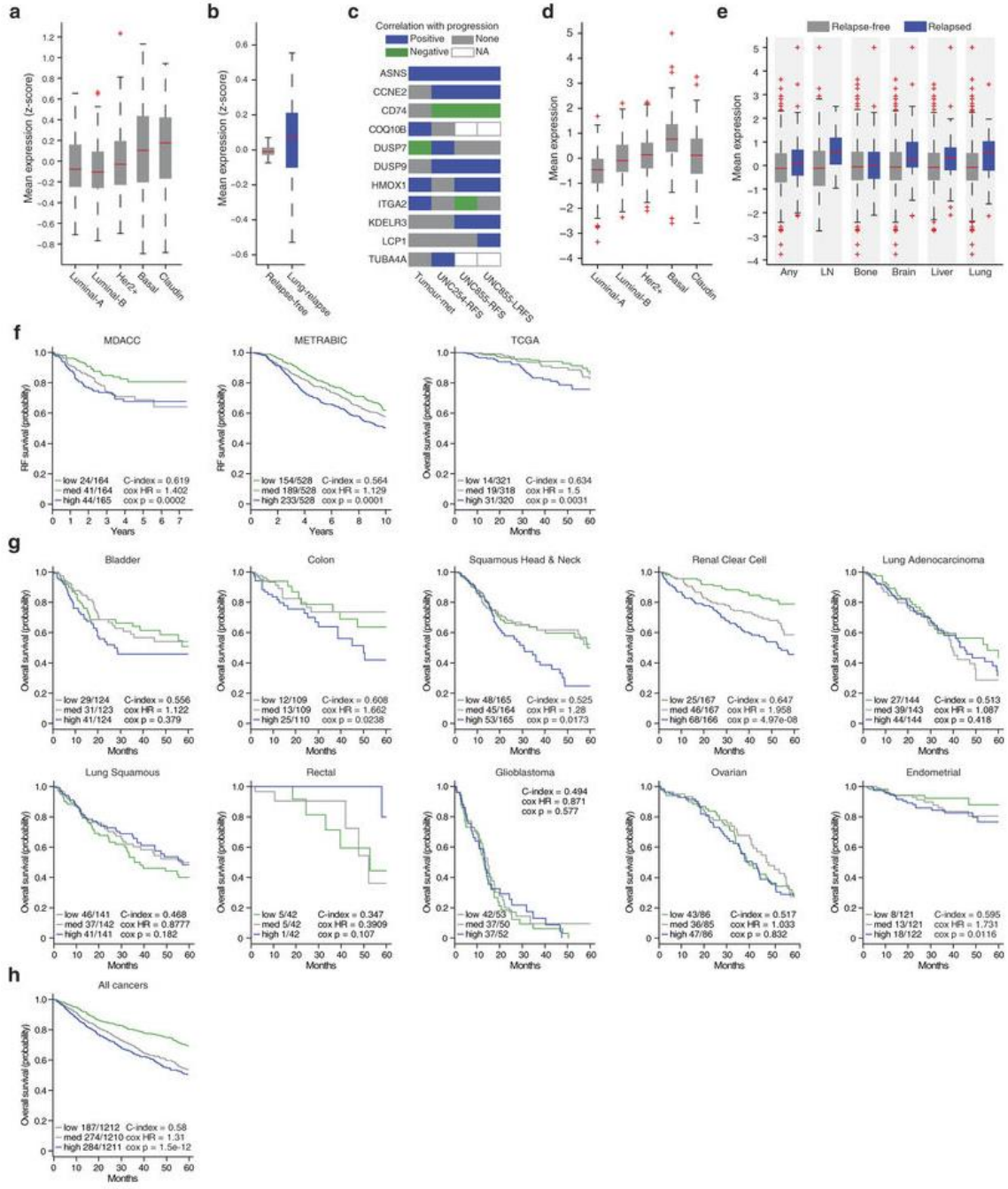


Figure 4: Asparagine availability regulates epithelial-to-mesenchymal transition
a, Relative protein abundances of EMT-up proteins in Asns-silenced and -expressing 4T1-T cells (n = 3 replicates per cell line, edges of the box are the 25th and 75th percentiles, and error bars extend to the values $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, where w is 1.5 and q_1 and q_3 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 parental 4T1 cells when grown in normal and Asparagine-supplemented media (n = 3 replicates per cell line, rank-sum p-value < 0.05). **c,** Relative expression levels of EMT-up and EMT-down genes in lung metastases vs. primary tumours derived from orthotopic injection of 4T1-T

cells (n = 4 mice, rank-sum p-value < 5.0×10^{-9} for EMT-up genes). **d,** Relative expression levels of Twist1, Cdh1, EMT-up and EMT-down genes in cells that were isolated from tumours and lungs that were derived from Asns-silenced vs. Asns-expressing 4T1-T cells (n = 4 replicates per condition, rank-sum p-value < 5.0×10^{-11} for EMT-down genes in the tumour and rank-sum p-value < 5.0×10^{-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 FDR < 0.05). **e,** Representative images of IHC stainings for Twist1 and Cdh1 in orthotopic tumours that were derived from Asns-silenced and –expressing 4T1-T cells. See source data.

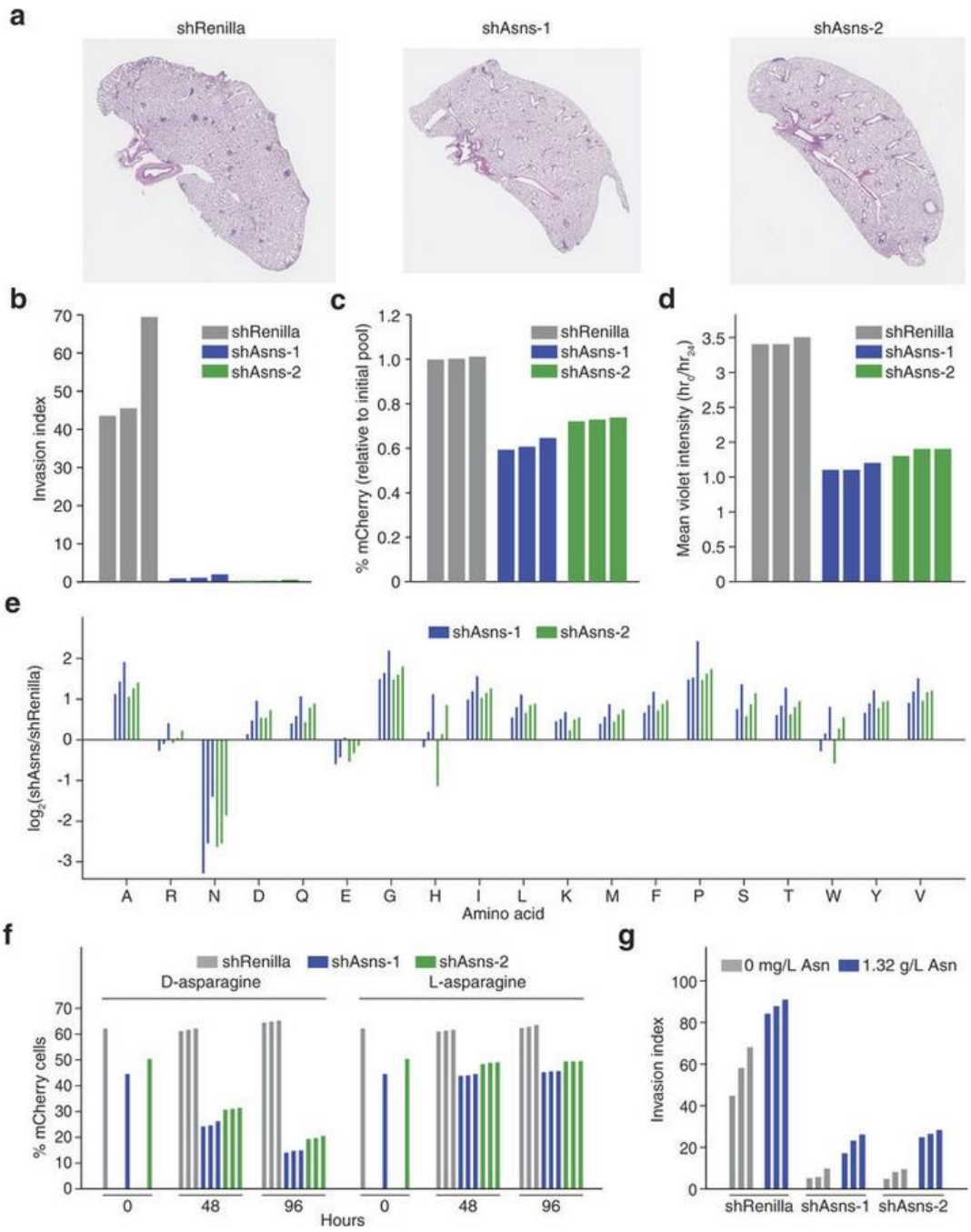
Extended Data Figure Legends



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

a, Expression level of genes identified as over-expressed in 4T1-T as compared to 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), where w is 1.5 and q1 and q3 are the 25th and 75th percentiles and this is also true for **b**, **d** and **e**, ANOVA p-value < 0.0001). **b,** Expression level of the same genes in disease free survivors and patients with relapse to the lung (rank-sum p-value < 0.01) **c,** For each gene that was identified in the screen, a prognostic value was calculated

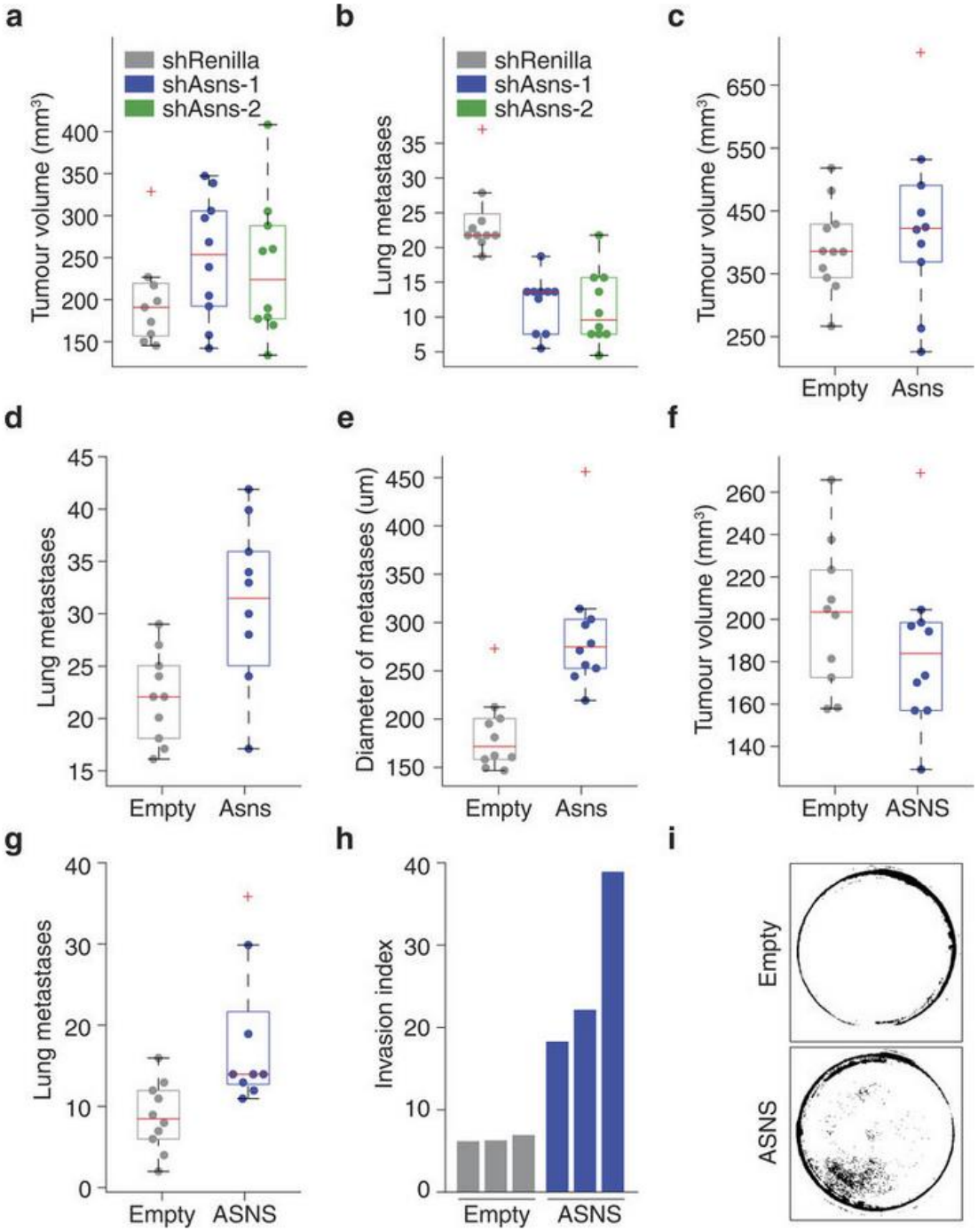
using three different datasets. One consisted of gene expression measurements in three patient-matched Basal tumour and metastasis pairs (Patients A1, A7 and A11). Here genes were classified as correlated with progression if expression was higher in each of the metastases and negatively correlated if 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 relapse free survival hazard ratios > 1, and negatively correlated if these ratios were significant (cox p-value < 0.05) and < 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. Shown are genes with human orthologues that were measured in the different datasets. **d,** Expression level of Asparagine Synthetase in the primary tumours of patients with different disease subtypes (ANOVA p-value < 0.0001). **e,** Expression level of Asparagine Synthetase in the primary tumours of patients with non-specific relapse and relapse to the lymph node, bone, brain, liver or lung in comparison to expression levels in patients without relapse to each corresponding site (rank-sum p-value < 0.005). **f,** Analysis of ASNS in 3 additional breast cancer patient sets (MDACC, METRABIC and TCGA). Shown are survival plots and relevant statistics (cox p-value < 0.01). **g,** Analysis of ASNS in the TCGA Pan Cancer expression data. Shown are survival plots and relevant statistics for the 10 non-breast solid-tumours represented in the dataset (cox p-value < 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-value < 0.0001).



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 intraveneously 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 Asparagine Synthetase. Cells were grown during the 24-hour period that the matrigel invasion assay described in **Fig. 2b** was being performed (n = 3

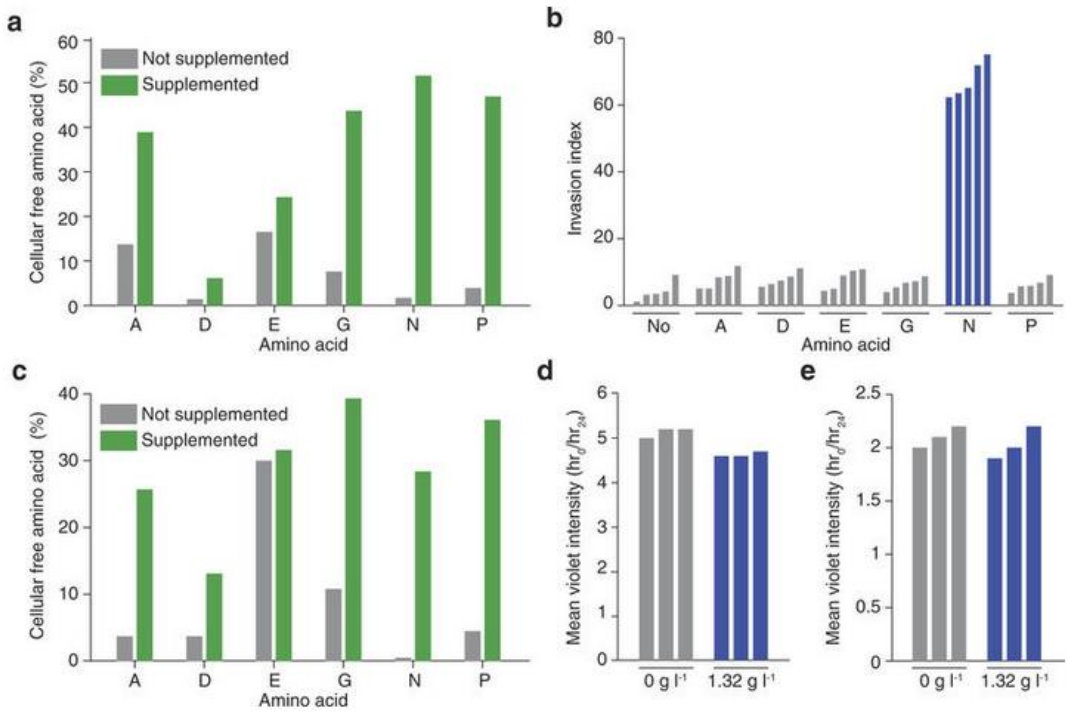
replicates per cell line). **d,** Violet cell-labelling intensity of Asns-silenced and –expressing 4T1-T cells, relative to the initial population. Cells were grown during the 24-hour period that the matrigel invasion assay described in **Fig. 2b** was being performed (n = 3 replicates per cell line). **e,** Free amino acid quantification by HPLC for each amino acid in Asns-expressing and –silenced cells. Shown are the log2-fold-changes in nanomoles per 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 L-asparagine or D-asparagine supplemented media 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 media that has been 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

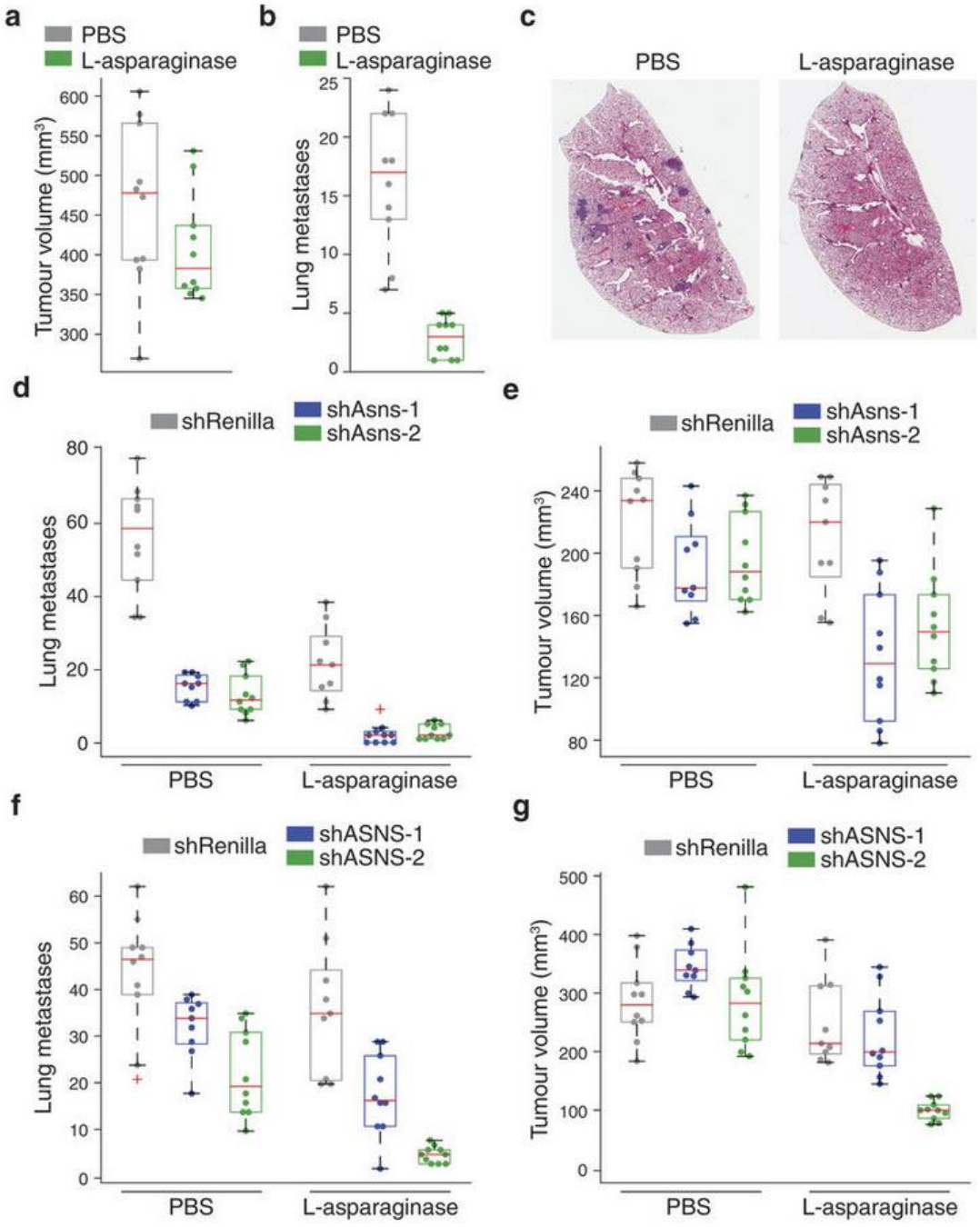
a, Volume measurements of tumours resulting from orthotopic injection of Asns-silenced and –expressing parental 4T1 cells (n = 10 mice per cell line, edges of the box are the 25th and 75th percentiles and error bars extend to the values q3 + w(q3 − q1) and q1 − w(q3 − q1), where w is 1.5 and q1 and q3 are the 25th and 75th percentiles and this is also the case for **b, c, d, e**, **f** and **g**). **b,** Quantification of lung metastases corresponding to the tumours described in a (rank-sum p-value < 0.002). **c,** Volume measurements of

tumours resulting from orthotopic injection of parental 4T1 cells with basal (Empty) or enforced (Asns) expression of Asns (n = 10 mice per cell line). **d,** Quantification of lung metastases corresponding to the tumours described in **c** (rank-sum p-value < 5.0×10^{-5}). **e,** Average diameters of the metastases of each mouse described in **d** (rank-sum p-value < 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-value < 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 impacts invasion and metastasis

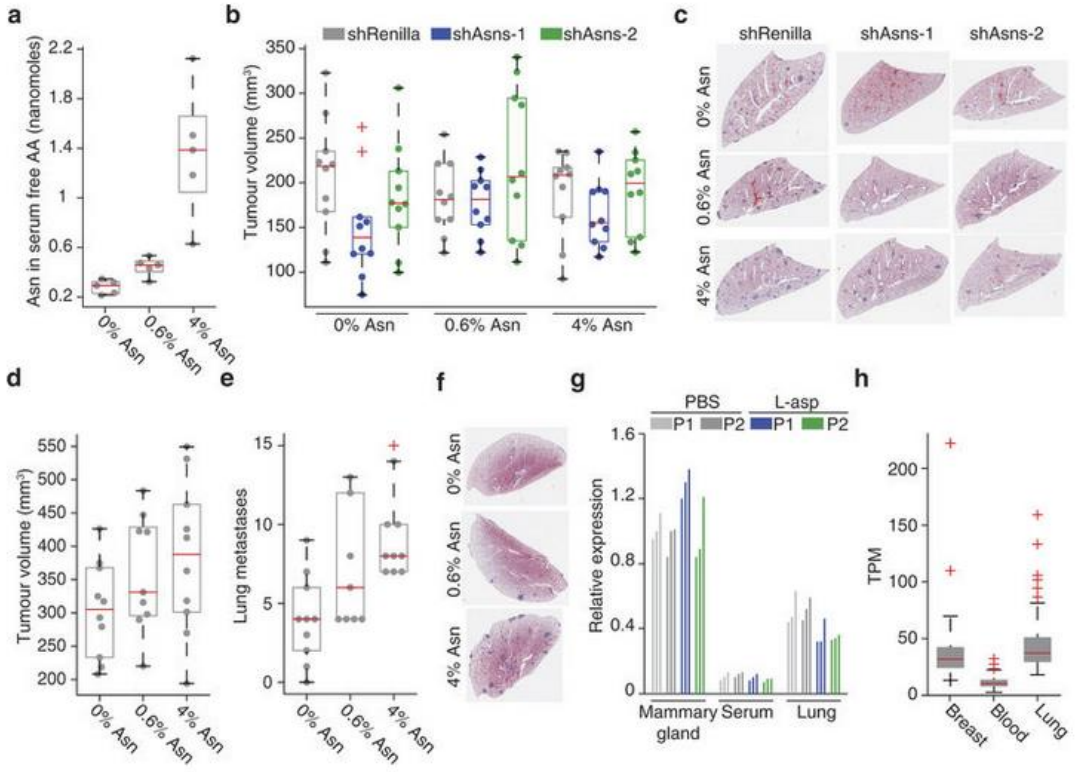
a, HPLC quantification of cellular free amino acid percentages for parental 4T1 cells when media is supplemented with each of the non-essential amino acids that is lacking in the DMEM culture media (n = 3 replicates per cell line). **b,** Quantification of MDA-MB-231 matrigel invasion rates under the same conditions as was described in **Fig. 3a** (n = 5 invasion chambers per condition, rank-sum p-value < 0.001). **c,** HPLC quantification of cellular free amino acid percentages for MDA-MB-231 cells when cultured in the media conditions described in **a** (n = 3 replicates per cell line). **d,** Violet cell-labelling intensity of parental 4T1 cells when they have been grown in asparagine-lacking and –supplemented media 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 they have been grown in asparagine-lacking and –supplemented media 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 impacts invasion and metastasis

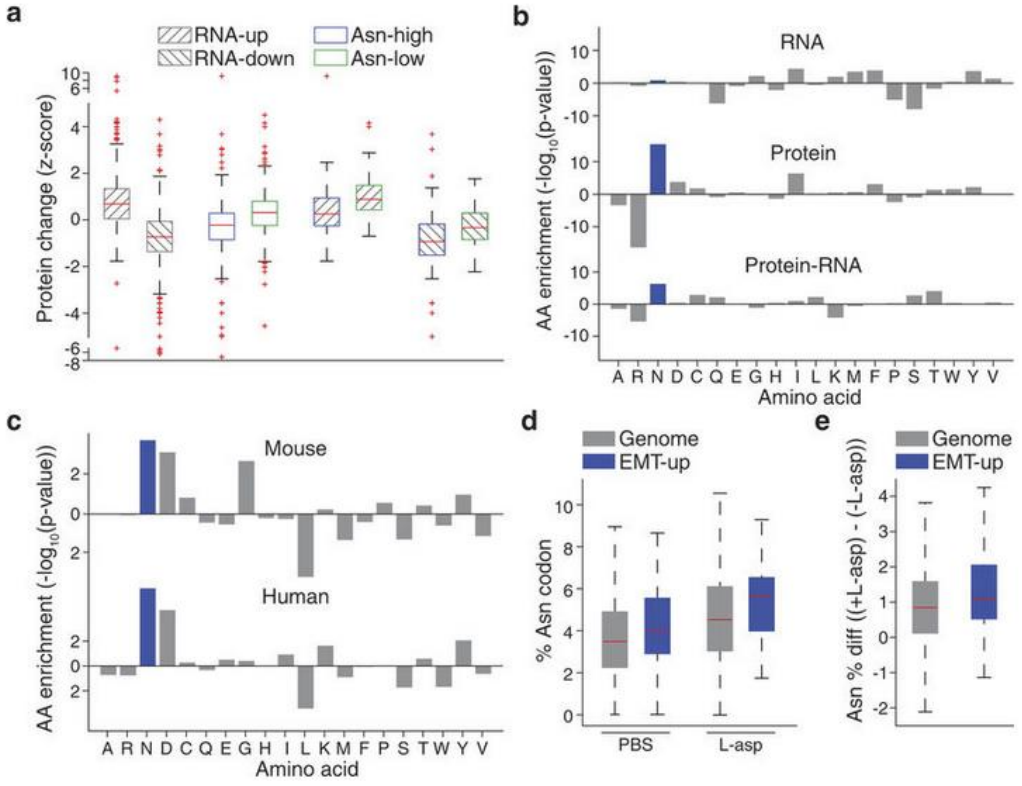
a, Tumour volumes resulting from the orthotopic injection of parental 4T1 cells. Half of the mice received L-asparaginase while the other half received an equivalent volume of PBS at the same injection rate (n = 10 mice per condition, edges of the box are the 25th and 75th percentiles and error bars extend to the values q3 + w(q3 − q1) and q1 − w(q3 − q1), where w is 1.5 and q1 and q3 are the 25th and 75th percentiles and this is also the case for **b**, **d**, **e**, **f** and **g**). **b,** Quantification of lung metastases detected in the animals described in a (rank-sum p-value < 0.001). **c,** Representative H&E stained lung sections

as described in **b**. **d,** Quantification of the lung metastases described in **Fig. 3b**, where Asns-silenced and –expressing 4T1-T cells were injected into mice. Half of the mice received L-asparaginase while the other half received an equivalent volume of PBS at the same injection rate (n=10 mice per condition, rank-sum p-value < 0.0005 for L-asparaginase vs. control for each line and for Asns-silenced vs. –unsilenced cells in each drug condition). **e,** Tumour volumes corresponding to the lung metastases described in **d** (rank-sum p-value < 0.005 for Asns-silenced vs. –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-value < 0.05 for ASNS-silenced vs. –expressing cells in both conditions and for silenced cells in treated vs. untreated mice). **g,** Tumour volumes corresponding to the mice described in **f** (rank-sum p-value < 0.05 for Asns-silenced vs. –expressing cells under both treatments and for PBS vs. L-asparaginase treated animals for each cell line). See source data.



Extended Data Figure 6: Tertiary validation that extracellular asparagine availability impacts invasion and metastasis

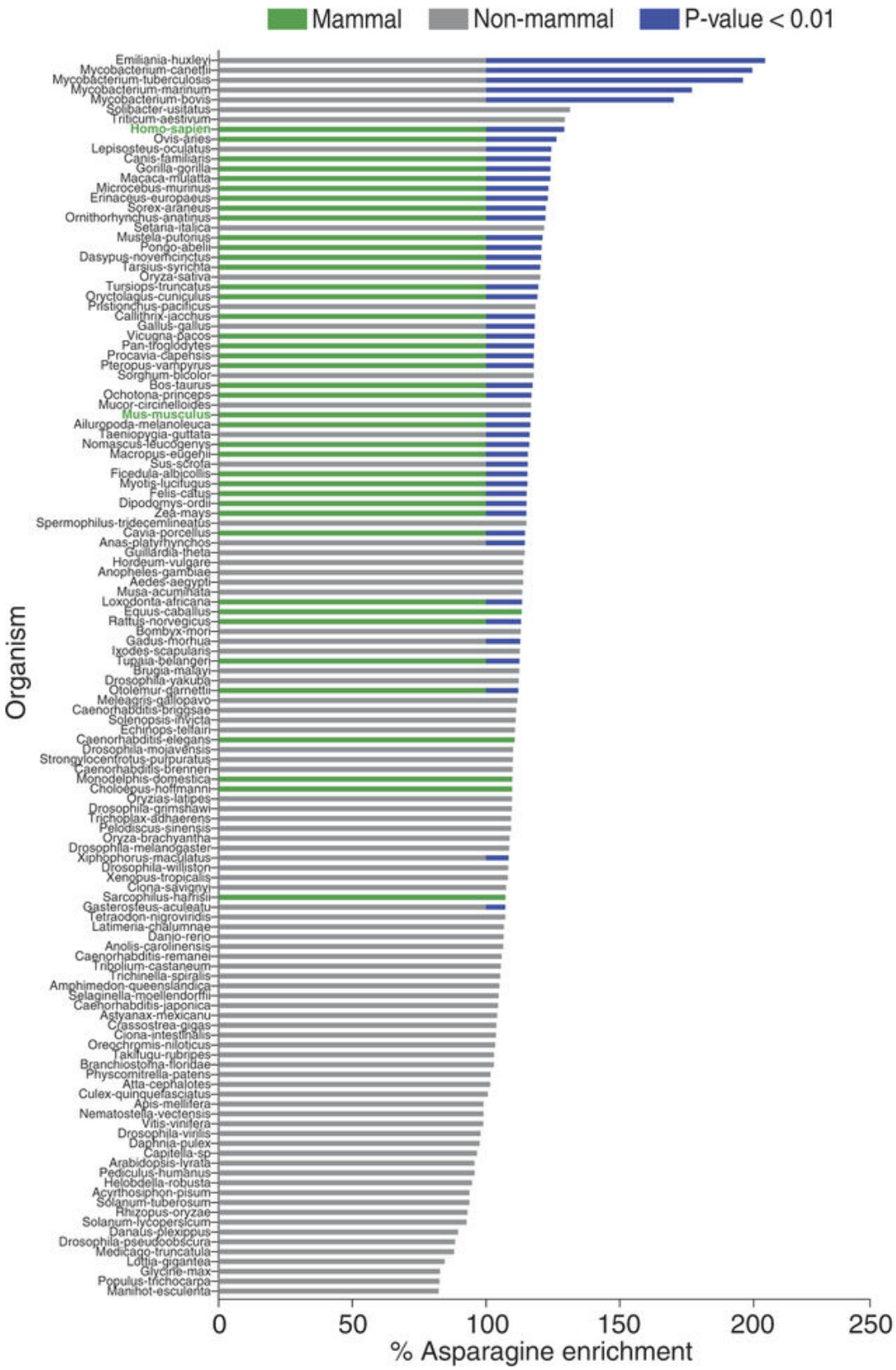
a, Asparagine content in the serum free amino acid pool, for mice that have been fed 0%, 0.6% or 4% asparagine diets (n = 5 mice per diet, edges of the box are the 25th and 75th percentiles and error bars extend to the values q3 + w(q3 − q1) and q1 − w(q3 − q1), where w is 1.5 and q1 and q3 are the 25th and 75th percentiles and this is also the case for **b**, **d**, **e** and **h**, rank-sum p-value < 0.05 between each diet). **b,** Volumes of orthotropic tumours corresponding to the lung metastases described in **Fig. 3c**, where Asns-silenced and –expressing 4T1-T cells were orthotopically 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, rank-sum p-value < 0.05 for mice receiving 0% vs. 0.6% and 0.6% vs. 4% diets). **e,** Quantification of metastases in the lungs of the animals described in **d** (rank-sum p-value < 0.05 between each diet). **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 that have been 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-value < 2.8×10^{-37} blood vs. breast and lung). See source data.



Extended Data Figure 7: Primary validation that asparagine availability regulates EMT

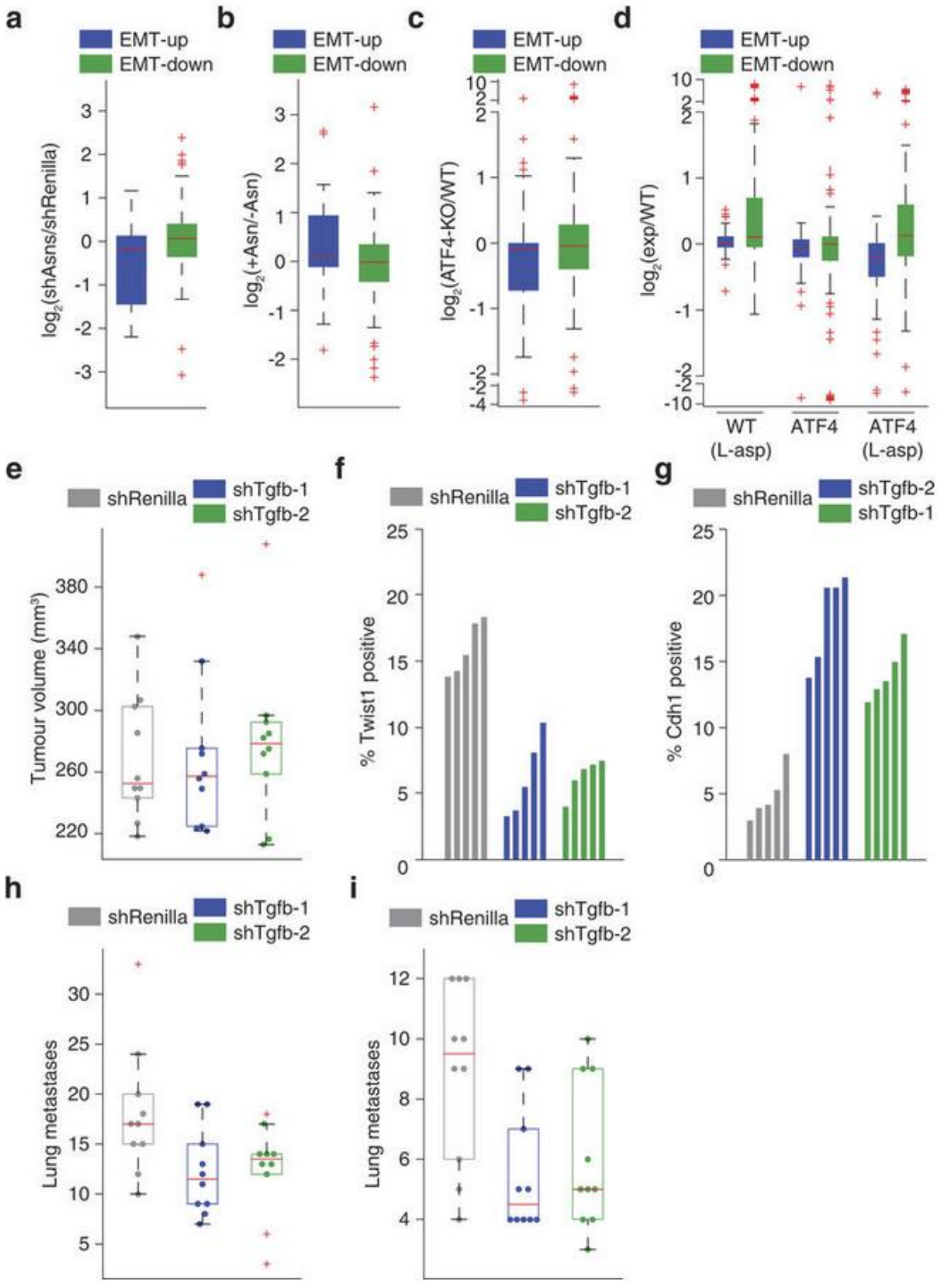
a, Protein-level changes between Asns-silenced and -expressing cells when genes are stratified by transcription-level changes (top and bottom 10% of genes based on log-fold change in Asns-silenced cells, Gene-up and -down, respectively) and asparagine content (top and bottom 10% of genes based on asparagine content, Asp-high and -low, respectively), edges of the box are the 25th and 75th percentiles and error bars extend to the values $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, where w is 1.5 and q_1 and q_3 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, while positive correlations indicate the amino acid is enriched. For protein minus RNA level expression changes, amino acids with positive correlations are enriched in proteins where depletion levels exceed what is predicted by corresponding RNA changes. Negative correlations indicate the amino acid is enriched in proteins where 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 murine and human EMT-up proteins (rank-sum p-value < 0.01 for both human and mouse). **d,** Position 15 asparagine codon enrichment in ribosome protected fragments (RPFs) from PC-3 cells grown with and without L-asparaginase, when all genes or only EMT-up genes are analyzed (outliers were not plotted to improve interpretability and this

is also the case for **e**, rank-sum p-value < 0.05 for EMT-up vs. all genes in both untreated and L-asparaginase treated cells). **e,** Increase in Asparagine codon representation in RPFs, when PC-3 cells are grown in L-Asparaginase (relative to without), when all genes or EMT-up genes are analyzed (rank-sum p-value < 0.05). See source data.


Mammal
Non-mammal
P-value < 0.01
Organism
Emiliania-huxleyi
Mycobacterium-canettii
Mycobacterium-tuberculosis
Mycobacterium-marinum
Mycobacterium-bovis
Solibacter-usitatus
Triticum-aestivum
Homo-sapien
Ovis-aries
Lepisosteus-oculatus
Canis-familiaris
Gorilla-gorilla
Macaca-mulatta
Microcebus-murinus
Erinaceus-europaeus
Sorex-araneus
Ornithorhynchus-anatinus
Setaria-italica
Mustela-putorius
Pongo-abelii
Dasypus-novemcinctus
Tarsius-syrichta
Oryza-sativa
Tursiops-truncatus
Oryctolagus-cuniculus
Pristionchus-pacificus
Callithrix-jacchus
Gallus-gallus
Vicugna-pacos
Pan-troglodytes
Procavia-capensis
Pteropus-vampyrus
Sorghum-bicolor
Bos-taurus
Ochotona-princeps
Mucor-circinelloides
Mus-musculus
Ailuropoda-melanoleuca
Taeniopygia-guttata
Nomascus-leucogenys
Macropus-eugenii
Sus-scrofa
Ficedula-albicollis
Myotis-lucifugus
Felis-catus
Dipodomys-ordii
Zea-mays
Spermophilus-tridecemlineatus
Cavia-porcellus
Anas-platyrhynchos
Guillardia-theta
Hordeum-vulgare
Anopheles-gambiae
Aedes-aegypti
Musa-acuminata
Loxodonta-africana
Equus-caballus
Rattus-norvegicus
Bombyx-mori
Gadus-morhua
Ixodes-scapularis
Tupaia-belangeri
Brugia-malayi
Drosophila-yakuba
Otolemur-garnettii
Meleagris-gallopavo
Caenorhabditis-briggsae
Solenopsis-invicta
Echinops-telfairi
Caenorhabditis-elegans
Drosophila-mojavensis
Strongylocentrotus-purpuratus
Caenorhabditis-brenneri
Monodelphis-domestica
Choloepus-hoffmanni
Oryzias-latipes
Drosophila-grimshawi
Trichoplax-adhaerens
Pelodiscus-sinensis
Oryza-brachyantha
Drosophila-melanogaster
Xiphophorus-maculatus
Drosophila-willistoni
Xenopus-tropicalis
Ciona-savignyi
Sarcophilus-harrisii
Gasterosteus-aculeatu
Tetraodon-nigroviridis
Latimeria-chalumnae
Danio-rerio
Anolis-carolinensis
Caenorhabditis-remanei
Tribolium-castaneum
Trichinella-spiralis
Amphimedon-queenslandica
Selaginella-moellendorffii
Caenorhabditis-japonica
Astyanax-mexicanu
Crassostrea-gigas
Ciona-intestinalis
Oreochromis-niloticus
Takifugu-rubripes
Branchiostoma-floridae
Physcomitrella-patens
Atta-cephalotes
Culex-quinquefasciatus
Apis-mellifera
Nematostella-vectensis
Vitis-vinifera
Drosophila-virilis
Daphnia-pulex
Capitella-sp
Arabidopsis-lyrata
Pediculus-humanus
Helobdella-robusta
Acyrthosiphon-pisum
Solanum-tuberosum
Rhizopus-oryzae
Solanum-lycopersicum
Danaus-plexippus
Drosophila-pseudoobscura
Medicago-truncatula
Lottia-gigantea
Glycine-max
Populus-trichocarpa
Manihot-esculenta
0
50
100
150
200
250
% Asparagine enrichment

Extended Data Figure 8: Conservation of asparagine enrichment in EMT promoting proteins

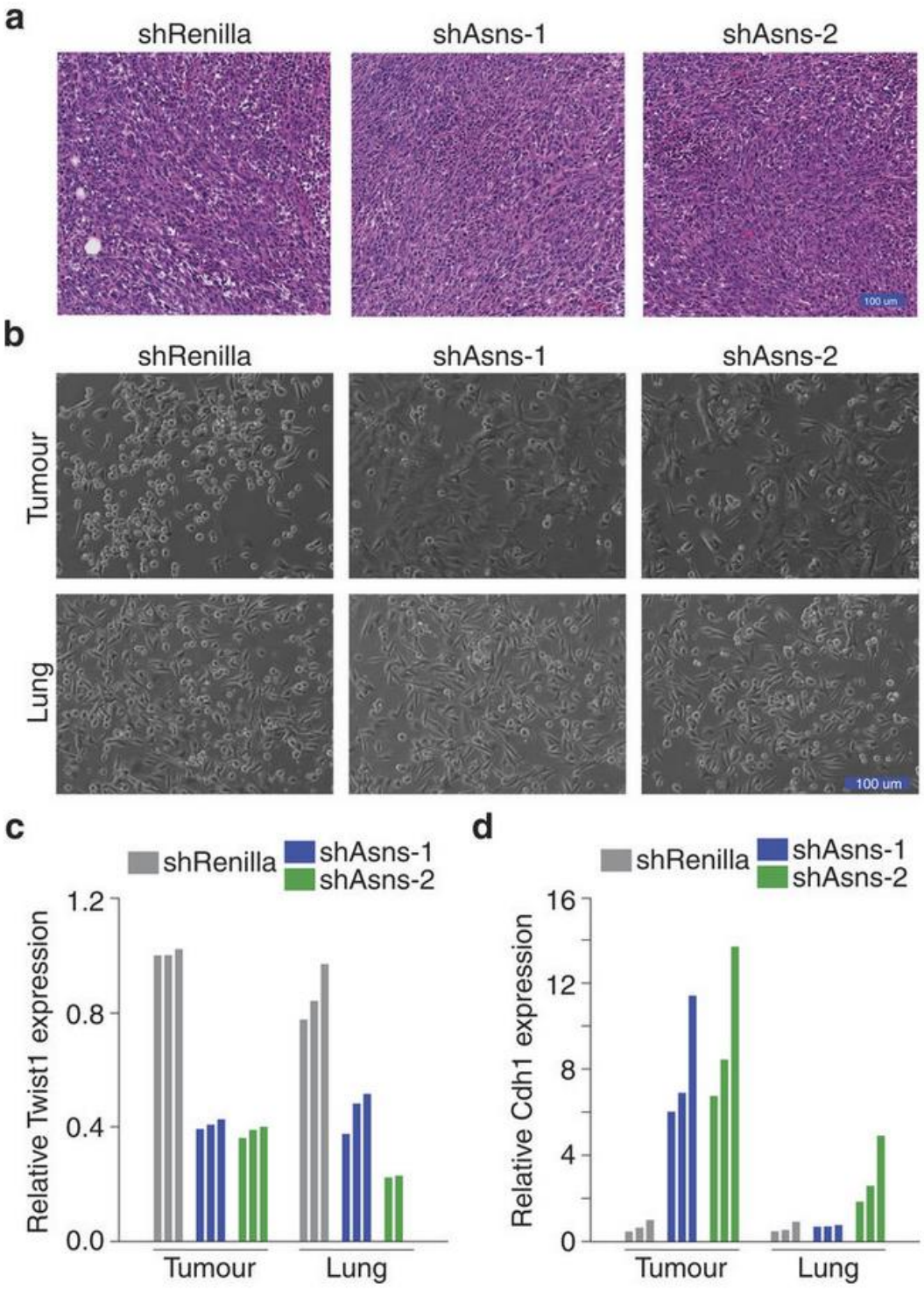
Asparagine enrichment analysis of epithelial-to-mesenchymal promoting protein orthologues in the 126 species listed in the Orthologous MAtrix database that harbour at least 10 orthologues (rank-sum p-value < 1.0x10-13 for all species and rank-sum p-value < 9.0x10-9 for mammals vs. other species).



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), where w is 1.5 and q1 and q3 are the 25th and 75th percentiles and this is

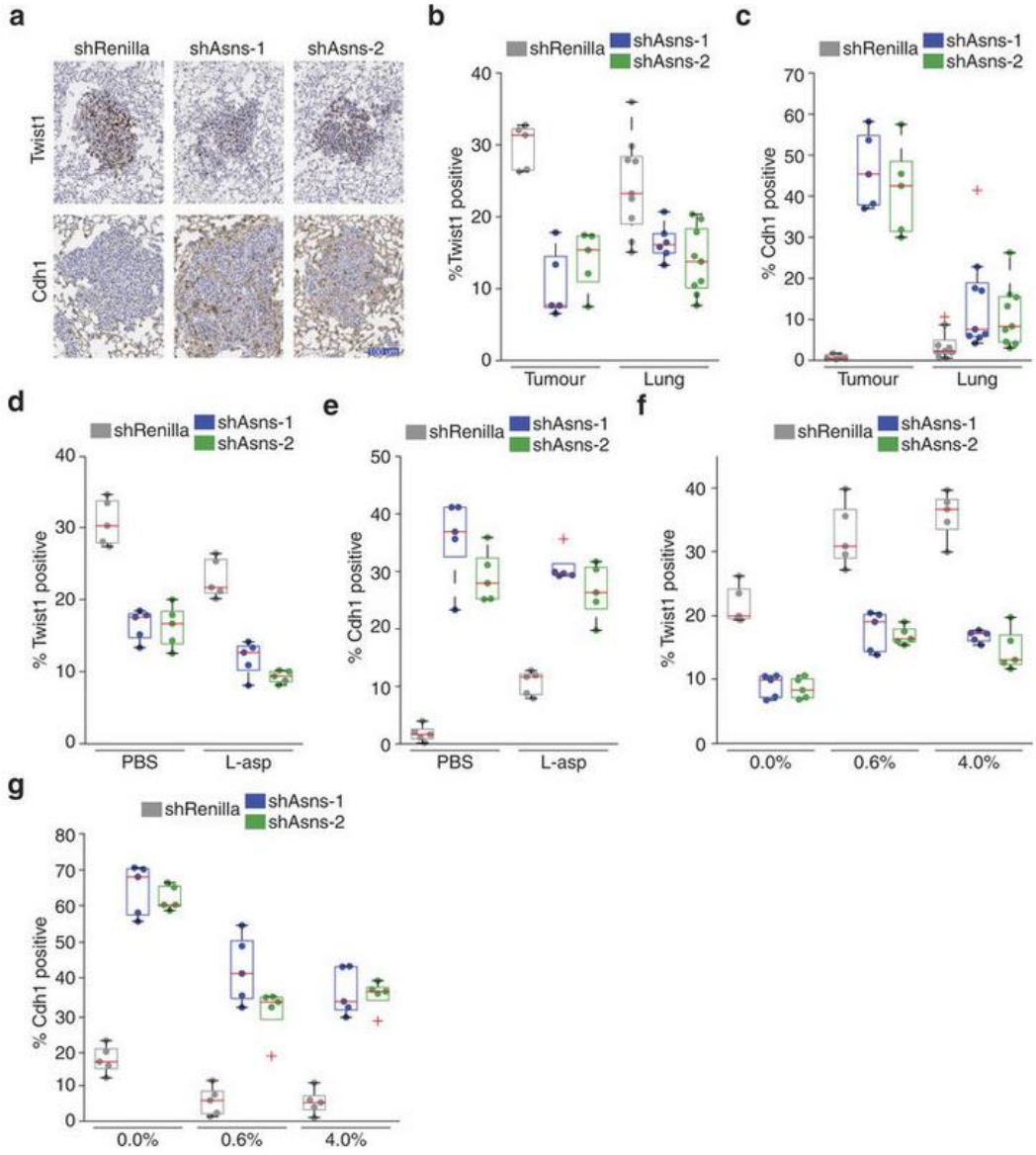
also the case for **b**, **c**, **d**, **e**, **h** and **i**, rank-sum p-value < 0.001 for EMT-up genes, DESeq FDR < 0.05 for Twist1 and Cdh1). **b,** Transcription-level changes in EMT-up and –down genes that occur in response to the media of parental 4T1 cells being supplemented with L-asparagine (n = 2 replicates per condition, rank-sum p-value < 0.005 for EMT-up genes). **c,** Gene expression changes in EMT-induced and EMT-repressed genes that result from ATF4 knockout in near haploid KBM-7 Chronic Myelogenous Leukemia (HAP-1 cells, rank-sum p-value < 0.05). **d,** Gene expression changes in EMT-induced and EMT-repressed genes, which result in the liver cells of homozygous ATF4 deleted mice when they have been treated with L-asparaginase (rank-sum p-value < 0.05 for EMT-down genes in WT+L-asp mice, and both EMT-up and –down genes in ATF4+L-asp mice). **e,** Volumes of tumours resulting from orthotopic injection of Tgf-ß-silenced and –expressing 4T1-T cells (n = 10 mice per cell line). **f,** Percent Twist1 positive regions based on IHC staining of sections from tumours described in **e** (n = 5 tumour sections per cell line, rank-sum p-value < 0.01 for Asns-silenced vs. –expressing cells). **g,** Percent Cdh1 positive regions based on IHC staining of sections from tumours described in **e** (n = 5 tumour sections per cell line, rank-sum p-value < 0.01 for Asns-silenced vs. –expressing cells). **h,** Quantification of metastases resulting from the tumours described in **e** (rank-sum p-value < 0.05). **i,** Quantification of metastases resulting from intravenous injection of Tgf-ß-silenced and –expressing cells (10 mice per cell line, rank-sum p-value < 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**, where 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 that

were 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 that were injected orthotopically with Asns-silenced 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**, where 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), where w is 1.5 and q1 and q3 are the 25th and 75th percentiles and this is also the case for **c**, **d**, **e**, **f** and **g**, rank-sum p-value < 0.01 and < 0.05 for Asns-silenced vs. –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-value < 0.01 and < 0.05 for Asns-silenced vs. –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 that are treated with PBS or L-asparaginase (n = 5 tumour sections per condition, rank-sum p-value < 0.01 for Asns-silenced vs. –unsilenced cells and rank-sum p-value < 0.05 for each cell line in treated vs. untreated mice). **e,** Quantification of Cdh1-positive regions in the tumours described in **d** (n = 5 tumour sections per condition, rank-sum p-value < 0.01 for Asns-silenced vs. –unsilenced cells and rank-sum p-value < 0.05 for each cell line in treated vs. untreated mice). **f,** Quantification of Twist1-positive regions in tumours resulting from orthotopic injection of Asns-expressing and –silenced cells into mice fed a 0%, 0.6% or 4% asparagine diet (n = 5 tumour sections per condition, rank-sum p-value < 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-value < 0.01 between Asns-silenced and –expressing cells and between diets).