

AWARD NUMBER: W81XWH-12-1-0300

TITLE: Understanding Tumor Dormancy as a Means of Secondary Prevention

PRINCIPAL INVESTIGATOR: Gregory Hannon

**CONTRACTING ORGANIZATION: Cancer Research UK
Cambridge CB2 0RE**

REPORT DATE: February 14th, 2019

TYPE OF REPORT: Final

**PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012**

DISTRIBUTION STATEMENT: Approved for public release; distribution is unlimited.

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.		
1. REPORT DATE (DD-MM-YYYY) FEB 2019	2. REPORT TYPE Final	3. DATES COVERED (From - To) 15SEP2012 - 14NOV2018
4. TITLE AND SUBTITLE Understanding Tumor Dormancy as a Means of Secondary Prevention		5a. CONTRACT NUMBER W81XWH-12-1-0300
		5b. GRANT NUMBER BC120182
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Gregory J. Hannon E-Mail: greg.hannon@cruk.cam.ac.uk		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Cancer Research UK Cambridge CB2 0RE		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT The purpose of this collaborative project was to establish a molecular definition of the dormant state of a cancer cell with the goal of understanding how this dormancy is broken, ultimately leading to recurrence in a patient that was stably in remission. Once our understanding of this is more complete, it is hoped that we can devise strategies for secondary prevention. We encountered considerable delays in this project mainly due to our move from CSHL to CRUK-		

CI which negated the use of our original strategy for the identification of candidate genes for dormancy. This led us to pursue an alternative strategy for the identification of dormant tumor cells and the characterization of their microenvironment: imaging mass cytometry or IMC. Over the past year, we have constructed antibody panels, which can be read by IMC, to identify and characterize the status of breast tumor cells and one to characterize immune infiltrates. This approach could be used as an adjunct to STPT, LCM, and RNAseq. We have also aimed to identify candidate dormancy regulators and find ways to manipulate these for patient benefit. Additionally, we have identified asparagine bioavailability as a major regulator of EMT, which could be manipulated to influence response to therapy, affecting the potential pool of residual/dormant disease, and its recognition by the innate and adaptive immune system.

15. SUBJECT TERMS

Breast cancer, Dormancy, metastasis, EMT, RNA seq, heterogeneity, ASNS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	67	19b. TELEPHONE NUMBER <i>(Include area code)</i>

TABLE OF CONTENTS

	<u>Page</u>
1. INTRODUCTION	5
2. KEYWORDS	5
3. ACCOMPLISHMENTS	5
4. IMPACT	10
5. CHANGES/PROBLEMS	12
6. PRODUCTS	13
7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS	16
8. SPECIAL REPORTING REQUIREMENTS	19
9. APPENDICES	19

1. **INTRODUCTION:** Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.

The primary purpose of our collaborative research project is to discover the genes and pathways that regulate breast cancer dormancy. Our plan was to use models of breast cancer wherein murine or human breast cancer cells establish and progress through dormancy and utilize innovative methods to perform laser capture microdissection of cancer cells in the dormant state as well as cancer cells that are progressing from dormancy. The molecular profiling of these cells at distinct stages of cancer progression, along with the profiling of various cell-types within the microenvironment that contact cancer cells was supposed to serve as the basis this project. A molecular description of the dormant state will eventually allow us to perform loss-of-function and gain-of-function studies of candidate drivers and suppressors of dormancy progression. Ultimately, we anticipate the identification of genes that could serve as attractive targets of therapeutic inhibition.

While the Tavazoie lab has been working towards human breast cancer profiling, the Hannon lab has been optimizing conditions for identifying and profiling dormant cells working towards two goals. One is the ability to locate dormant cells and the second is the ability to capture and profile, using RNA seq, single cells.

2. **KEYWORDS:**

Breast cancer, Dormancy, metastasis, EMT, RNA seq, heterogeneity, ASNS

3. **ACCOMPLISHMENTS:** The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction.

What were the major goals of the project?

List the major goals of the project as stated in the approved SOW. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The broad goals of this collaborative proposal were:

- 1) Establish a molecular definition of the dormant state
- 2) Undertake functional approaches toward understanding how dormancy is broken
- 3) Conduct pre-clinical studies of strategies for secondary prevention

What was accomplished under these goals?

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include

pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

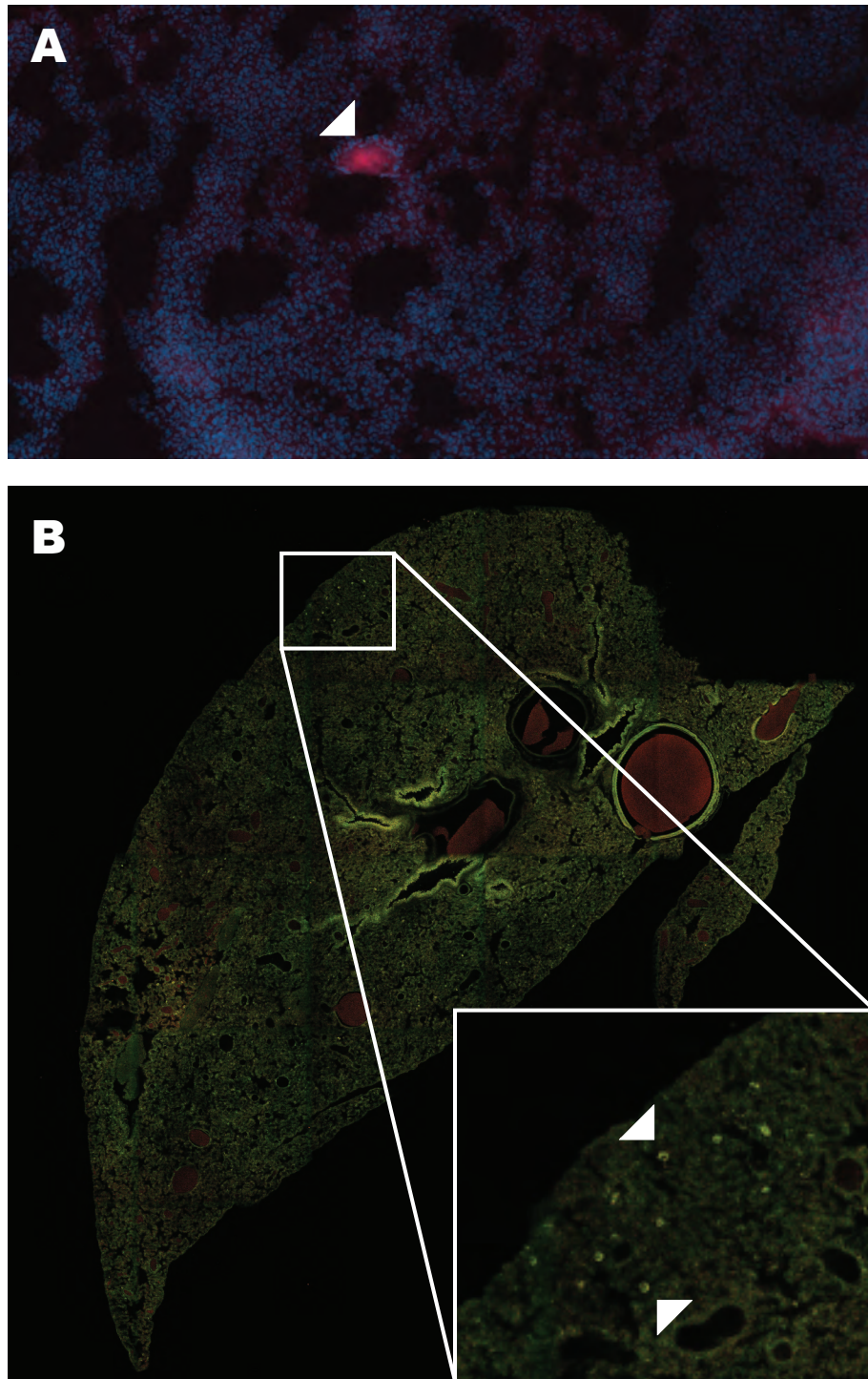
The course and duration of this project were heavily impacted by the move of the Hannon laboratory from CSHL to CRUK-CI. The main contributing factor being the extremely long delay in obtaining the STPT microscope due to the strict procurement regulations imposed by the University of Cambridge. The availability of the STPT constituted the basis for the realization of Aim 1. Given that Aim 1 was a prerequisite for Aim 2 and 3, this delay greatly affected our ability to deliver on the Aims as originally proposed. This led us looking for alternative ways to achieve the broad goals set out for this project and also to some promising outcomes as described below.

Task 1 – Broadly to identify dormant tumor cells and characterize these and their environment

We surveyed different strategies to identify dormant cells in secondary sites in order to then characterize them molecularly. We started by using label-retaining strategies (H2B-GFP, DiI dye, etc.) to identify cells that are quiescent and therefore able retain a fluorescent label for extended periods. We found that the immunogenicity of the GFP fluorophore in the H2B-GFP was causing a significant reduction in the metastatic potential of 4T1 mammary tumor cells and therefore we switched to the use of DiI. This lipophilic dye, not only allows the identification of dormant cells, but it is well tolerated by the immune system and it does not require genetic modification of the tumor cells. An example of metastasis observed in the lungs of mice injected with DiI-stained 4T1 cells is shown in Fig. 1A. Another way to identify dormant cells is to follow their behavior over time through direct visualization of tumor cells using non-invasive methods in live animals. This has become possible with the introduction of a more sensitive form of luciferase, the Akaluc, which is sensitive enough to allow the visualization of even single tumor cells via non-invasive bioluminescence imaging. We have shown that we are able to reveal metastasis in lungs 3 weeks after the orthotopic injection in the mammary fat pad of 4T1 cells containing Aka-Luc. As part of Aim 1, we have therefore established different approaches to allow the identification of dormant cells in secondary sites.

Although we were able to identify DiI stained metastasis in some cryosections as shown in Fig. 1A, our goal was to be able to use Serial Two-Photon Tomography (STPT). The STPT acquires fluorescent images of each layer of tissue, before sectioning this layer and automatically collecting the section on an indexed slide. The use of the STPT has therefore several advantages such as its speed and the fact that the images are acquired prior sectioning, therefore limiting possible artifacts. The analysis of the images acquired at the STPT would allow to distinguish dormant cells and micrometastasis using the fluorescent methods described above. Since all the sections are indexed, the slide containing the sections with cells of interest could be easily retrieved. These represent significant advantages compared to the alternative of manually sectioning all secondary sites and separately image each section to then look for dormant cells. We finally received the STPT at the end of 2017 and we spent the final year of the grant optimizing the embedding, sectioning and collection of slices from different samples, including lungs. We are now able to section and collect automatically and reliably 25 micron sections and manually 15 micron sections, which are more amenable for some of the downstream processing.

Fig.1 - Imaging of tumor cells in lungs. A) DiI stained 4T1 micrometastasis (arrowhead) in lung cyosection of a BALB/c mouse 21 days after orthotopic injection in the mammary fat pad. Nuclei are stained with DAPI. B) Single GFP-positive 4T1 cells (arrowheads) are observed after STPT sectioning in the lung of BALB/c mouse 18 days after orthotopic injection in the mammary fat pad. Inset shows higher magnification. Blood vessels were perfused with Rhodamine-dextran (red).



An example of lung tissue sectioned 18 days after orthotopic injection of GFP-positive 4T1 tumor cells in the mammary fat pad imaged and sectioned on our STPT is shown in Fig. 1B. Single GFP-positive cells are visible in a close-up of the section, reinforcing the fact that the STPT is a fast and reliable method to identify even single dormant tumor cells for downstream analysis.

The next step after the STPT would have been microdissection with LCM to produce RNA-seq libraries from dormant cells, micrometastasis, growing metastasis, as well as the microenvironment surrounding them. Comparing the RNA expression profiles of these different samples should allow us to compile a molecular definition of the dormant state and find candidate genes potentially responsible for the initiation and maintenance of dormancy and for the exit from dormancy. We already have in place in the lab all the methods to successfully isolate RNA from LCM microdissected tissue and to produce low-input libraries from it as we consistently use these methods for other projects. However, the inability to use the STPT prevented us from applying these methods to dormant cells so far. We are now completely geared to perform these experiments, regrettably when the grant period has ended. We will anyway continue to pursue this avenue, sourcing alternative funding, to be able to bring this important project to completion.

As mentioned in the last report, given the difficulties in using the STPT-LCM approach to produce molecular signatures for dormant cells, we also pursued the alternative approach of using Imaging Mass Cytometry (IMC) to characterize both the status of breast tumor cells at secondary sites, as well as the immune infiltrates interacting with them. We have already optimized and validated two different panels of antibodies that are designed to characterize either the tumor cells or the immune infiltrate. We are now geared to start using them on dormancy models or also on human tissue samples within a collaboration with Alana Welm and Cyrus Ghajar. As for the above, we will continue to pursue this project even if this award is now ended.

Tasks 2 and 3 – to identify candidate dormancy regulators and find ways to manipulate these for patient benefit.

Because we were unable to pursue the identification of candidate genes through the strategies originally set out in the proposal, we pursued alternative avenues that brought us to the identification of promising candidates. In our yearly reports, we described our use of the 4T1 model of breast cancer to identify possible drivers of metastasis in mice. We identified Asparagine Synthase as an important mediator of metastasis in this model. However, we performed our original experiments using NGS immunocompromised mice. One of the main advantages of the 4T1 model is that it is transplantable in fully immunocompetent BALB/c mice. Given that tumor microenvironment and immune system are major components influencing both metastatic potential and metastatic dormancy, we are now expanding our previous studies on 4T1 cells to examine their behavior in recipient mice with different immunocompetency. As mentioned in our last report, our preliminary results suggest that in mice with a higher degree of immunocompetency (Balb/c and Balb/c nude), 4T1 cells tend to metastasize preferentially through the lymphatic system as they are readily found in lymph nodes, whereas in NSG mice they seem to spread prevalently through the bloodstream. We are now in the process of performing a more detailed analysis of the metastatic potential of different clones in immunocompetent mice, much like we did in Wagenblast et al., Nature (2015). Our hope is to identify drivers of metastatic behavior in immunocompetent mice and to compare our finding with the immunocompromised ones to establish whether lympho-angiogenesis and lymphatic system

can play a role in mammary tumor dissemination in an immunocompetent setting. This could potentially give insights also on metastatic dormancy.

One of the main finding from our studies on 4T1 tumors was the identification of Asparagine Synthetase (Asns) as a driver of metastasis in NSG mice, strongly implicating 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 tumor types. One mechanism underlying our findings is likely a link between asparagine bioavailability and EMT, which can be observed both *in vitro* and *in vivo* (Knott et al. 2018. Nature 554:378-381). Following up on this work, we have now established a mass-spectrometry based assay to measure the ASNS activity in cells. We made good progress on this and we are now able to compare levels of ASNS in different cells. This method is sensitive and robust enough to measure changes in ASNS activity in different conditions. In the future we plan to use this method to screen for ASNS inhibitors that could possibly be used as a mean to reduce the ability of breast cancer cells to metastasize and perhaps also influence their ability to enter or exit the dormant state.

A vast literature has associated EMT with resistance to therapeutic intervention. This encompasses cytotoxic chemotherapies as well as targeted therapies. Recent studies have even suggested that cells that have undergone EMT can evade immunotherapy. For a cell to remain dormant and serve as a seed of later disease it must evade a patient's initial treatment. The work described above has led us to the hypothesis that asparagine bioavailability can influence the potential of cells colonizing secondary sites. In high asparagine conditions, cells can adopt a mesenchymal phenotype and potentially occupy the perivascular niches that have been previously proposed as a site for DTC preservation. Restriction of asparagine through any of a number of mechanisms may negate this potential in the colonizing cell but might also reverse EMT later at the dormant niche and in micrometastases, revealing these cells to the immune system or to therapeutic intervention. Thus, we feel that despite the change in direction that this work has experienced for external factors, it still represents a substantive step forward toward the overall goals of this grant.

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

This work has supported the training of a number of individuals. Several technicians have been trained as a part of this program. One of them has recently started her graduate studies. Clare Rebbeck has been promoted to Senior Research Associate during her time as part of this award. A postdoc, Ashley Nicholls, is acquiring new skills in drug discovery as we search of an ASNS inhibitor and another postdoc, Claire Mulvey, is acquiring new skills optimizing the STPT embedding and sectioning as well as the IMC.

How were the results disseminated to communities of interest?

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

Wagenblast et al (2015). Nature 520:358-62.

Knott et al. (2018). Nature 554:378-381.

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

This is not applicable, since this is the final report for this grant. Nevertheless, now that we finally have all the instruments in place to realize our original goals, we will definitely seek alternative source of funding to be able to further pursue this important project.

4. **IMPACT:** Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:

What was the impact on the development of the principal discipline(s) of the project?

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that an intelligent lay audience can understand (Scientific American style).

Nothing to report

What was the impact on other disciplines?

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Nothing to report

What was the impact on technology transfer?

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use, including:

- *transfer of results to entities in government or industry;*
- *instances where the research has led to the initiation of a start-up company; or*
- *adoption of new practices.*

Nothing to report

What was the impact on society beyond science and technology?

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

- *improving public knowledge, attitudes, skills, and abilities;*
- *changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or*
- *improving social, economic, civic, or environmental conditions.*

The observation that asparagine availability impacts metastasis could provide strategies for improving primary therapy and for reducing relapse.

5. **CHANGES/PROBLEMS:** The PD/PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:

The problems we have had acquiring the necessary equipment have impacted both the timeline and the trajectory of our work in this grant. Despite the fact that this grant unfortunately run out, we will continue after the end of the grant to study tumor dormancy. The fact that we could not proceed with the project as we originally planned, provide the impetus to attempt our alternative approach that has provided very interesting and potentially important new directions.

Actual or anticipated problems or delays and actions or plans to resolve them

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

Not applicable

Changes that had a significant impact on expenditures

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

The main contributor to changes in expenditures during the duration of the grant was the delay in the procurement of the STPT microscope. This resulted in the inability to perform the aims as originally set out and caused us to look for alternative strategies that were therefore funded through this grant.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or

equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Significant changes in use or care of human subjects

No significant changes

Significant changes in use or care of vertebrate animals

No significant changes

Significant changes in use of biohazards and/or select agents

No significant changes

6. **PRODUCTS:** *List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state “Nothing to Report.”*

- **Publications, conference papers, and presentations**
Report only the major publication(s) resulting from the work under this award.

Journal publications. List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).

Wagenblast et al (2015). Nature 520:358-62.

Knott et al. (2018). Nature 554:378-381.

Books or other non-periodical, one-time publications. Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like. Identify for each one-time publication: author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (e.g., book, thesis or dissertation); status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).

Not applicable

Other publications, conference papers and presentations. Identify any other publications, conference papers and/or presentations not reported above. Specify the status of the publication as noted above. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (*) if presentation produced a manuscript.

Not applicable

- **Website(s) or other Internet site(s)**

List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

Not applicable

- **Technologies or techniques**

Identify technologies or techniques that resulted from the research activities. Describe the technologies or techniques were shared.

Not applicable

- **Inventions, patent applications, and/or licenses**

Identify inventions, patent applications with date, and/or licenses that have resulted from the research. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.

Not applicable

- **Other Products**

Identify any other reportable outcomes that were developed under this project. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment and /or rehabilitation of a disease, injury or condition, or to improve the quality of life. Examples include:

- data or databases;

- *physical collections;*
- *audio or video products;*
- *software;*
- *models;*
- *educational aids or curricula;*
- *instruments or equipment;*
- *research material (e.g., Germplasm; cell lines, DNA probes, animal models);*
- *clinical interventions;*
- *new business creation; and*
- *other.*

Not applicable

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate “no change”.

Example:

Name: Mary Smith
Project Role: Graduate Student
Researcher Identifier (e.g. ORCID ID): 1234567
Nearest person month worked: 5

Contribution to Project: Ms. Smith has performed work in the area of combined error-control and constrained coding.

Funding Support: The Ford Foundation (Complete only if the funding support is provided from other than this award.)

Name: Clare Rebbeck
 Project Role: Senior Research Associate (Post Doc)
 Contribution to project: Project design, animal work, cell culture & imaging

Name: Elena Rozhkova
 Project Role: Post Doc
 Contribution to project: Animal work, cell culture & imaging

Name: Megan Keane
 Project Role: Research Assistant
 Contribution to project: Animal work, cell culture & imaging

Name: Ashley Nicholls
Project Role: Research Associate (Post Doc)
Contribution to project: Project design, animal work, cell culture & imaging

Name: Aarthi Gopinathan
Project Role: Research Associate (Post Doc)
Contribution to project: Animal work, cell culture & imaging

Name: Claire Mulvey
Project Role: Research Associate (Post Doc)
Contribution to project: STPT optimisation, cell culture & imaging

Name: Natasha Narayanan
Project Role: Student
Contribution to project: Animal work, cell culture & imaging

Name: Cristina Jauset Gonzalez
Project Role: Graduate Student
Contribution to project: Animal work, cell culture & imaging

Name: Sophia Wild
Project Role: Graduate Student
Contribution to project: Animal work, cell culture & imaging

Name: Tatjana Kovacevic
Project Role: Graduate Student
Contribution to project: Animal work, cell culture & imaging

Name: Fiona Nugent
Project Role: Research Assistant (technician)
Contribution to project: Animal work, cell culture & imaging

Name: Jose' Franco Alvarez
Project Role: Research Assistant (technician)
Contribution to project: Animal work, cell culture & imaging

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

If the active support has changed for the PD/PI(s) or senior/key personnel, then describe what the change has been. Changes may occur, for example, if a previously active grant has closed and/or if a previously pending grant is now active. Annotate this information so it is clear what has changed from the previous submission. Submission of other support information is not necessary for pending changes or for changes in the level of effort for active support reported previously. The awarding agency may require prior written approval if a change in active other support significantly impacts the effort on the project that is the subject of the project report.

Not applicable

What other organizations were involved as partners?

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe partner organizations – academic institutions, other nonprofits, industrial or commercial firms, state or local governments, schools or school systems, or other organizations (foreign or domestic) – that were involved with the project. Partner organizations may have provided financial or in-kind support, supplied facilities or equipment, collaborated in the research, exchanged personnel, or otherwise contributed.

Provide the following information for each partnership:

Organization Name:

Location of Organization: (if foreign location list country)

Partner’s contribution to the project (identify one or more)

- *Financial support;*
- *In-kind support (e.g., partner makes software, computers, equipment, etc., available to project staff);*
- *Facilities (e.g., project staff use the partner’s facilities for project activities);*
- *Collaboration (e.g., partner’s staff work with project staff on the project);*
- *Personnel exchanges (e.g., project staff and/or partner’s staff use each other’s facilities, work at each other’s site); and*
- *Other.*

Tavazoie lab, Rockefeller University.

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: *For collaborative awards, independent reports are required from BOTH the Initiating Principal Investigator (PI) and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to <https://ers.amedd.army.mil> for each unique award.*

QUAD CHARTS: *If applicable, the Quad Chart (available on <https://www.usamraa.army.mil>) should be updated and submitted with attachments.*

9. **APPENDICES:** *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.*

A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis

Elvin Wagenblast¹, Mar Soto¹, Sara Gutiérrez-Ángel¹, Christina A. Hartl¹, Annika L. Gable¹, Ashley R. Maceli¹, Nicolas Erard^{1,2}, Alissa M. Williams¹, Sun Y. Kim¹, Steffen Dickopf¹, J. Chuck Harrell³, Andrew D. Smith⁴, Charles M. Perou³, John E. Wilkinson⁵, Gregory J. Hannon^{1,2} & Simon R. V. Knott^{1,2}

Cancer metastasis requires that primary tumour cells evolve the capacity to intravasate into the lymphatic system or vasculature, and extravasate into and colonize secondary sites¹. Others have demonstrated that individual cells within complex populations show heterogeneity in their capacity to form secondary lesions^{2–5}. Here we develop a polyclonal mouse model of breast tumour heterogeneity, and show that distinct clones within a mixed population display specialization, for example, dominating the primary tumour, contributing to metastatic populations, or showing tropism for entering the lymphatic or vasculature systems. We correlate these stable properties to distinct gene expression profiles. Those clones that efficiently enter the vasculature express two secreted proteins, Serpine2 and Slpi, which were necessary and sufficient to program these cells for vascular mimicry. Our data indicate that these proteins not only drive the formation of extravascular networks but also ensure their perfusion by acting as anticoagulants. We propose that vascular mimicry drives the ability of some breast tumour cells to contribute to distant metastases while simultaneously satisfying a critical need of the primary tumour to be fed by the vasculature. Enforced expression of SERPINE2 and SLPI in human breast cancer cell lines also programmed them for vascular mimicry, and SERPINE2 and SLPI were overexpressed preferentially in human patients that had lung-metastatic relapse. Thus, these two secreted proteins, and the phenotype they promote, may be broadly relevant as drivers of metastatic progression in human cancer.

Until now, the most detailed studies of tumour heterogeneity have been retrospective^{6–9}. For example, single cell analyses of human breast tumours have illustrated evolutionary paths of genetic diversification⁹. In such cases, genetic variation could not be associated with differences in the behaviour and capabilities of clonal populations and their specific contributions to disease. We therefore wished to complement such studies by creating an experimental model of tumour heterogeneity.

To this end, we marked individual mouse mammary carcinoma 4T1 cells with a molecular barcode via retroviral infection (Fig. 1a and Extended Data Fig. 1a). We drew from a complex mixture five different cohorts of 100,000 cells each, and introduced these orthotopically into immunocompromised recipients (NOD-SCID-*Il2rg*^{−/−} (NSG) mice). After 24 days, primary tumours, brachial lymph nodes, blood, lungs, livers and brains were collected, and the barcode populations within each tissue were quantified (Fig. 1b and Extended Data Fig. 1f). We asked whether the subset of clones that engrafted in all samples (~1,400 clones) showed consistent behaviour in terms of contributions to aspects of disease progression across all five experiments.

Two conclusions were drawn from this analysis. First, clone abundance within the primary tumour did not correlate with abundance in circulating tumour cells (CTCs) or secondary lesions. Second, distinct

groups of clones contributed to lymph node and blood-borne metastases. Significant overlap existed between abundant clones in the blood-borne metastases and CTCs (Fig. 1b and Extended Data Fig. 1g, $P < 0.001$, hypergeometric test). However, no significant overlap was observed when comparing these sets to the prominent clones in the lymph node (Fig. 1b and Extended Data Fig. 1h). Indeed, others have reported that 20–30% of patients with distant relapse are free of axillary lymph-node metastases¹⁰. Thus, clonal populations within the 4T1 cell line reproducibly contribute to different aspects of disease progression.

We wished to understand the properties of these clones, which underlay their differential capabilities. We therefore established 23 clonal lines from another barcoded population (Fig. 2a and Extended Data Figs 1b and 2). All relevant barcode integration sites were mapped to ensure that no known oncogenes or tumour suppressors were altered during the barcoding process (Supplementary Information). The clones spanned a range of *in vitro* growth rates and cellular morphologies. After minimal propagation, they were pooled and orthotopically injected into NSG mice. In addition, the pool was propagated on adherent culture plates. Primary tumours and aliquots from the *in vitro* system were removed after 14 and

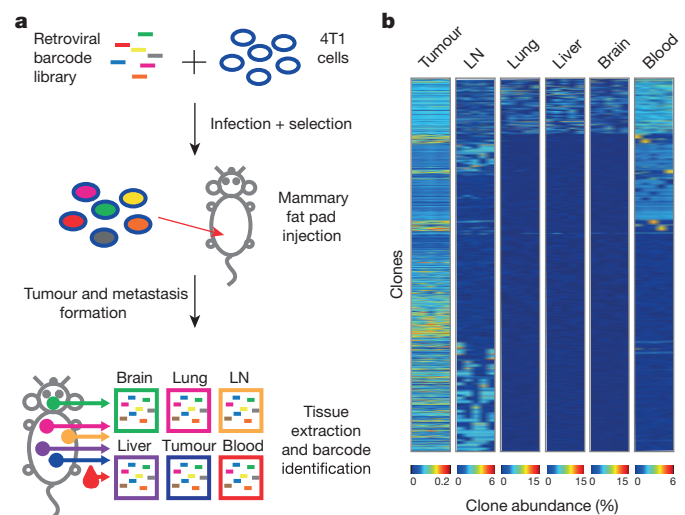


Figure 1 | Clonal analysis of 4T1 transplantation by molecular barcoding. **a**, Retroviral barcoding strategy for identifying clonal populations within the 4T1 cell line. **b**, Relative proportions of clones that engrafted in all animals in the lymph node (LN), lung, liver, brain and blood. Columns represent independent experiments ($n = 5$ mice). Shown are the ~1,400 clones that successfully engrafted in all animals.

¹Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ²CRUK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ³Department of Genetics and Pathology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁴Molecular and Computational Biology, University of Southern California, Los Angeles, California 90089, USA. ⁵Department of Pathology, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA.

24 days. In addition, at 24 days, the brachial lymph nodes, blood, lungs, livers and brains were isolated.

At 14 days, the clonal profiles of the *in vitro* samples and the primary tumours were found to be highly similar (Fig. 2b). However, at 24 days, while the *in vitro* population maintained its distribution, the primary tumour evolved along a different trajectory with clone 4T1-I dominating. Even when engrafted individually, 4T1-I showed accelerated growth between the 14- and 24-day time points, indicating that this phenotype is not dependent on clonal interactions (Extended Data Fig. 3a).

Examination of metastatic sites and CTCs showed that different clones had different capacities to contribute, and this did not correlate with their abundance in the primary tumour (Fig. 2c). Clones that were relatively less represented in the primary tumour entered the bloodstream and survived as CTCs, and a subset of these had the additional ability to colonize secondary sites. The latter clones differ still from those that colonized lymph nodes. The 4T1-T clone that dominates sites colonized by blood-borne routes was also best at forming lung metastases when injected individually (Extended Data Fig. 3b). Intravasation seemed a key gating step since intracardiac injection of the pool led to an entirely different clonal distribution in CTCs and lung metastases (Extended Data Fig. 3c).

The proclivities of each clonal line were general properties of most of their constituent cells. This was demonstrated by infection of lines 4T1-L, 4T1-E and 4T1-T with secondary, independent barcode libraries containing mCherry, allowing for populations of cells within each line to be monitored at each stage of disease (Fig. 2d and Extended Data Fig. 3d). Each secondarily barcoded line was separately pooled with the remaining lines and injected orthotopically. Similar numbers of subclones were identified for each clonal line within the tumours,

indicating that they engraft at comparable rates. A large proportion of the engrafted 4T1-E and 4T1-T subclones were able to contribute to the CTC population. Furthermore, many 4T1-T CTCs were able to extravasate and colonize the lung. The ability of the 4T1-T clone to form lung metastases was confirmed by mCherry staining of lung tissue sections from these experiments (Fig. 2e). Finally, these properties appear to be stable as they remained after the clones had been propagated for more than 20 doublings (Extended Data Fig. 3e).

To ascertain the specific drivers of this phenotype, we intersected the set of genes significantly overexpressed in clones 4T1-E and 4T1-T with a set that was found to be upregulated in lung metastases relative to matched primary tumours (Fig. 3a). Expression levels of the resultant 12 candidates were additionally examined in human patients, comparing those that did or did not relapse with lung metastases¹¹. Of the 10 genes with associated patient data, the human orthologues of *Serpine2* and *Slpi* (*SERPINE2* and *SLPI*, respectively) emerged as the most significantly overexpressed in relapsed patients. The 4T1 cell line is used to model aggressive breast cancer subtypes such as basal, Her2 and claudin-low. Notably, it is precisely these tumour types, and not luminal cancers, that show increased *SERPINE2* and *SLPI* expression in patients that relapse (Extended Data Fig. 4a, $P < 0.005$, Wilcoxon rank-sum test for summed expression). Additionally, a hazard analysis determined that amplified expression of these genes was significantly associated with relapse in the lung ($P < 0.0002$, Supplementary Information). Both *SERPINE2* and *SLPI* have been implicated in breast cancer progression, however, *Slpi* has also been proposed to suppress tumour growth^{12–16}.

To validate *Serpine2* and *Slpi* as drivers of metastatic progression, 4T1-T populations were individually infected with two short hairpin RNAs (shRNAs) targeting each gene or with a control shRNA

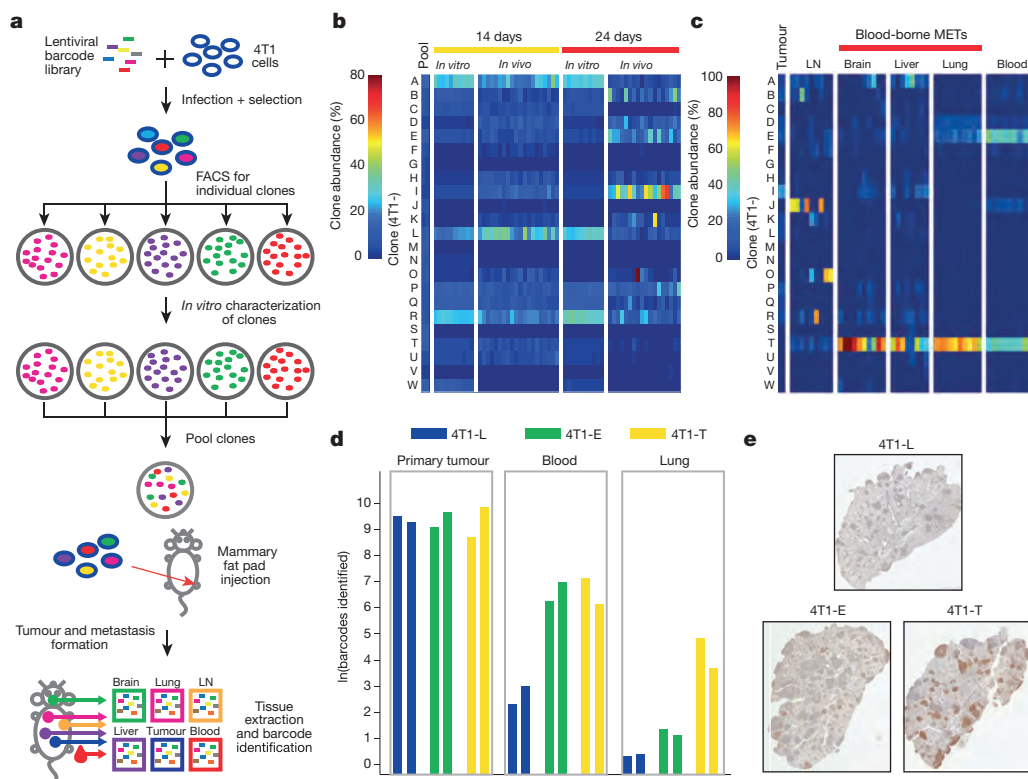


Figure 2 | Focused analysis of a subset of 4T1 clones throughout metastatic disease progression. **a**, Strategy for isolating individual molecularly barcoded 4T1 clones. **b**, Relative proportions of the clonal lines *in vitro* and in orthotopic primary tumours at 14 and 24 days (*in vitro* $n = 10$ cell lines, *in vivo* $n = 18$ mice). **c**, Relative proportions of clonal lines among the vascular CTCs and secondary lesions in lymph node, brain, liver and lungs in animals corresponding to the tumours extracted at 24 days in **b** ($n = 8$ mice). **d**, Subclonal

analysis of 4T1-L, 4T1-E and 4T1-T cells via secondary barcode library infection. Clonal lines were separately infected with a second barcode library, pooled with the 22 other lines and then orthotopically injected into two mice. Indicated is the number of clones identified in 100 bootstrap samples of 1 million sequence reads. **e**, Immunohistochemistry analysis for mCherry in lung metastases resulting from each of the three pooled injections discussed in **d**. mCherry is expressed in the secondary barcode library. Original magnification, $\times 20$.

(Extended Data Fig. 1c and Supplementary Information). After orthotopic injection of a pooled collection of the resultant lines, a significant depletion of the *Serpine2* and *Slpi* shRNA expressing cells in CTCs and lung metastases was observed (Extended Data Fig. 4b, $P < 0.01$, Wilcoxon rank-sum). When each shRNA-expressing line was assessed separately in a polyclonal setting, *Serpine2*- and *Slpi*-depleted 4T1-T clones were significantly decreased in their contribution to CTCs and lung metastases (Fig. 3b, $P < 0.02$, Wilcoxon rank-sum). Also, a reduction in shRNA-expressing lung nodules was observed after silencing of either gene (Extended Data Fig. 4c, d, $P < 0.01$, Wilcoxon rank-sum). Finally, when *Serpine2* and *Slpi* were silenced in parental 4T1 cells, singly or in combination (Extended Data Fig. 1c, d), significant reductions in lung metastases were observed. Lungs corresponding to single gene knockdowns showed a ~50% reduction in metastases and those corresponding to double knockdowns showed a ~60% reduction in secondary lesions (Extended Data Fig. 4e, f, $P < 0.01$ and $P < 0.005$, respectively, Wilcoxon rank-sum).

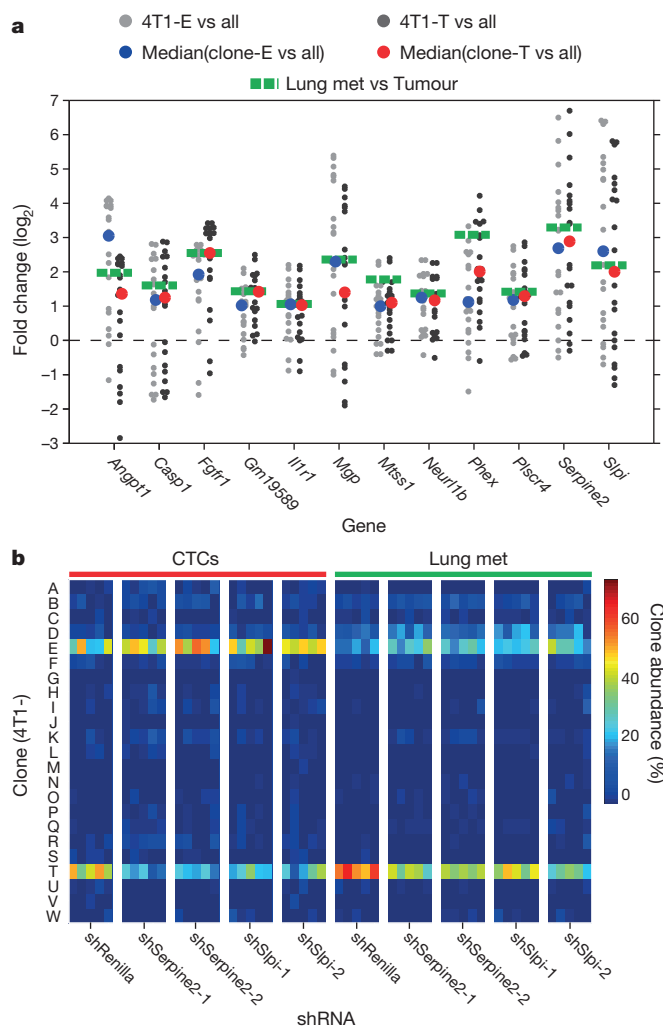


Figure 3 | Serpine2 and Slpi are regulators of intravasation into the cardiovascular system. **a**, Genes upregulated in the intravasating clones 4T1-E and 4T1-T. RNA sequencing (RNA-seq) was performed for all clonal lines ($n = 2$ per cell line) as well as for two pairs of matched primary tumours and lung metastases. Grey dots represent the fold change of each gene in 4T1-E (light grey) or 4T1-T (dark grey) relative to each of the other clonal lines (the median values are plotted as blue and red dots for 4T1-E and 4T1-T, respectively). The green dotted lines represent the mean fold change in the tumour and lung metastases (met) pairs. **b**, Relative proportions of clonal lines in the CTCs and lung where 4T1-T has been infected with non-targeting shRNAs and shRNAs targeting *Serpine2* and *Slpi* ($P < 0.02$, Wilcoxon rank-sum, $n = 5$ mice).

Thus, *Serpine2* and *Slpi* probably act at the intravasation step, a hypothesis supported by the finding that intracardiac injection rescued the metastatic potential of shRNA-expressing cells (not shown). Notably, a recent study also implicated Serpins, including SERPINE2, in metastasis of breast cancer to the brain¹⁷. In this study, knockdown of *SERPINE2* did not block metastasis; however, cells were introduced by intracardiac injection, bypassing the requirement for increased SERPINE2 and SLPI expression according to our model.

In tumours derived from 4T1-E or 4T1-T cell lines, we observed an increase in vessels with focal loss of the endothelial cell marker CD31 (not shown). Quantitative analysis of vascular leakiness revealed that vessels within 4T1-T tumours were significantly more leaky than those of parental 4T1 and 4T1-L tumours (Extended Data Fig. 5a, b, $P < 0.05$, Wilcoxon rank-sum). Finally, silencing of *Serpine2* or *Slpi* in these cells resulted in reduced vascular leakiness (Extended Data Fig. 5c, $P < 0.03$, Wilcoxon rank-sum).

Focal loss of CD31 in vessels and a high degree of tumour leakiness has been associated with a phenomenon termed vascular mimicry, in which tumour cells differentiate into endothelial-like cells and form extracellular-matrix-rich tubular structures to carry blood from the vasculature to hypoxic regions of the tumour^{18–20}. We proposed that the propensity of 4T1-E and 4T1-T cells to intravasate was the result of a heightened capacity for vascular mimicry, placing tumour cells in direct contact with blood.

Analysis of serial sections of tumours derived from mCherry expressing 4T1-T cells revealed vast networks of structures consistent with vascular mimicry ($PAS^+/CD31^-$ fluid-filled channels lined by mCherry⁺ tumour cells; Fig. 4a, b and Extended Data Fig. 6a). An equivalent number of $PAS^+/CD31^-$ channels were identified in 4T1-E-derived tumours. By contrast, a much lower number of such structures were identified in tumours derived from other clones (Fig. 4b). When mCherry-expressing 4T1-T cells were pooled with the remaining clones and injected orthotopically, most tumour cells surrounding the resulting $PAS^+/CD31^-$ vessels were mCherry-positive (Extended Data Fig. 6b).

Clones 4T1-E and 4T1-T were robust in their capacity to form tubular structures when grown on matrigel, a characteristic consistent with increased vascular mimicry in tumour cells (Extended Data Fig. 7a). In addition, *Serpine2*- and *Slpi*-depleted 4T1-T cells displayed fewer $PAS^+/CD31^-$ channels in their derived tumours, and formed significantly fewer tubular structures *in vitro* (Fig. 4c and Extended Data Fig. 7b, $P < 0.02$ and $P < 0.0002$, Wilcoxon rank-sum).

Enforced expression of *Serpine2* and *Slpi* in parental 4T1 cells or non-intravasating clones 4T1-B, -F, -N and -S increased *in vitro* formation of tubular structures (with the exception of 4T1-B overexpression of *Slpi*, Extended Data Fig. 1e, 7c, $P < 0.03$, Wilcoxon rank-sum). For each clonal line (with the exception of 4T1-F), this was accompanied by an enhanced capacity to contribute to CTCs in the context of a heterogeneous clone mixture (for this experiment lacking 4T1-E, Extended Data Fig. 7d, $P < 0.05$, Wilcoxon rank-sum). Tumours derived from the parental 4T1 cells overexpressing *Serpine2* and *Slpi* had significantly more $PAS^+/CD31^-$ channels than cells infected with an empty vector, and this was accompanied by an increase in lung metastases (Extended Data Fig. 7e–g, $P < 0.01$, Wilcoxon rank-sum, and $P < 0.05$, Friedman).

Serpine2 and *Slpi* also enabled vascular mimicry in human basal and/or claudin-low breast cancer lines MDA-MB-231 and -436 *in vitro* (Extended Data Fig. 8a, b, $P < 0.05$, Friedman) and *in vivo* (Fig. 4d and Extended Data Fig. 8c, $P < 0.0002$ and $P < 0.002$, respectively, Wilcoxon rank-sum). This was accompanied by an increase in lung metastatic burden (Extended Data Fig. 8d, e, $P < 0.05$, Wilcoxon rank-sum). A barcode analysis of these lines (as described for Fig. 2d) determined that this resulted from an expansion in the number of clones that were able to form CTCs (Fig. 4e and Extended Data Fig. 8f, $P < 0.04$ and $P < 0.02$, respectively, Wilcoxon rank-sum).

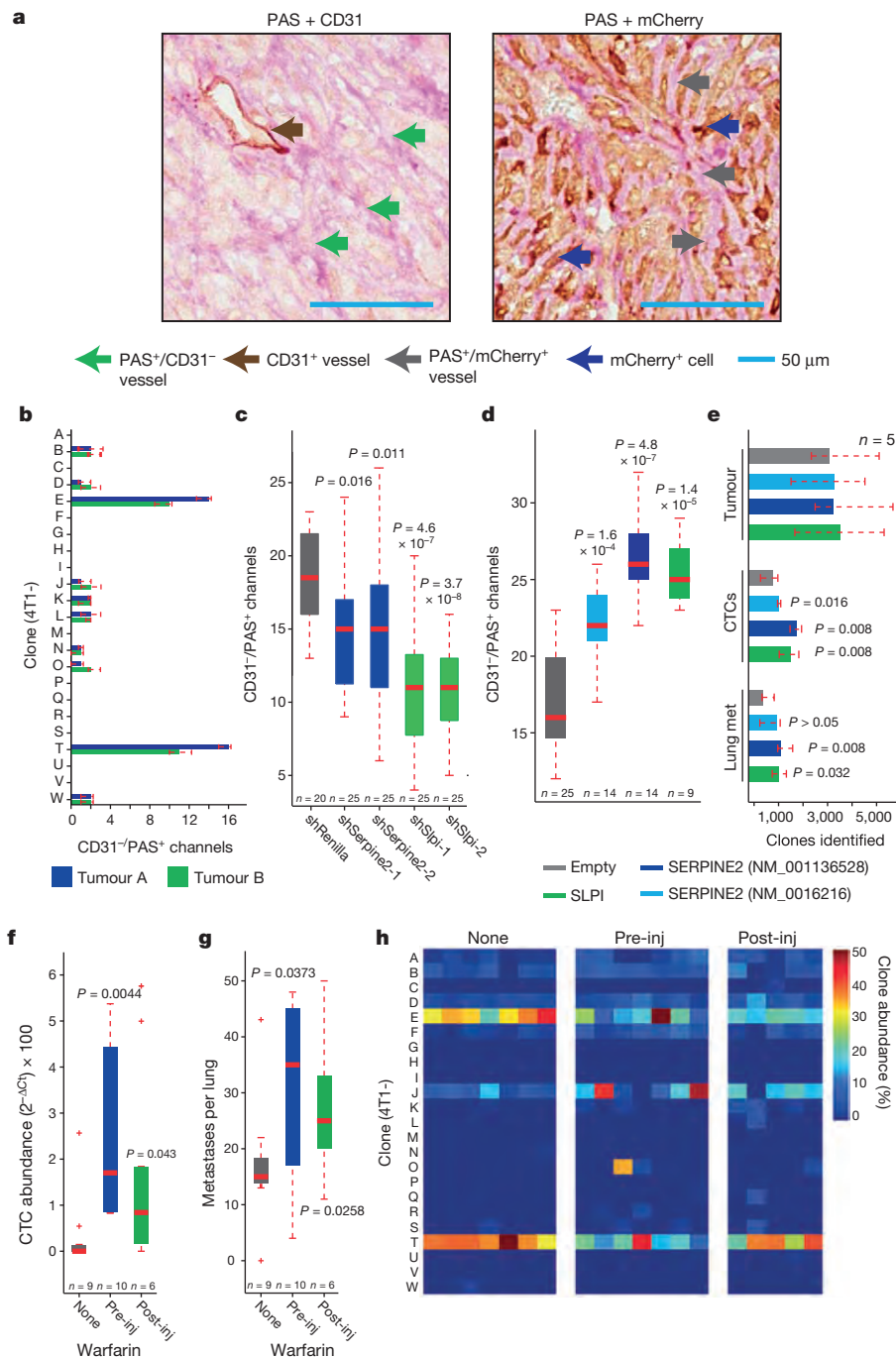


Figure 4 | Vascular mimicry drives metastatic progression. **a**, Serial sections of a primary tumour derived from 4T1-T cells constitutively expressing mCherry stained with PAS and either CD31 or mCherry. Brown arrow indicates CD31-positive blood vessels, green arrows CD31-negative/PAS-positive channels, blue arrows show mCherry-positive tumour cells, and grey arrows show PAS⁺ channels lined by mCherry-positive tumour cells. Scale bars, 50 μm. **b**, Quantification of CD31⁺/PAS⁺ channels in tumours derived from each clonal line (median of $n = 5$ fields plotted). **c**, Quantification of CD31⁺/PAS⁺ channels in 4T1-T-derived primary tumours that have been infected with a non-targeting shRNA and shRNAs targeting *Serpine2* or *Slpi* ($P < 0.02$, Wilcoxon rank-sum, $n = 20$ fields for shRenilla and $n = 25$ fields for shSerpine2-1, shSerpine2-2, shSlpi-1 and shSlpi-2). **d**, Quantification of CD31⁺/PAS⁺ channels in MDA-MB-231-derived primary tumours cells that have been infected with an empty vector or vectors for overexpression of SERPINE2 or SLPI ($P < 0.0002$, Wilcoxon rank-sum, $n = 25$ fields for empty, $n = 14$ fields for SERPINE2 (NM_001136528 and NM_0016216) and $n = 9$ fields for SLPI). **e**, Sub-clonal analysis of MDA-MB-231 cells that have been infected with an empty vector or vectors for overexpression of SERPINE2 or

SLPI ($P < 0.04$, Wilcoxon rank-sum, $n = 5$ mice). **f**, Cardiovascular CTC abundance (measured by qPCR of the barcode vector) in animals injected orthotopically with all 23 clonal lines and administered regular drinking water or water containing 10 mg ml⁻¹ warfarin ($P < 0.05$, Wilcoxon rank-sum, $n = 9$, 10 and 6 mice for mice administered no warfarin, warfarin pre-injection and warfarin post-injection, respectively). **g**, Numbers of lung metastatic nodules identified in the animals described in **f** ($P < 0.04$, Wilcoxon rank-sum, $n = 9$, 10 and 6 mice for mice administered no warfarin, warfarin pre-injection and warfarin post-injection, respectively). **h**, Relative proportions of each clone in the cardiovascular CTCs of animals that were orthotopically injected with the 23-clone pool and administered regular drinking water or water containing 10 mg ml⁻¹ warfarin (either pre- or post-injection, $P < 0.002$, Wilcoxon rank-sum, $n = 7$, 7 and 5 mice for mice administered no warfarin, warfarin pre-injection and warfarin post-injection, respectively). For all box plots, the edges of the box are the twenty-fifth and seventy-fifth percentiles. The 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 twenty-fifth and seventy-fifth percentiles, respectively.

Both Serpine2 and Slpi are anticoagulants and previous studies have reported that such factors are amplified in tumours with pronounced vascular mimicry²¹. Anticoagulants could have a role in maintaining flow in extravascular channels by preventing clotting at the vascular–extravascular interface. To test whether anticoagulants promote intravasation and metastasis, we orthotopically injected the pool of clones into warfarin-treated mice. These animals had significantly reduced levels of cleaved prothrombin factors fragments 1 and 2 (F1 and F2) in the blood, and increased vascular leakiness in their tumours (Extended Data Fig. 9a–c, $P < 0.01$ and $P < 0.000005$, Wilcoxon rank-sum). However, no significant change in PAS⁺/CD31⁺ vessels was observed (not shown). While tumour volumes remained stable, quantitative PCR (qPCR) for the barcode vector in whole blood revealed a significant escalation in the number of CTCs (Fig. 4f, $P < 0.05$, Wilcoxon rank-sum). This was also accompanied by an increase in lung metastatic burden (Fig. 4g, $P < 0.04$, Wilcoxon rank-sum). By contrast, when cells were injected intracardially, a significant decrease in metastases was observed, indicating that anticoagulants strongly promote intravasation (Extended Data Fig. 9d, $P < 0.05$, Wilcoxon rank-sum).

If the anticoagulant activity of Serpine2 and Slpi is important for intravasation, one would expect cells overexpressing these proteins to have less of a competitive advantage if coagulation were reduced globally within the tumour. We therefore clonally profiled the tumours, blood and lung metastases from warfarin-treated animals. Although no difference was observed between the primary tumours, clones 4T1-E and 4T1-T were reduced relatively in the CTCs and lung metastases of the warfarin-treated groups (Fig. 4h and Extended Data Fig. 9e, f, $P < 0.002$ and $P < 0.003$, Wilcoxon rank-sum). Those clones with relatively increased abundance in CTCs of warfarin-treated animals all showed some native capacity to form extravascular networks *in vivo*. These results hint that the anticoagulant action of Serpine2 and Slpi promotes extravascular network perfusion and consequently intravasation.

We have described a mouse model of breast tumour heterogeneity, which allowed us to probe the molecular basis of stable differences in the ability of clonal populations to contribute to various aspects of the disease. In this model, the ability to form CTCs, and ultimately metastases, is closely linked to the capacity for vascular mimicry. Tumour cell lined vasculature has shown a strong clinical correlation with advanced stage disease and poor clinical outcome²². In our model, vascular mimicry is driven by increased expression of two secreted proteins, Serpine2 and Slpi. Very little is currently known of the molecular determinants that enable vascular mimicry, and, to our knowledge, Serpine2 and Slpi are among the first validated drivers of this process. These proteins drive the formation of tubules *in vitro* and extravascular networks *in vivo*. In addition, we have shown that they probably have an additional role, by acting as anticoagulants at the vascular/extravascular interface to maintain perfusion of the tumour lined networks. Together, these properties are likely to promote the passage of red blood cells into the tumour and cancer cells into the bloodstream. Thus, our findings reveal a process that links fulfilment of the needs of the primary tumour with metastatic progression.

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.

Received 19 September 2014; accepted 11 March 2015.

Published online 8 April 2015.

1. Vanharanta, S. & Massague, J. Origins of metastatic traits. *Cancer Cell* **24**, 410–421 (2013).

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. Miller, B. E., Miller, F. R., Wilburn, D. & Heppner, G. H. Dominance of a tumor subpopulation line in mixed heterogeneous mouse mammary tumors. *Cancer Res.* **48**, 5747–5753 (1988).
4. Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895 (1977).
5. Fidler, I. J. Tumor heterogeneity and the biology of cancer invasion and metastasis. *Cancer Res.* **38**, 2651–2660 (1978).
6. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
7. Lohr, J. G. *et al.* Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell* **25**, 91–101 (2014).
8. Mullighan, C. G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**, 1377–1380 (2008).
9. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
10. Braun, S. *et al.* Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer. *N. Engl. J. Med.* **342**, 525–533 (2000).
11. 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).
12. Amiano, N. O. *et al.* Anti-tumor effect of SLPI on mammary but not colon tumor growth. *J. Cell. Physiol.* **228**, 469–475 (2013).
13. Fayard, B. *et al.* The serine protease inhibitor protease nexin-1 controls mammary cancer metastasis through LRP-1-mediated MMP-9 expression. *Cancer Res.* **69**, 5690–5698 (2009).
14. Martin, K. J., Patrick, D. R., Bissell, M. J. & Fournier, M. V. Prognostic breast cancer signature identified from 3D culture model accurately predicts clinical outcome across independent datasets. *PLoS ONE* **3**, e2994 (2008).
15. Rosso, M. *et al.* Secretory Leukocyte Protease Inhibitor (SLPI) expression downregulates E-cadherin, induces beta-catenin re-localisation and triggers apoptosis-related events in breast cancer cells. *Biol. Cell* **106**, 308–322 (2014).
16. Sayers, K. T., Brooks, A. D., Sayers, T. J. & Chertov, O. Increased secretory leukocyte protease inhibitor (SLPI) production by highly metastatic mouse breast cancer cells. *PLoS ONE* **9**, e104223 (2014).
17. Valiente, M. *et al.* Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* **156**, 1002–1016 (2014).
18. Maniotis, A. J. *et al.* Vascular channel formation by human melanoma cells *in vivo* and *in vitro*: vasculogenic mimicry. *Am. J. Pathol.* **155**, 739–752 (1999).
19. Hendrix, M. J., Sefter, E. A., Hess, A. R. & Sefter, R. E. Vasculogenic mimicry and tumour-cell plasticity: lessons from melanoma. *Nature Rev. Cancer* **3**, 411–421 (2003).
20. Folberg, R., Hendrix, M. J. & Maniotis, A. J. Vasculogenic mimicry and tumor angiogenesis. *Am. J. Pathol.* **156**, 361–381 (2000).
21. Ruf, W. *et al.* Differential role of tissue factor pathway inhibitors 1 and 2 in melanoma vasculogenic mimicry. *Cancer Res.* **63**, 5381–5389 (2003).
22. Cao, Z. *et al.* Tumour vasculogenic mimicry is associated with poor prognosis of human cancer patients: a systemic review and meta-analysis. *Eur. J. Cancer* **49**, 3914–3923 (2013).

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

Acknowledgements This work was supported by the Howard Hughes Medical Institute as well as grants from the NIH (G.J.H.). 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. Coblenz, L. Bianco for support with mouse work; J. Ratcliff and P. Moody for assistance with flow cytometry; D. Hoppe, A. Nourjanova, R. Puzis for histology support and S. Hearn for microscopy assistance. We thank E. Hodges and E. Lee for support with next-generation sequencing; K. Chang for the lentiviral barcode library; E. Mardis and C. Sawyers for comments on the manuscript. E.W. is a Starr Centennial Scholar and is supported by a fellowship from the Boehringer Ingelheim Fonds. J.C.H. and C.M.P. were supported by funds from the NCI Breast SPORE program (P50-CA58223-09A1), the Breast Cancer Research Foundation and the Triple Negative Breast Cancer Foundation. S.R.V.K. is supported by a fellowship from The Hope Funds For Cancer Research.

Author Contributions E.W., G.J.H. and S.R.V.K. designed the experiments. E.W., M.S., S.G., C.A.H., A.L.G., A.R.M., N.E., A.M.W., S.Y.K., S.D. and S.R.V.K. performed experiments and analysed data. J.C.H. and C.M.P. assisted with human expression data. A.D.S. helped with analysis. J.E.W. performed histological analysis. E.W., G.J.H. and S.R.V.K. wrote the paper. G.J.H. and S.R.V.K. supervised the research.

Author Information All raw and processed data is available through the Gene Expression Omnibus (GEO) under the accession number GSE63180. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.edu).

METHODS

Cell culture. The mouse mammary tumour cell line 4T1 (ATCC) and any derived clonal cell lines were cultured in DMEM high glucose (Life Technologies) supplemented with 5% fetal bovine serum (FBS; Thermo Scientific), 5% fetal calf serum (FCS; Thermo Scientific), non-essential amino acids (Life Technologies) and penicillin-streptomycin (Life Technologies). Human breast tumour cell lines MDA-MB-231 and MDA-MB-436 (ATCC) were cultured in DMEM high glucose supplemented with 10% FBS, non-essential amino acids and penicillin-streptomycin. Human umbilical vein endothelial cells (HUVECs) (Lonza) were cultured in EBM-2 media with the EGM-2 Bulletkit (Lonza). HUVECs were used within three passages. Platinum-A (Cell BioLabs) and 239-FT (Life Technologies) packaging cell lines were cultured in DMEM high glucose supplemented with 10% FBS and penicillin-streptomycin.

Virus production. All retroviral vectors were packaged using platinum-A packaging cells. The lentiviral barcode library was packaged using 293-FT lentivirus packaging cells. Cells were plated on 15 cm adherent tissue culture plates (Corning) ~5 h before transfection at a confluency of ~70%. A transfection mixture was prepared with viral vector (75 µg), VSV-G (7.5 µg), 2 M calcium chloride (187.5 µl) (Sigma-Aldrich) and, when transfecting shRNA-containing vectors, 20 nM siRNAs targeting Pasha (200 µl) (Qiagen). The mixture was brought to 1.5 ml with H₂O and then added drop-wise to the same amount of 2× HBS while being bubbled. One litre of 2× HBS was prepared with 280 mM NaCl, 50 mM HEPES, 1.5 mM Na₂HPO₄, 12 mM dextrose and 10 mM KCl (Sigma-Aldrich), then adjusted to a pH of 7.02. After the transfection mixture was added to the HBS, vigorous bubbling continued for 30–60 s. After letting the resultant mixture stand for 15 min, it was added to the packaging cells along with 100 mM chloroquin (7.5 µl) (Sigma-Aldrich). After 14 h, media was replaced. Thirty hours after media change, virus was collected and filtered through a 0.45-µm filter (EMD Millipore) and stored at 4 °C.

Establishment of clonal cell lines. Around 30 million 4T1 cells were infected with the lentiviral barcode library (Extended Data Fig. 1b) at a multiplicity of infection (MOI) of 0.3. Single cells were sorted using the FACSaria IIU cell sorter (BD Biosciences) into 96-well plates. Clonal cell lines were minimally expanded and frozen down. The barcode of each individual clonal cell line was determined by Sanger sequencing. Forward primer: 5'-CAGAATCGTTGCCTGCACATCTTGGAAAC-3' and reverse primer: 5'-ATCCAGAGTTGATTGTTCCAGACGCGT-3'.

Clonal cell line proliferation rates. Proliferation assays were performed by counting viable cells over 72 h. In total, 1×10^5 cells were plated in duplicates and were counted using the MACSQuant Analyzer (Miltenyi Biotec).

Chromosomal integration site. Genomic DNA from each clone was isolated using the QIAamp DNA Blood Mini Kit (Qiagen). Chromosomal integration sites were determined using the Lentiviral Integration Site Analysis Kit (Clontech).

Pooling experiments. For clonal pooling experiments, clonal cell lines were counted in duplicates using the MACSQuant Analyzer (Miltenyi Biotec). Equal numbers of cells were pooled together for injection. A pre-injection pool was collected to validate equal representation of each clone before injection. Tumour, lung, brain, liver and brachial lymph node were collected from mice for further processing. Blood was collected through cardiac perfusion with PBS and 0.5 M EDTA, pH 8.0, was added as an anticoagulant.

Animal studies. All mouse experiments were approved by the Cold Spring Harbour Animal Care and Use Committee. Female 6–7-week-old NOD-SCID-*Il2rg*^{-/-} (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ, NSG) were purchased from JAX. All orthotopic injections were performed using 1×10^5 mouse mammary tumour cells re-suspended in 20 µl of a 1:1 mix of PBS and growth-factor-reduced Matrigel (BD Biosciences). For human breast cancer cells MDA-MB-231 and MDA-MB-436, 2×10^6 cells were re-suspended in 50 µl of a 1:1 mix of PBS and Matrigel. Injections were done into mammary gland 4. For intracardiac injections, 1×10^5 mouse mammary tumour cells were re-suspended in 200 µl of PBS and injected into the left cardiac ventricle. For tail-vein injections, 5×10^5 mouse mammary tumour cells were re-suspended in 100 µl PBS.

Primary tumour volume was measured using the formula $V = 1/2(L \times W^2)$, in which L is length and W is width of the primary tumour. Warfarin (10 mg l⁻¹, Sigma-Aldrich) was administered with drinking water and changed every 3 days. For all animal studies where a P value was to be reported, a minimum of five animals per condition were used. This allows for standard non-parametric tests (for example, Wilcoxon rank-sum) to detect strong effects. Animals were assigned to treatment groups through random cage selection. Animals that succumbed to tumour cell injections were excluded from analysis. No statistical methods were used to predetermine sample size.

Quantification of lung metastatic burden. The lung metastatic burden of individual clones and MDA-MB-231 cells injected into the mammary gland was evaluated in five-micrometre sections stained with a standard haematoxylin and eosin protocol. Quantification was performed using ImageJ Software (NIH) converting images to 8-bit. Upper and lower thresholds for each image were adjusted to determine total lung area and adjusted again to determine the metastatic area.

Both values were used to obtain relative metastatic areas. For all other experiments, the lung metastatic burden was evaluated by counting the number of metastatic nodules in the lung. For this, five-micrometre sections were stained with a standard haematoxylin and eosin protocol.

Barcode and shRNA analysis. Genomic DNA was isolated using phenol chloroform extraction for all tissues except blood. Genomic DNA for blood was isolated using the QIAamp DNA Blood Mini Kit (Qiagen).

The barcodes of the retroviral library (Extended Data Fig. 1a) were amplified using a one-step PCR protocol. For each sample, 96 individual PCR reactions of 200 ng of genomic DNA were carried out using KOD Polymerase (EMD Millipore). Forward primer: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and reverse primer: 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3'. The reverse primer contained a barcode (NNNNNN) that enabled multiplexing with standard Illumina Truseq chemistry and software. The PCR was carried out for 30 cycles and PCR products were purified using the PCR purification kit (Qiagen). PCR products were size selected on an E-gel SizeSelect 2% agarose gel (Life Technologies), and sequenced on the Illumina HiSeq sequencer generating 22-nucleotide single-end (SE) reads.

The barcodes of the lentiviral library (Extended Data Fig. 1b) were amplified using a two-step PCR protocol. For each sample, eight individual PCR reactions of 200 ng of genomic DNA were carried out using KOD Polymerase (EMD Millipore). Forward primer 1: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGAATCGTTGCCTGCACATCTTGGAAAC-3' and reverse primer 1: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCCAGAGTTGATTGT-TCCAGACGCGT-3'. The first PCR was carried out for 25 cycles. PCR products were purified using the PCR purification kit (Qiagen). The second PCR was performed using 500 ng of PCR product from the first PCR. Forward primer 2: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3 and reverse primer 2: 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3'. The reverse primer contained a barcode (NNNNNN) that enabled multiplexing with standard Illumina Truseq chemistry and software. The second PCR was carried out for 25 cycles and PCR products were again purified using the PCR purification kit (Qiagen). PCR products were size selected on an E-gel SizeSelect 2% agarose gel (Life Technologies), and sequenced on the Illumina HiSeq sequencer generating 22-nucleotide single-end (SE) reads.

For Figs 1, 2d, 4e and Extended Data Fig. 8f, the vector library was sequenced at high depth. For each experiment the corresponding fastq file was aligned to the vector library with the Bowtie software, allowing three mismatches. Each experimental read was then assigned to the most abundant vector sequence that it mapped to. For Fig. 1 only sequences that were present with a count greater than or equal to five in all tumours were analysed. For Fig. 2d, the error bars represent the number of clones that were identified when 100 bootstrapped samples of 1 million reads each were processed for each of the two tumours, blood and lung samples. For Fig. 4e and Extended Data Fig. 8f, bootstrapping was also performed as described above and each sample (tumour, blood sample or pair of lungs) was assigned the median of the clones identified in the corresponding random samplings.

For Extended Data Fig. 1g, h, a mixed Gaussian model was fitted to the summed distributions described in Extended Data Fig. 1f. Abundant clones were identified as those that then subsequently clustered into the Gaussian with the larger mean.

The shRNAs were amplified using the same two-step PCR protocol as described above for the lentiviral barcode library. Forward primer 1: 5'-CAGAATCGTTGCCTGCACATCTTGGAAAC-3' and reverse primer 1: 5'-CTGCT-AAAGCGCATGCTCCAGACTGC-3'. Forward primer 2: 5'-AATGATACGGCGACCACCGAGATCTACACTAGCCTGCGCAGTAGTGAAGCCACAG-ATGTA-3' and reverse primer 2: 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGCTA-AAGCGCATGCTCCAGACTGC-3'. The reverse primer contained a barcode (NNNNNN) that enabled multiplexing.

NGS libraries that failed to produce more than 5,000 sequences were excluded from any further analysis as this was taken as evidence of poor quality.

Barcode quantification. Barcode libraries were de-convoluted using the Bowtie software allowing three mismatches²³. Barcode counts were then quantified from the resultant .sam file with a simple shell script containing the unix commands, cut, sort and uniq-c.

Isolation of matched tumour and lung metastatic cells. Tumour and lung tissue were harvested from mice injected with the pool of 23 clonal cell lines. Tissue was minced and treated in DMEM high glucose containing 1× collagenase/hyaluronidase buffer (StemCell) and 10 U DNase I (Sigma) for 1 h at 37 °C. Cells were washed in HBSS (Life Technologies) twice and then re-suspended in 4T1 cell culture media containing 60 µM 6-thioguanine. Cells were passaged for 5 days until all stromal cells died.

RNA-seq library preparation. Total RNA was purified and DNase treated using the Qiagen RNeasy Mini Kit. RNA integrity (RNA Integrity score >9) and quantity was measured on an Agilent Bioanalyzer (RNA Nano kit). The NuGEN Ovation RNA-Seq V2 protocol was carried out on 100 ng of total RNA. cDNA was fragmented using the Covaris LE220 sonicator according to the manufacturer's instruction to yield a target fragment size of 200 bp. The fragmented cDNA was subsequently processed using the NuGEN Ovation Ultralow DR Multiplex System. Two technical replicates were used per sample.

Analysis of RNA-seq data. Each sample was sequenced on the Illumina HiSeq sequencer generating 76-nucleotide single-end reads. Reads were aligned to the mm10 genome using the Bowtie-2 alignment tool under default parameters²⁴. Mapped reads were then assigned to genes using HTSeq-count (using the latest version of RefSeq.gtf file for gene coordinates)²⁵. Resultant counts were then normalized and compared using DESeq²⁶. For a gene to be considered overexpressed it had to show an at least twofold change with a false discovery rate (FDR) < 0.05.

Analysis of clinical data. A matrix where rows correspond to genes and columns to patients was quantile normalized to ensure that each patient profile had an equivalent empirical distribution. For the analysis of SERPINE2 and SLPI in relapsed and non-relapsed patients the across-patient profiles were z-score normalized and then the values of the relapsed patients compared to the non-relapse for each tumour subtype and site of relapse combination.

shRNA knockdown and cDNA overexpression. Mouse and human cell lines were transduced with amphotropically packaged retroviruses (Extended Data Fig. 1c–e). For shRNA knockdown studies, 4T1-T cells and parental 4T1 cells were selected with 500 $\mu\text{g ml}^{-1}$ hygromycin for 1 week.

For overexpression studies, all mouse clonal cells lines were selected with 1,000 $\mu\text{g ml}^{-1}$ G418 for 1 week, the parental 4T1 cell line was selected with 600 $\mu\text{g ml}^{-1}$ G418 for 1 week. MDA-MBA-231 cells were selected with 1,500 $\mu\text{g ml}^{-1}$ G418 and MDA-MD-436 cells were selected with 1,000 $\mu\text{g ml}^{-1}$ G418 for 1 week. shSerpine2-1: 5'-TGCTGTTGACAGTGAGCGACAGGTCTTCAATCAGATCATATAGTGAAGCCACAGATGTATATGATCTGATTGAAGACCTGGTGCCTACTGCCTCGGA-3'. shSerpine2-2: 5'-TGCTGTTGACAGTGAGCGACAGGTCTTCAATCAGATGTATATGATCTGATTGAAGACCTGGTGCCTACTGCCTCGGA-3'. shSlpi1-1: 5'-TGCTGTTGACAGTGAGCGATGCGTGAATCCTGTTCCCATATAGTGAAGCCACAGATGTATATGGAACAGGATTCACGCACTGCCTACTGCCTCGGA-3'. shSlpi2-2: 5'-TGCTGTTGACAGTGAGCGATCAGGCAAGATGTATGATGCTTGTGAAGCCACAGATGTAAGCATCATACATCTTGCCTGAGTGCC-TACTGCCTCGGA-3'.

Barcode complexity studies. Parental 4T1 cells were infected with the retroviral barcode library and were selected with 500 $\mu\text{g ml}^{-1}$ hygromycin for 1 week. 4T1-E, 4T1-L and 4T1-T cells were infected with the retroviral barcode library. 4T1-E and 4T1-L cells were selected with 1,000 $\mu\text{g ml}^{-1}$ hygromycin for 1 week. 4T1-T cells were selected with 500 $\mu\text{g ml}^{-1}$ hygromycin for 1 week.

After infection and selection of MDA-MB-231 and MDA-MB-436 cells with overexpression constructs, cells were infected with the lentiviral barcode library and selected with 1 $\mu\text{g ml}^{-1}$ puromycin for 5 days.

qRT-PCR. Total RNA was purified and DNase treated using the RNeasy Mini Kit (Qiagen). Synthesis of cDNA was performed using SuperScript III Reverse Transcriptase (Sigma). Quantitative PCR analysis was performed on the Eppendorf Mastercycler ep realplex. All signals were quantified using the ΔC_t method and were normalized to the levels of *Gapdh*.

qRT-PCR primers. Mouse *Slpi* (exon 1–2): 5'-GACTGTGGAAGGAGGCAAA-3', 5'-GGCATTTGGCTTCTCAAG-3'. Mouse *Slpi* (exon 3–4): 5'-CAGTGTGACGGCAATACAAG-3', 5'-GCCAATGTCAGGGATCAGG-3'. Mouse *Serpine2* (exon 3–4): 5'-TCTGCCTCTGAGTCCATCA-3', 5'-AACCAGAC-TTCCACAAACC-3'. Mouse *Serpine2* (exon 5–6): 5'-TCATCCCTCACATCACTACCA-3', 5'-CTTTCAGTGGCTCCTTCAGAT-3'. Mouse *Gapdh* (exon 2–3): 5'-AATGGTGAAGGTGGTGTG-3', 5'-GTGGAGTCACTAGTGAACATGTAG-3'. Human *SLPI* (exon 1–2): 5'-TGTGGAAGGCTCTGGAAAG-3', 5'-TGGCACTCAGGTTCTGTATC-3'. Human *SERPINE2* (exon 5–6): 5'-GCCATGGTGATGAGATACGG-3', 5'-GCACTTCAATTTAGAGGCAT-3'. Human *GAPDH* (exon 2–3): 5'-ACATCGCTCAGACACCATG-3', 5'-TGATGTTGAGTCAATGAAGGG-3'.

mCherry analysis. For immunohistochemistry, five-micrometre sections of paraffin-embedded lungs or primary tumours were deparaffinized in xylene, rehydrated in an alcohol series and immersed in distilled water. The sections were treated with high-temperature antigen retrieval in citrate buffer (pH 6). The slides were then blocked with 2.5% ready-to-use normal horse serum from ImmPRESS Anti-Rabbit Ig (peroxidase) Polymer Detection Kit (MP-7401, Vector Laboratories) for 1 h and then incubated with primary antibody RFP Antibody Pre-adsorbed (1:200) (600–401–379, Rockland) overnight at 4 °C. After washing, the slides were incubated with secondary antibody from the previous kit

for 30 min, rinsed and developed with chromogen ImmPACT DAB Peroxidase Substrate for staining (SK-4105, Vector Laboratories) until the desired intensity was achieved. Slides were counterstained with haematoxylin and coverslipped. Sections were then scanned on the Aperio Light Field Slide Scanner (Aperio) for further quantification. Omission of the primary antibody was used as a negative control in both cases. All quantification was performed in a blinded setting.

Vascular leakage. To visualize vascular leakage in the primary tumour, 100 μl of dextran Alexa 647, 10 kDa (1 mg ml⁻¹ in PBS) (Life Technologies) were injected into mice by tail vein injection. Three minutes later, mice were perfused with 4% paraformaldehyde (PFA). After fixation, tumours were collected and placed in 4% PFA overnight at 4 °C. After this, samples were infiltrated with 20% sucrose overnight at 4 °C. Tumours were frozen in OCT compound (Sakura Finetek) and 25- μm thick sections were cut, washed, incubated with DAPI (1 mg ml⁻¹) (Sigma-Aldrich) and mounted in ProLong Gold antifade reagent (Life Technologies). Sections were examined under the LSM 780 Confocal microscope (Zeiss).

An average of 5–7 fields were taken from each sample. Images were quantified using ImageJ software (NIH). For quantifying fluorescence, the threshold of each picture was adjusted to the lowest possible value in the DAPI channel to measure total tissue area. The dextran threshold was fixed in each picture at a determined value based on the average intensity of all samples processed. The dextran-positive area was then normalized to the total tissue area in order to calculate the leakiness index. All quantification was performed in a blinded setting.

CD31 analysis. Four-micrometre sections of paraffin-embedded primary tumours were de-paraffinized in xylene, rehydrated in an alcohol series and immersed in distilled water. The sections were then treated with high-temperature antigen retrieval in citrate buffer (pH 6), blocked with 2.5% ready-to-use normal horse serum from ImmPRESS Anti-Rabbit Ig (peroxidase) Polymer Detection Kit (MP-7401, Vector Laboratories) for 1 h and incubated with primary antibody against CD31 (1:400) (28364, Abcam) overnight at 4 °C. After washing, the slides were incubated with secondary antibody from the previous kit for 30 min, rinsed, and developed with chromogen ImmPACT DAB Peroxidase Substrate for staining (SK-4105, Vector Laboratories) until the desired intensity was achieved. Slides were then stained with Periodic Acid-Schiff (PAS) Kit (Sigma) according to manufacturer's instructions. Sections were then scanned on the Aperio Light Field Slide Scanner (Aperio) for further analysis. Omission of the primary antibody was used as a negative control.

Vascular mimicry. PAS staining, haematoxylin and eosin staining, and CD31 immunohistochemistry were used to evaluate the presence and extent of mimicry as previously described^{18–20}. Five random $\times 40$ fields per tumour were scored for the number and size of areas with morphology consistent with mimicry. The criteria used was (1) PAS positive channels that contain red cells and fluid, (2) the absence of CD31 staining in these channels, and (3) the polarization of tumour cells on an indistinct or imperceptible matrix lining vascular channels with red cells and/or fluid and no evidence of endothelialization or tumour cells lining vascular spaces with no evidence of a matrix. All quantification was performed in a blinded setting.

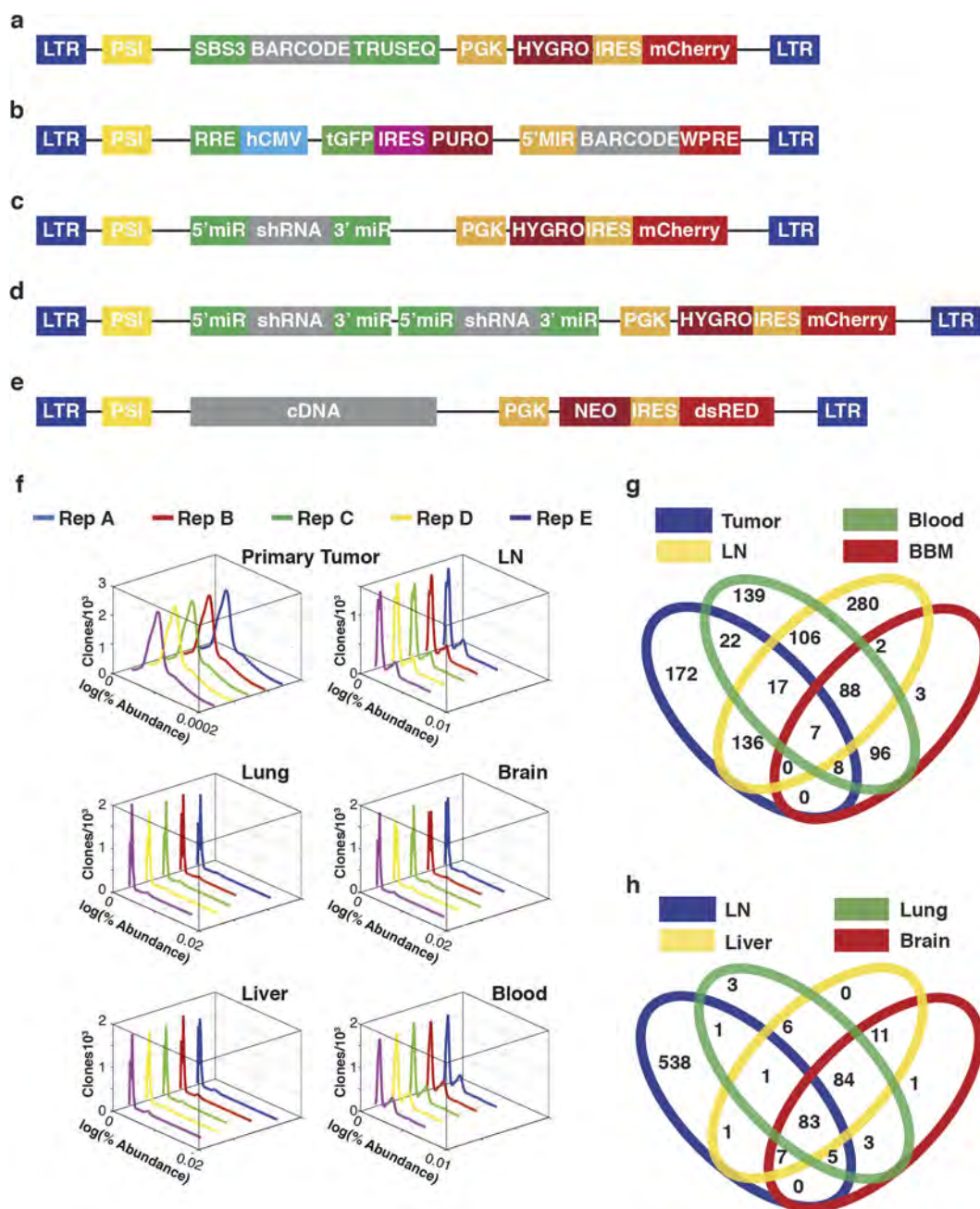
Tube formation assay. The 96-well plates were coated with 50 μl of growth-factor-reduced Matrigel (BD Biosciences) and 5×10^3 cells were re-suspended in EBM-2 media and plated in each well. All cells were plated in four replicates. Morphological studies were performed after 8 h using the Zeiss Axio Observer inverted microscope. For Extended Data Fig. 7a, the morphological analysis was performed after overnight incubation.

Prothrombin fragment 1+2 ELISA. Blood was collected from animals that were treated with warfarin by cardiac heart puncture using 3.8% sodium citrate as an anticoagulant. Samples were centrifuged at 1,000g for 15 min at 4 °C. Blood plasma was isolated and stored at -80 °C for further processing. Plasma samples were analysed for prothrombin fragment 1+2 using the Mouse Prothrombin Fragment 1+2 (F1+2) ELISA kit (Kamiya Biomedical Company) according to manufacturer's instructions.

qPCR for circulating tumour cells. Genomic DNA for blood was isolated using the QIAamp DNA Blood Mini Kit (Qiagen) and quantified using Prime Time qPCR assays (IDT). All samples were processed in triplicates. Each reaction consisted of 50 μl , containing 25 μl of iTaq Universal Supremix (BioRad), 2.5 μl of barcode primers and probe (primer 1: 5'-ATCCAGAGGTTGATTGTCCAGACGCGT-3', primer 2: 5'-CAGAATCGTTGCTGCACATCTTGAAAC-3', FAM probe: 5'-/56-FAM/AAAGGCTCGA/ZEN/GACGTAGTCAGACGT/3IABkFQ/-3'), 2.5 μl of housekeeping (NM_172901.2) primers and probe (primer 1: 5'-GACTTGTAAACGGCAGGCAGATTTGTG-3', primer 2: 5'-GAGGTGTGGGTACCTCGACATC-3', HEX probe: 5'-/5HEX/CCGTGTCGC/ZEN/TCTGAAGGGCAATAT/3IABkFQ/-3, IDT) and 20 μl of gDNA sample (100 ng). The cycling conditions were 1 cycle of denaturation at 95 °C for 3 min, followed by 40 cycles of amplification (95 °C for 15 s, 68 °C for 1 min). qPCR analysis was performed on the Eppendorf

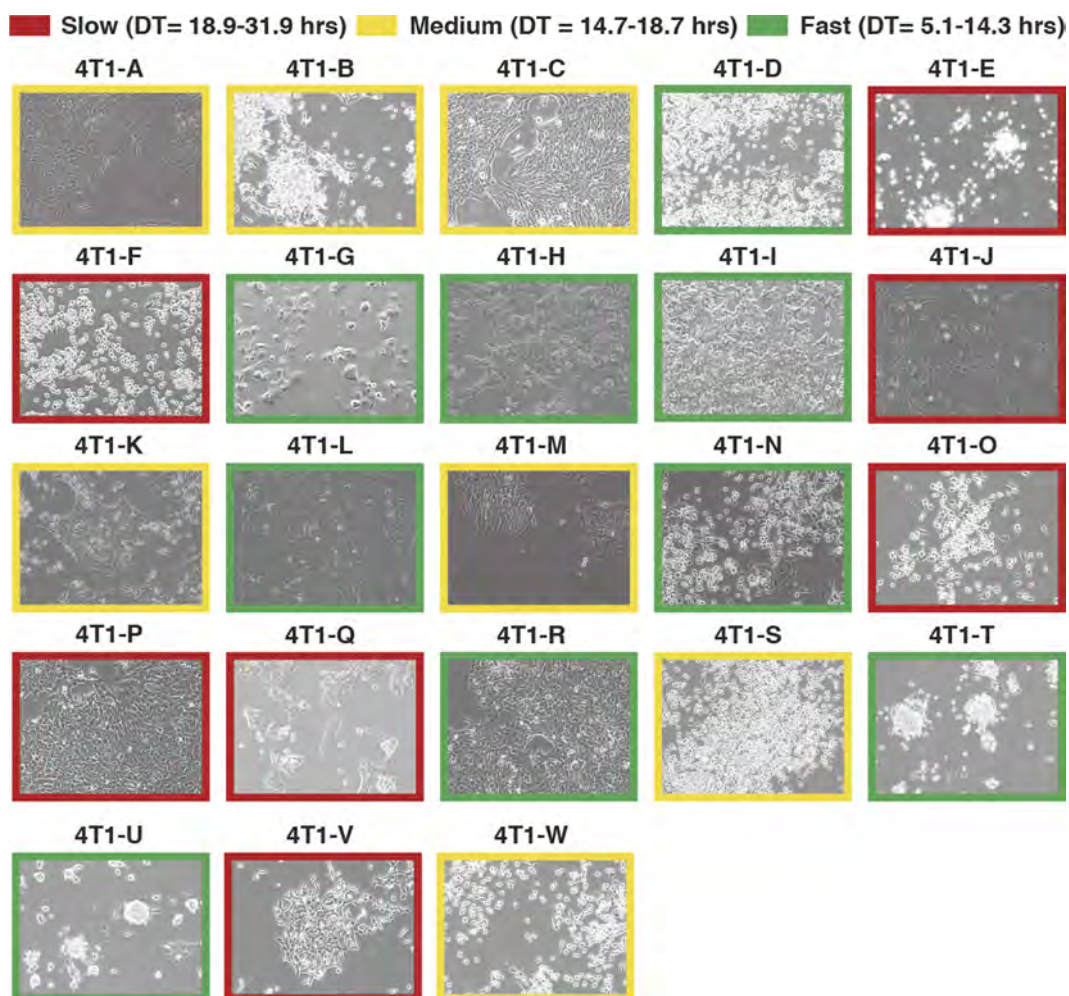
Mastercycler ep realplex. All signals were quantified using the ΔC_t method and were normalized to the levels of the housekeeping gene.

23. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
24. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
25. Anders, S., Pyl, P. T. & Huber, W. HTSeq — A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
26. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).



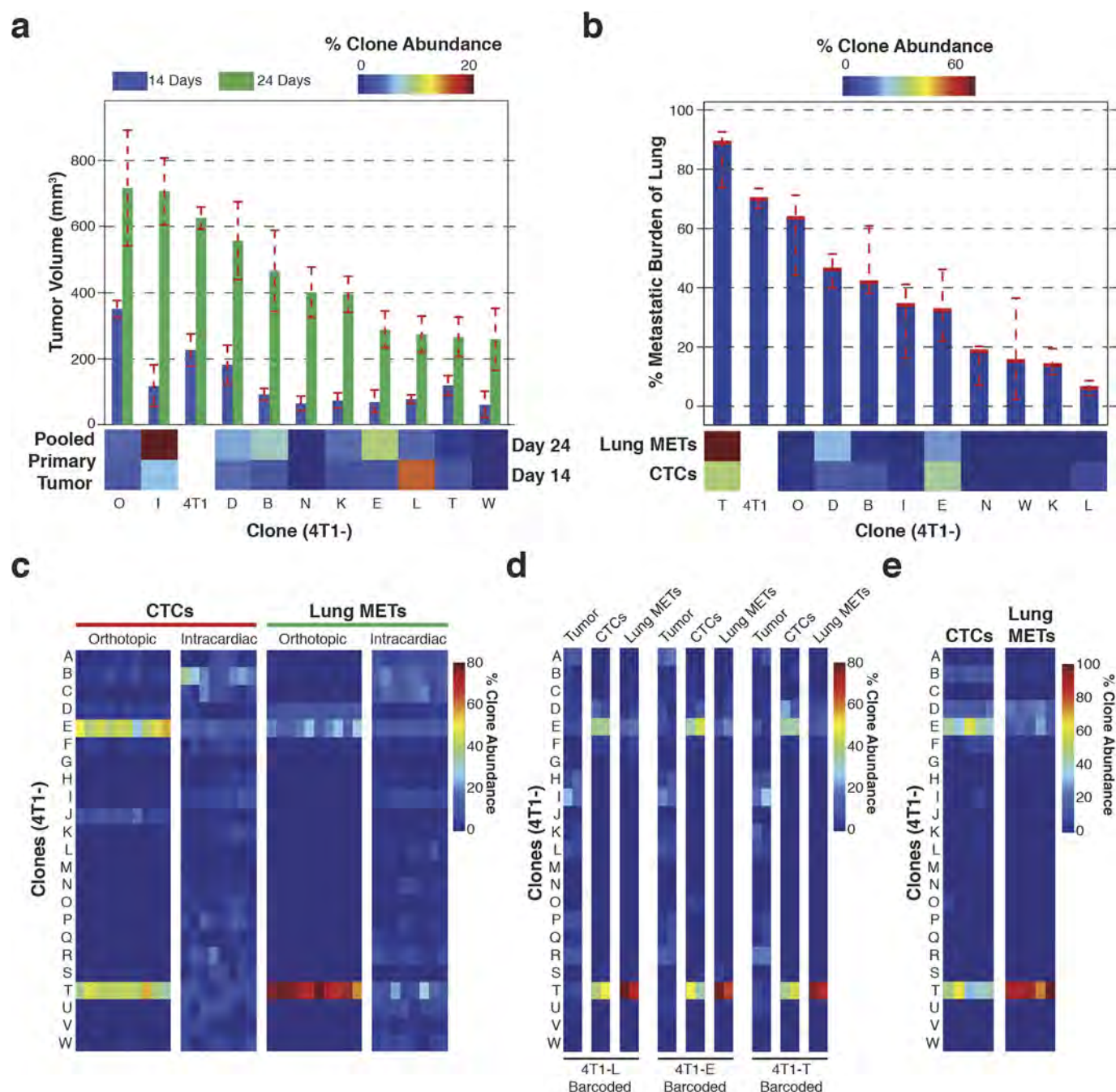
Extended Data Figure 1 | Overview of vector plasmids and clonal abundances in the primary tumour and secondary sites after 4T1 transplantation. **a**, Schematic of the retroviral barcode vector (used for Figs 1, 2d, 4e and Extended Data Fig. 8f). **b**, Schematic of the lentiviral barcode vector (used for Fig. 2a–c). **c**, Schematic of the retroviral shRNA vector used for single gene knockdown. **d**, Schematic of the tandem retroviral shRNA vector used for double gene knockdown. **e**, Schematic of the retroviral cDNA

vector used for gene overexpression. **f**, Distribution of clone abundances in the primary tumours and secondary sites for the 4T1 clones that engrafted and contributed to tumour formation in all five animals after orthotopic injection. **g**, Overlap of abundant clones in the primary tumour, lymph nodes, blood and in all blood-borne metastases (BBM: lung, liver and brain) ($P < 0.001$, hypergeometric test). **h**, Overlap of the abundant clones in the lymph nodes, liver, lung and brain.



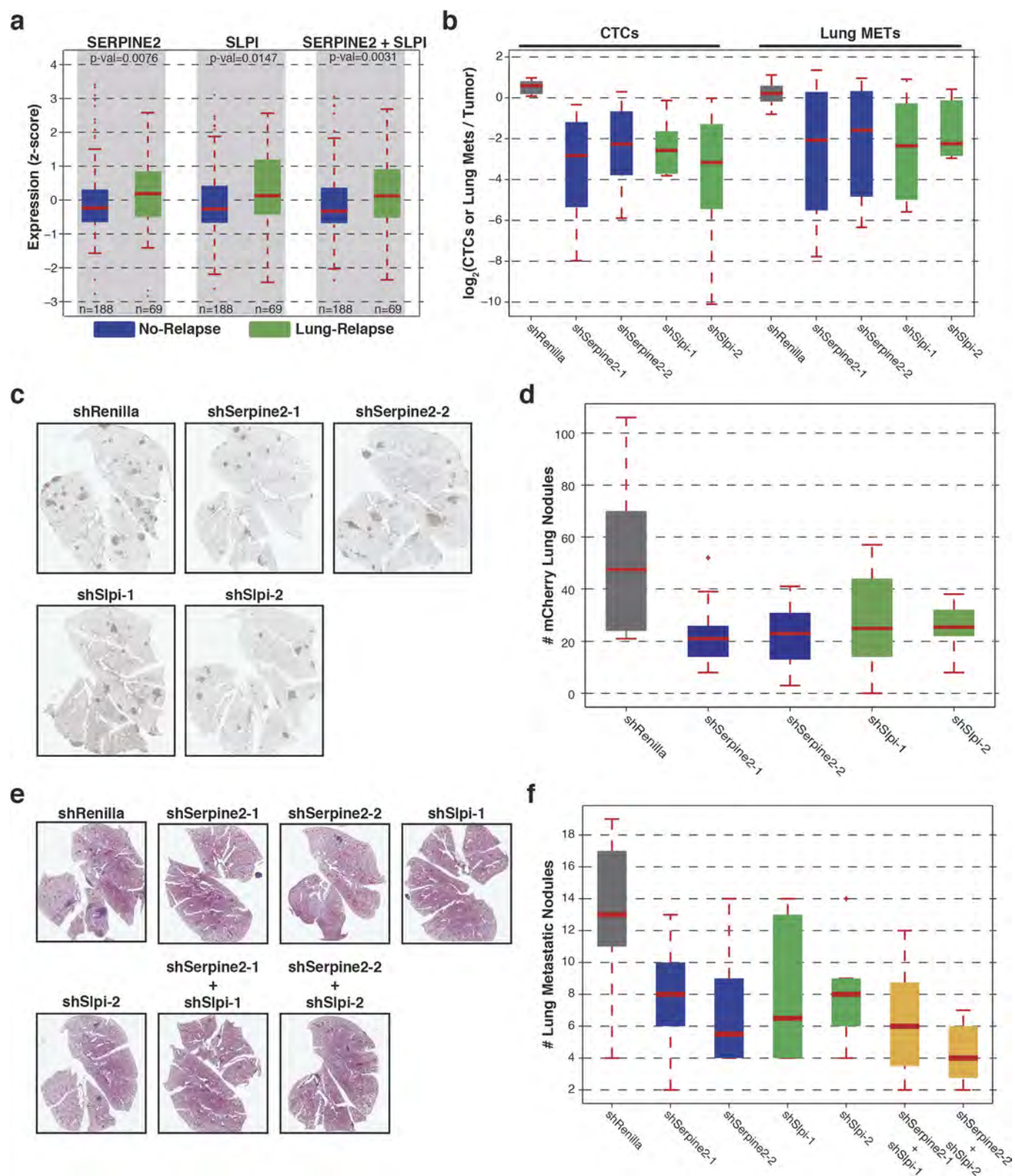
Extended Data Figure 2 | Morphology and proliferation rates of all clonal lines grown. Microscopic phase-contrast images of the 23 clonal lines

discussed in Figs 2–4. Doubling times (DT), as calculated over a 3-day period, are colour-coded in the image borders. Original magnification, $\times 10$.



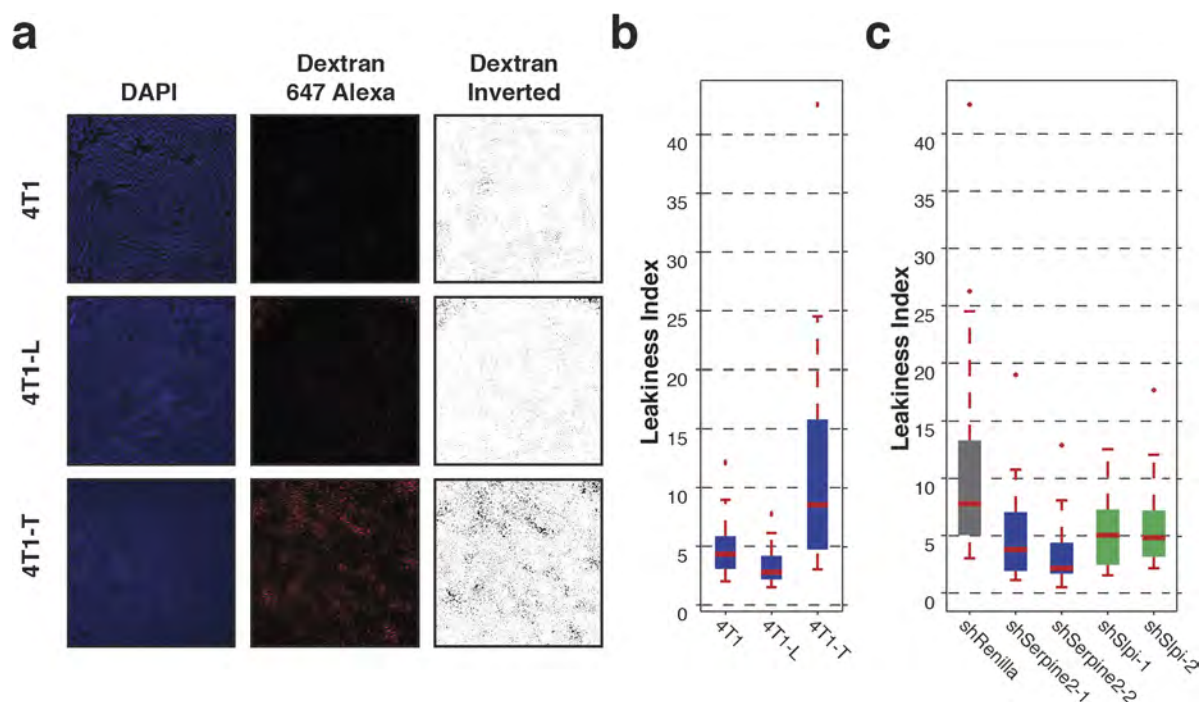
Extended Data Figure 3 | Tumour growth and metastases rates for individually injected clones and clonal composition of intracardiacally injected pool. **a**, Primary tumour volumes resulting from orthotopically injected individual clonal cell lines. Measurements were taken 14 and 24 days after injection ($n = 4$ mice). For comparison, the clonal composition of the primary tumour when the pool of clones is injected orthotopically is shown below. **b**, The percentage of lung metastatic burden for the animals discussed in **a** 24 days after injection ($n = 4$ mice). For comparison, the clonal composition of the lung when the pool of clones is injected orthotopically is shown below. **c**, A comparison of the clonal composition of CTCs and lung metastases when

the pool of 23 clonal lines was orthotopically injected (data from Fig. 2c) versus injected into the left cardiac ventricle of mice ($n = 10$ mice for orthotopic injections and $n = 8$ mice for intracardiac injections). **d**, Clonal compositions of the primary tumours, CTCs and lung metastases analysed in Fig. 2d. **e**, Clonal composition of CTCs and lung metastases after orthotopic injection when the individual clones were frozen down 3–5 times and cultured for a week each time ($n = 5$ mice). For all bar graphs, 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 twenty-fifth and seventy-fifth percentiles.



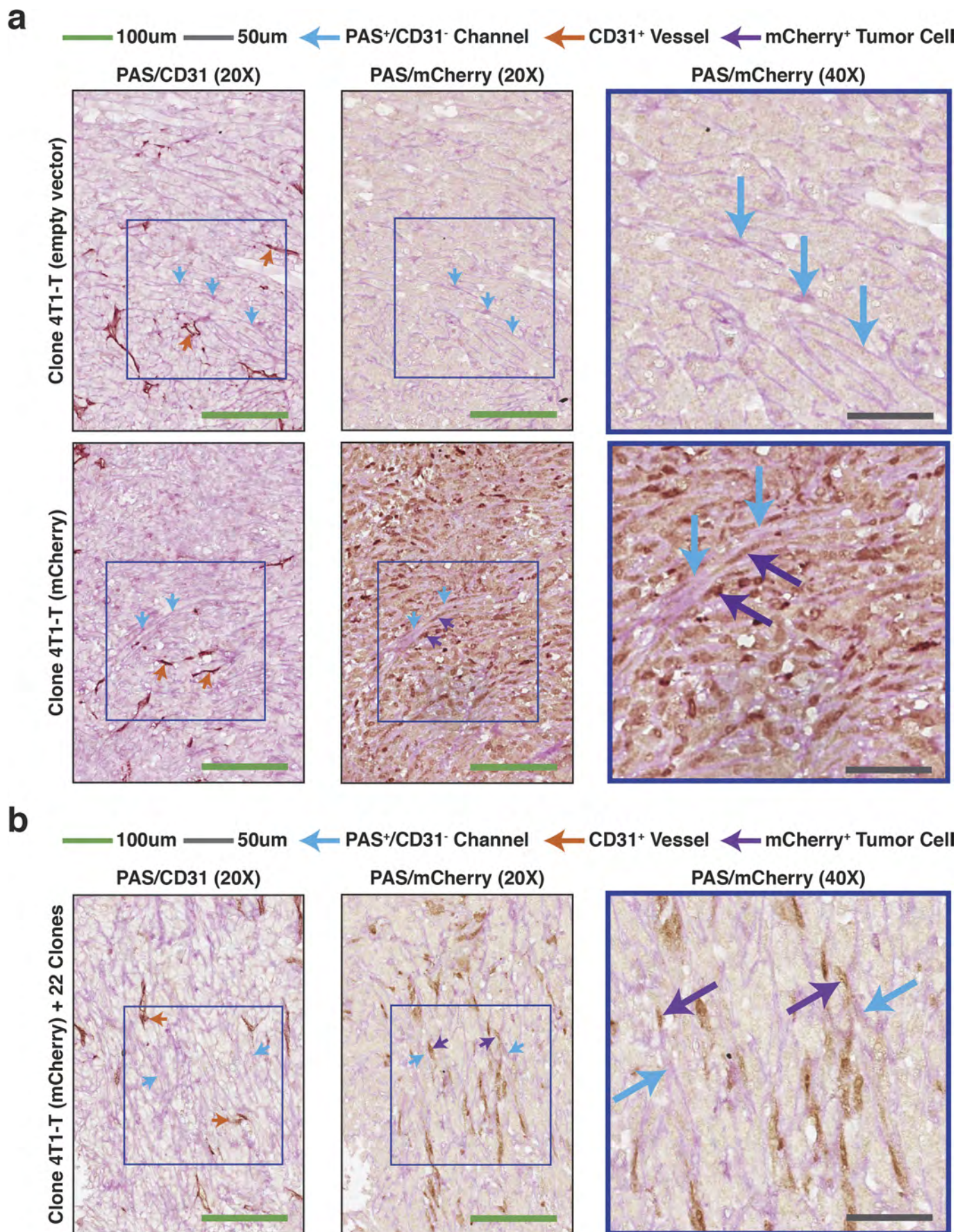
Extended Data Figure 4 | SERPINE2 and SLPI expression in human patients and *Serpine2* and *Slpi* shRNAs abundances in lung and CTCs after orthotopic and intracardiac injection. **a**, Analysis of *Serpine2* and *Slpi* human orthologues SERPINE2 and SLPI, respectively in basal, Her2 or claudin-low breast cancer patients with no relapse as compared to patients with relapse in the lung ($P < 0.005$, Wilcoxon rank-sum, n indicated in figure). **b**, A non-targeting shRNA and two targeting each of *Serpine2* and *Slpi* were infected separately into 4T1-T cells. After selection, the separately infected cells were pooled in equal amounts with the remaining 22 clonal lines and orthotopically injected into mice. The proportions of shRNAs shown are that of the CTCs and lung metastases in comparison to those in the primary tumour ($P < 0.01$, Wilcoxon rank-sum, $n = 10$ mice). **c**, 4T1-T cells were infected with a non-targeting shRNA or shRNAs targeting *Serpine2* or *Slpi*. The shRNAs were contained in a vector that constitutively expressed mCherry. The infected cells

were then pooled with the other 22 clones and injected orthotopically ($n = 10$ mice). Shown are representative images of the resultant mCherry positive lung metastatic nodules. **d**, Quantification of all mCherry positive metastatic lung nodules in all mice ($P < 0.01$, Wilcoxon rank-sum, $n = 10$ mice). **e**, 4T1-parental cells were infected with a non-targeting shRNA or shRNAs targeting *Serpine2* or *Slpi*. In each case the cells were then orthotopically injected into mice. Shown are representative images of haematoxylin-and-eosin-stained lung sections ($n = 10$ mice). **f**, Quantification of lung metastatic nodules in all haematoxylin-and-eosin-stained sections described in **e** ($P < 0.01$ and $P < 0.005$ for single- and double-knockdowns, respectively, Wilcoxon rank-sum, $n = 10$ mice). For all box plots, the edges of the box are the twenty-fifth and seventy-fifth percentiles. The 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 twenty-fifth and seventy-fifth percentiles.



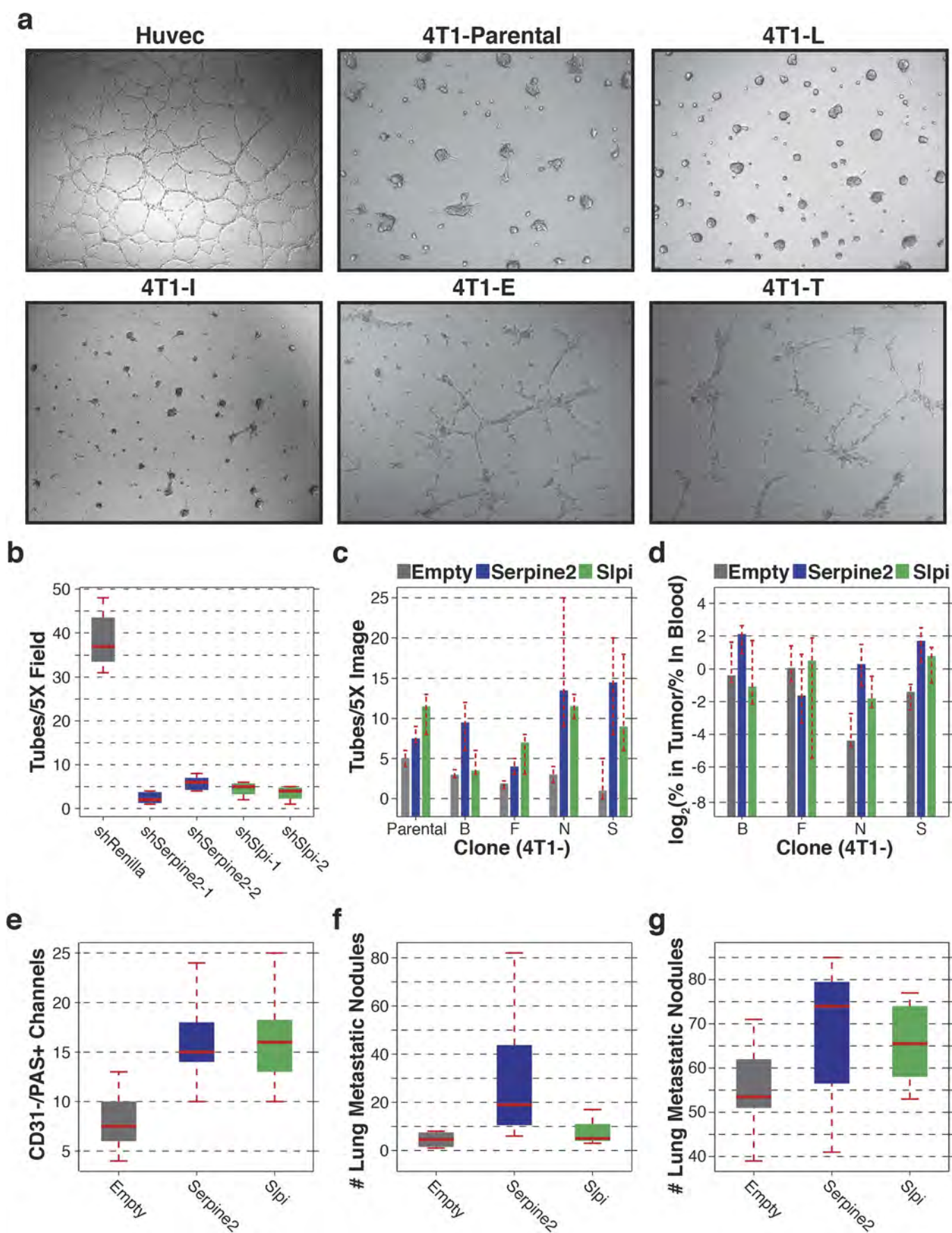
Extended Data Figure 5 | Leakiness index of clonal cell lines. **a**, DAPI and dextran Alexa 647 (+ dextran inverted)-stained tumour sections from orthotopic tumours derived from parental-4T1, 4T1-L and 4T1-T cells. **b**, Leakiness index of primary tumours resulting from orthotopic injection of parental-4T1, 4T1-L and 4T1-T cells ($P < 0.05$, Wilcoxon rank-sum, $n = 15$ mice). **c**, The leakiness index of primary tumours derived from 4T1-T cells that

have been infected with either a non-targeting shRNA or an shRNA targeting *Serpine2* or *Sipi* ($P < 0.03$, Wilcoxon rank-sum, $n = 18$ mice). For all box plots, the edges of the box are the twenty-fifth and seventy-fifth percentiles. The error bars extend to the values $q_3 + w(q_3 - q_1)$, and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 are the twenty-fifth and seventy-fifth percentiles.



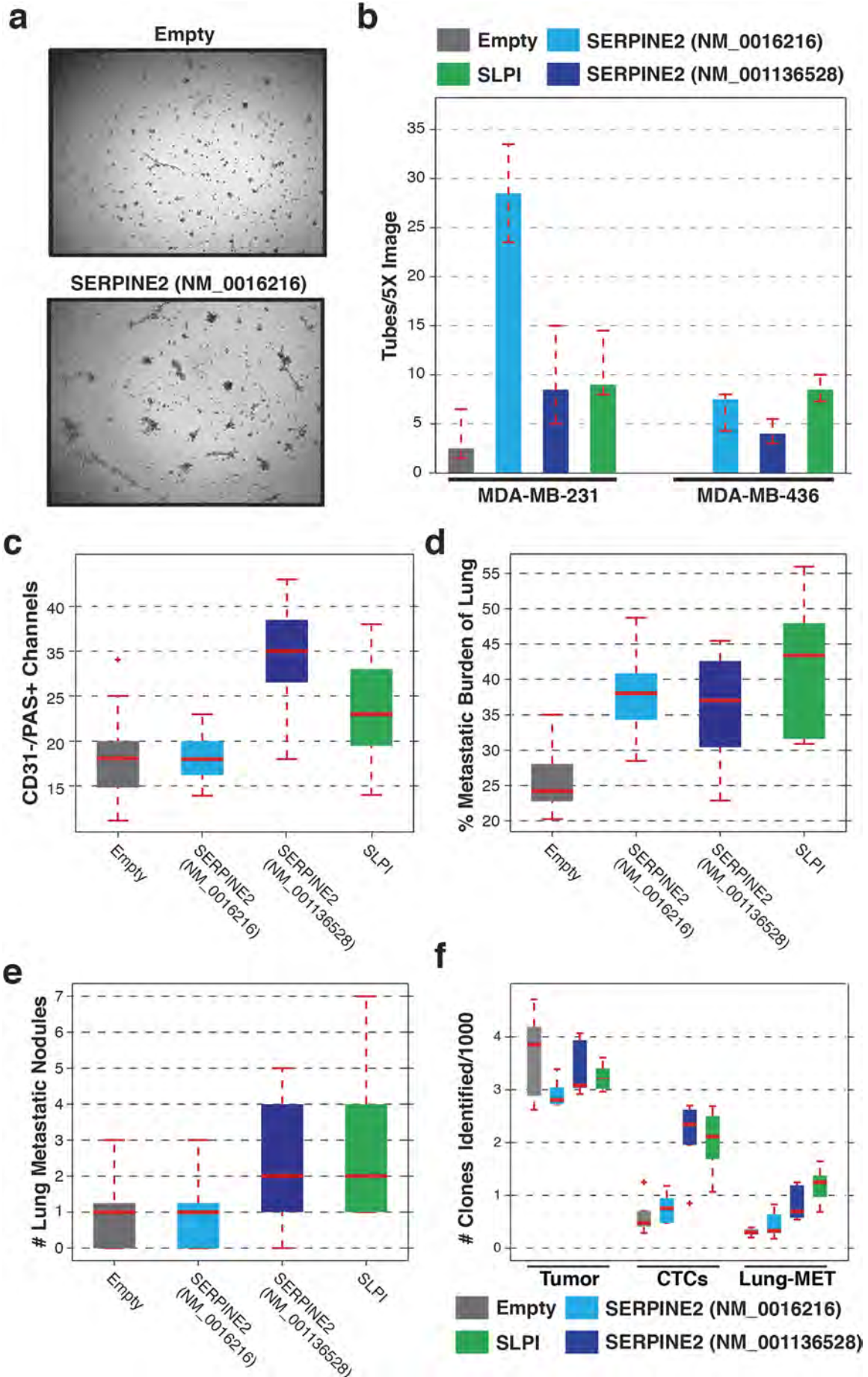
Extended Data Figure 6 | Vascular mimicry in 4T1-T and pooled primary tumours. **a**, PAS/CD31 (left) and PAS/mCherry (right) staining of serially sectioned 4T1-T derived primary tumours, where the tumour cells constitutively express an empty vector or mCherry. Orange arrows indicate CD31⁺ endothelial blood vessels and blue arrows show PAS⁺/CD31⁻ channels. The purple arrows indicate mCherry-positive tumour cells adjacent to PAS⁺ channels (green scale bars = 100 μ m, grey scale bars = 50 μ m).

b, PAS/CD31 (left) and PAS/mCherry (right) staining of serially sectioned primary tumours resulting from orthotopic injection of a pool of the 23 clones, in which the 4T1-T clonal line constitutively expresses mCherry. Orange arrows indicate CD31⁺ endothelial blood vessels and blue arrows show PAS⁺/CD31⁻ channels. The purple arrows indicate mCherry-positive tumour cells adjacent to PAS⁺ channels (green scale bars = 100 μ m, grey scale bars = 50 μ m).



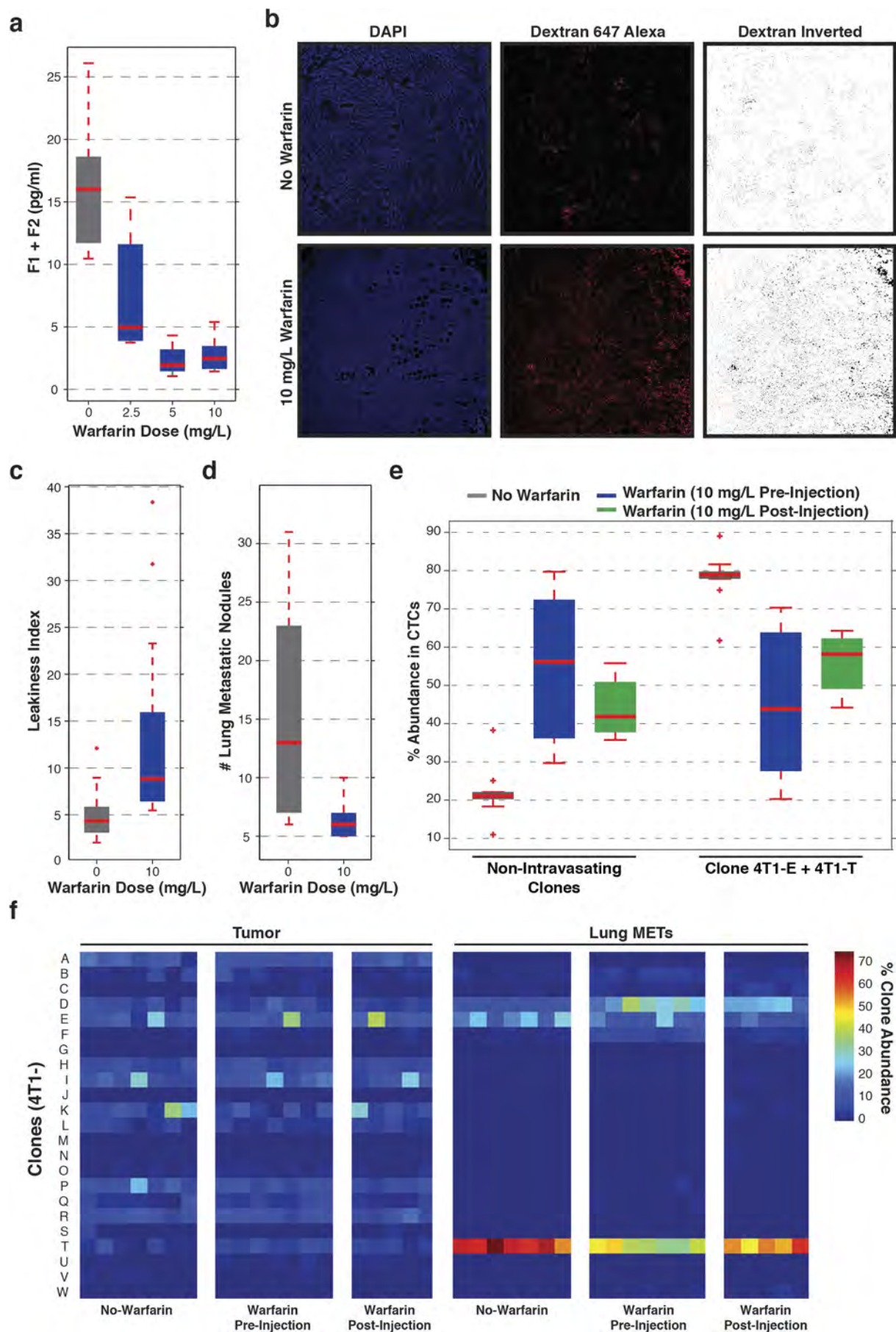
Extended Data Figure 7 | Overexpression of Serpine2 and Slpi in parental 4T1 and clonal cell lines. **a**, Images of HUVECs, parental 4T1, 4T1-L, 4T1-I, 4T1-E and 4T1-T cells grown on Matrigel to assess tube formation ability. **b**, Number of tubular structures identified per $\times 5$ microscopic field when 4T1-T cells had been infected with a non-targeting shRNA or shRNAs targeting *Serpine2* or *Slpi* and grown on Matrigel ($P < 0.0002$, Wilcoxon rank-sum, $n = 8$ fields). **c**, Number of tubular structures identified per $\times 5$ microscopic image when non-intravasating 4T1 clonal lines had been infected with an empty retroviral vector or vectors for overexpression of *Serpine2* or *Slpi* ($P < 0.03$, Wilcoxon rank-sum, with the exception of 4T1-B overexpressing *Slpi*, $n = 4$ fields). **d**, Relative CTC proportions of the cells described in **c** after they had been pooled with the remaining clonal lines and orthotopically injected

($P < 0.05$, Wilcoxon rank-sum, with the exception of 4T1-F and 4T1-B *Slpi*, $n = 10$ mice). **e**, PAS⁺/CD31[−] channels in primary tumours resulting from orthotopic injection of parental 4T1 cells infected with an empty retroviral vector or vectors for overexpression of *Serpine2* or *Slpi* ($P < 0.01$, Wilcoxon rank-sum, $n = 10$ fields). **f**, Quantification of the number of lung metastatic nodules resulting from the tumours described in **e** after 18 days ($n = 4$ mice). **g**, Quantification of the number of lung metastatic nodules resulting from the tumours described in **e** after 24 days ($P < 0.05$, Friedman, $n = 10$ mice). For all box plots, the edges of the box are the twenty-fifth and seventy-fifth percentiles. The error bars extend to the values $q_3 + w(q_3 - q_1)$, and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 are the twenty-fifth and seventy-fifth percentiles.



Extended Data Figure 8 | Overexpression of SERPINE2 and SLPI in human breast cancer cell lines. **a**, MDA-MB-231 cells infected with an empty overexpression vector plated on Matrigel (top) and MDA-MB-231 cells overexpressing SERPINE2 (NM_0016216) plated on Matrigel (bottom). **b**, Full quantification of tubular structures in basal cell lines MDA-MB-231 and MDA-MB-436 when infected with an empty overexpression vector or when infected with vectors for overexpressing SERPINE2 or SLPI ($P < 0.05$, Friedman test, for SERPINE2 isoform NM_0016216 and SLPI, $n = 4$ fields). **c**, Quantification of CD31⁺/PAS⁺ channels in MDA-MB-436-derived primary tumours cells that have been infected with an empty retroviral vector or vectors for overexpression of SERPINE2 or SLPI ($P < 0.002$, Wilcoxon rank-sum, for SERPINE2 isoform NM_001136528 and SLPI, $n = 20$ fields for empty, SERPINE2 NM_001136528 and NM_0016216 and $n = 25$ fields for SLPI). **d**, The percentage of lung metastatic burden after orthotopic injection of MDA-MB-231 cells that have been infected with an empty retroviral vector or vectors

for overexpression of SERPINE2 or SLPI ($P < 0.05$, Wilcoxon rank-sum, $n = 9$ mice). **e**, Quantification of the number of lung metastatic nodules after orthotopic injection of MDA-MB-436 cells that have been infected with an empty retroviral vector or vectors for overexpression of SERPINE2 or SLPI (Wilcoxon rank-sum $P < 0.05$ for SERPINE2 isoform NM_001136528 and SLPI, $n = 9$ mice). **f**, Sub-clonal analysis of MDA-MB-436 cells that have been infected with an empty retroviral vector or vectors for overexpression of SERPINE2 or SLPI. Each cell line was infected with a barcode library and then the number of clones were quantified in the primary tumour, cardiovascular CTCs and lung metastatic cells ($P < 0.02$, Wilcoxon rank-sum, for SERPINE2 isoform NM_001136528 and SLPI in CTCs and lung metastasis, $n = 5$ mice). For all box plots, the edges of the box are the twenty-fifth and seventy-fifth percentiles. The 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 twenty-fifth and seventy-fifth percentiles.



Extended Data Figure 9 | Effect of warfarin on leakiness, metastasis and clonal abundance. **a**, Levels of prothrombin fragments F1 and F2 (active component of blood coagulation), as quantified by ELISA, after administration of warfarin in the drinking water of mice ($P < 0.01$, Wilcoxon rank-sum, $n = 4$ mice). **b**, DAPI and dextran Alexa 647 staining to visualize vascular leakage in primary tumours derived from the pool of 23 clonal lines. Animals were administered drinking water either with or without 10 mg ml^{-1} warfarin. **c**, Leakiness index of primary tumours resulting from orthotopic injection of the pool of 23 clonal lines. After injection mice were administered drinking water with no warfarin or water containing 10 mg ml^{-1} warfarin ($P < 0.000005$, Wilcoxon rank-sum, $n = 10$ mice). **d**, Counts of lung metastatic nodules in animals that were injected intravenously with the pool of 23 clonal lines administered drinking water with no warfarin or water containing

10 mg ml^{-1} of warfarin ($P < 0.05$, Wilcoxon rank-sum, $n = 10$ mice). **e**, The pool of 23 clonal lines was injected orthotopically and the mice administered drinking water with no warfarin or 10 mg ml^{-1} warfarin. The percentage abundance of each clone in the cardiovascular CTCs of animals (either pre- or post-injection, $P < 0.002$, $n = 7$ mice for no warfarin, $n = 7$ mice for pre-injection and $n = 5$ mice for post-injection). **f**, The clone proportions in the resultant primary tumours and lung metastases described in (e) ($P < 0.003$, Wilcoxon rank-sum, for lung metastases, $n = 7$ mice for no warfarin, $n = 7$ mice for pre-injection and $n = 5$ mice for post-injection). For all box plots, the edges of the box are the twenty-fifth and seventy-fifth percentiles. The error bars extend to the values $q_3 + w(q_3 - q_1)$, and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 are the twenty-fifth and seventy-fifth percentiles.

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. Kim², Mar Soto², Michel Wagner⁷, Marc-Olivier Turgeon⁷, Lisa Fish^{8,9,10}, Nicolas Erard¹, Annika L. Gable², Ashley R. Maceli², Steffen Dickopf², Evangelia K. Papachristou¹, Clive S. D'Santos¹, Lisa A. Carey¹¹, John E. Wilkinson¹², J. Chuck Harrell¹³, Charles M. Perou¹⁴, Hani Goodarzi^{8,9,10}, George Poulgiannis^{7,15} & Gregory J. Hannon^{1,2,6}

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 sites¹. 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 subtypes³ (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 cells⁴. These were placed 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

¹Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ²Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ³Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048, USA. ⁴Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ⁵Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ⁶New York Genome Center, 101 6th Avenue, New York, New York 10013, USA. ⁷Division of Cancer Biology, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. ⁸Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94158, USA. ⁹Department of Urology, University of California, San Francisco, San Francisco, California 94158, USA. ¹⁰Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California 94158, USA. ¹¹Division of Hematology and Oncology, University of North Carolina at Chapel Hill, 170 Manning Drive, CB7305, Chapel Hill, North Carolina 27599, USA. ¹²Department of Pathology, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA. ¹³Department of Pathology, Virginia Commonwealth University, Richmond, Virginia 23284, USA. ¹⁴Department of Genetics and Pathology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁵Division of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, London SW7 2AZ, UK.

*These authors contributed equally to this work.

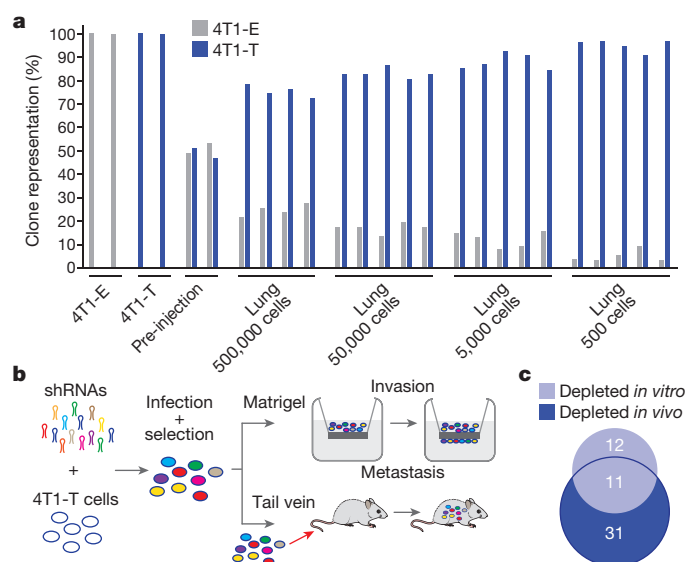


Figure 1 | Identification of metastatic drivers. **a**, Relative proportions of 4T1-E and -T cells extracted from the lungs of NSG mice, into which mixtures of cells were introduced via tail vein at different concentrations. Each bar represents a sample or independent mouse. **b**, RNAi screening scheme 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 $P < 0.01$).

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 non-essential 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).

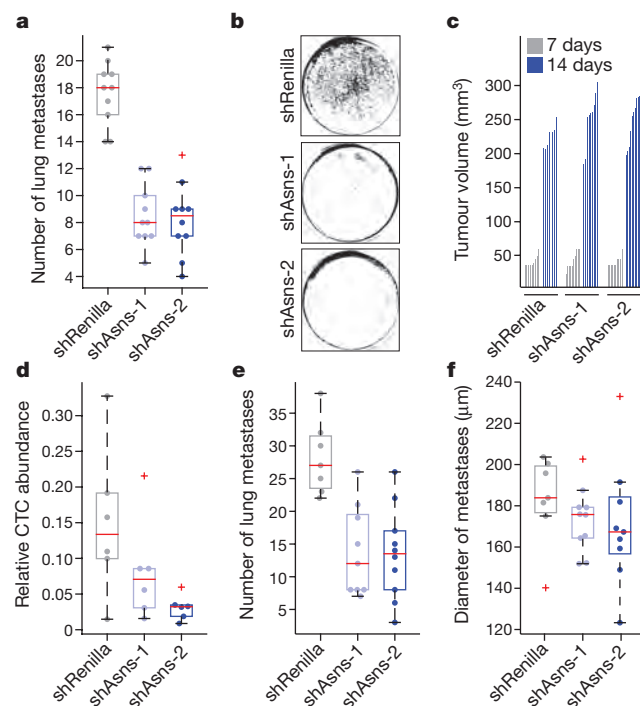


Figure 2 | Validation of Asns as a driver of invasion and metastasis.

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 $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 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.

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 site⁹. 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 asparagine¹⁰. 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)

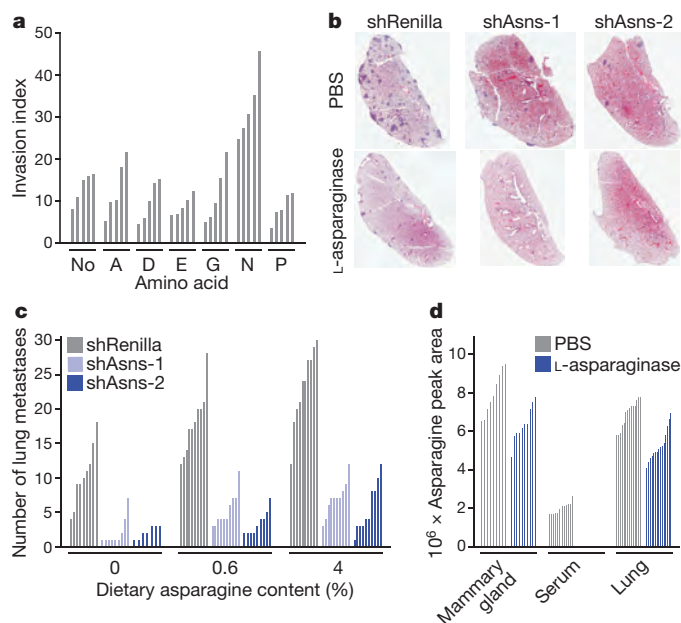


Figure 3 | Extracellular asparagine availability drives invasion and metastasis. **a**, Quantification of parental 4T1 cell invasion rates, as measured by the Matrigel invasion assay, in culture medium supplemented with the indicated NEAAs ($n = 5$ invasion chambers, rank-sum $P < 0.01$). **b**, Representative 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, rank-sum $P < 0.05$ 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.

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 patients¹¹.

To understand the mechanism by which asparagine availability might affect invasion and metastasis, we examined expression changes

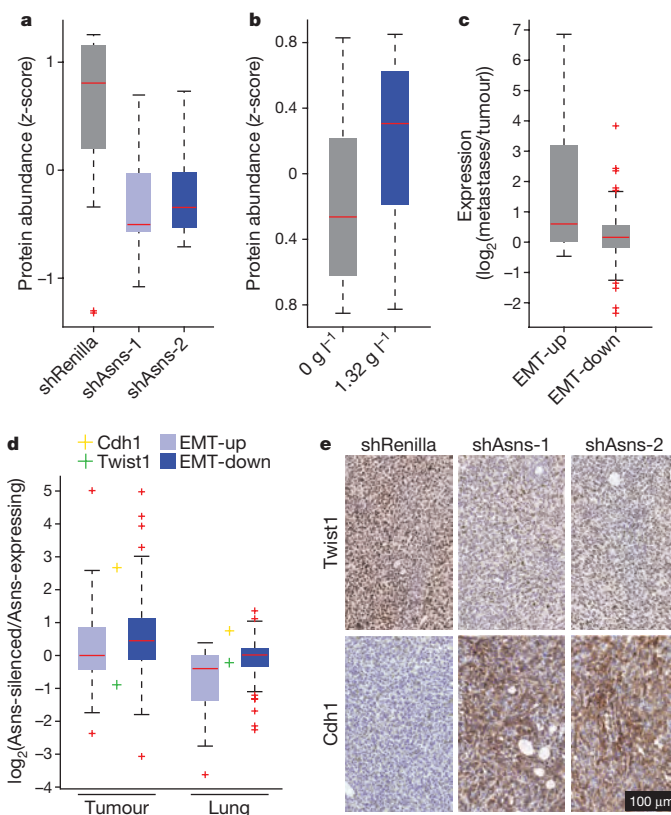


Figure 4 | Asparagine availability regulates EMT. **a**, Relative protein abundances of EMT-up proteins in Asns-silenced and -expressing 4T1-T cells ($n = 3$ replicates per cell line, edges of the box are the 25th and 75th percentiles, and error bars extend to the values $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which 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 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.

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 changes¹² (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 upon 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 asparagine¹² (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 EMT¹⁶ (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.

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.

Received 5 July 2016; accepted 15 December 2017.

Published online 7 February 2018.

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. *Nat. Commun.* **7**, 11457 (2016).
7. Richards, N. G. & Schuster, S. M. Mechanistic issues in asparagine synthetase catalysis. *Adv. Enzymol.* **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. USA* **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-Baghadi, 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- β signaling. *Int. J. Cancer* **91**, 76–82 (2001).

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

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. Coblenz, and L. Bianco for support with mouse work; K. Cheng, J. Bourbonniere, D. Hoppe, A. Nourjanova, and R. Puzis for support with histology; E. Hodges and E. Lee for support with next-generation sequencing; and J. Johnson for assistance with HPLC. This work was also performed with the assistance of the Cancer Research UK, Cambridge Institute Proteomics Core Facility. 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 National Institutes of Health (NIH) (NCI R00 CA194077). Work in the G.P. laboratory is supported by the Institute of Cancer Research, London and a Cancer Research UK 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 Department of Defense Breast Cancer Research Program (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 analyses. E.K.P. and C.S.D. assisted with proteomic analyses. L.F. and H.G. assisted with ribosomal profiling analyses. M.W., M.T., and G.P. performed metabolite profiling experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. 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. Correspondence and requests for materials should be addressed to G.J.H. (Greg.Hannon@cruc.cam.ac.uk).

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

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 described¹.

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 described¹.

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 $500 \mu\text{g ml}^{-1}$ 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 pre-injection 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 CAG 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 described⁴. 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 portal¹⁷. 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 reported¹. Cells were either grown in 4T1 cell culture medium containing $60 \mu\text{M}$ 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 described¹. 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 parameters¹⁸. 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 $500 \mu\text{g ml}^{-1}$ hygromycin for 5 days and MDA-MB-231 cells were selected with $2 \mu\text{g ml}^{-1}$ 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 $600 \mu\text{g ml}^{-1}$ G418 and MDA-MB-231 cells were selected with $1,500 \mu\text{g ml}^{-1}$ G418.

cDNA overexpression genes. Mouse Asns: NM_012055.1. Human ASNS: NM_001673.2.

shRNA knockdown sequences. Mouse shAsns-1: TGCTGTTGACAGTGACGCC ACTGCCAATAAGAAAGTATATAGTGAAGCCACAGATGTATATACTTTCT TATTGGCAGTGTTGCCTACTGCCTCGGA.

Mouse shAsns-2: TGCTGTTGACAGTGACGCCACTATGAAGTTTGGG TTTATAGTGAAGCCACAGATGTATAATCCAAACTTCATAGTGTTCGCTACTGCCTCGGA.

Mouse shTgfb1-1: TGCTGTTGACAGTGAGCGCCAGTATATATATGTTCTTCAATAGTGAAGCCACAGATGTATTGAAGAACATATATATCTGTGCTTCTGCTACTGCTCGGA.

Mouse shTgfb1-2: TGCTGTTGACAGTGAGCGAAGTATATATGTTCTTCAAATAGTGAAGCCACAGATGTATTGAAGAACATATATATCTGTGCTTCTGCTCGGA.

Human shASNS-1: TGCTGTTGACAGTGAGCGCCAGAAGCTAAAGGTC TTGTTATAGTGAAGCCACAGATGTATAACAAGACCTTAGCTTCTGATGCTACTGCTCGGA.

Human shASNS-2: TGCTGTTGACAGTGAGCGCAGCAATGACAGAAGTGGATATAGTGAAGCCACAGATGTATATCCATCTTCTGTCATTGCTTTGCTTACTGCTCGGA.

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 ΔC_t 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'-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 AGC 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 CTCs. CTCs were quantified as previously described¹. 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'-GACTATTGAAGC TGTCCTTCC-3'; primer 2, 5'-CGCAGCTTACCTTGTAGAT-3'.

HEX probe: 5'-/56-FAM/TTCAAGTGG/ZEN/GAGCGCGTGATG AA/31ABkFQ/-3'.

Housekeeping probes and primers: primer 1, 5'-GACTTGTAAACGGGCA GGCAGATTGTG-3'; primer 2, 5'-GAGGTGTGGGTACCTCGACATC-3'.

HEX probe: 5'-/5HEX/CCGTGTCGC/ZEN/TCTGAAGGGCAATAT/31ABkFQ/-3'.

Quantification of lung metastasis. For each lung, 5 μ m 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 described¹. 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 $\times$ 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. 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 μ g 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 μ m, 2.1 $\times$ 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 μ m $\times$ 2 cm C18, 5 μ m, 100 Å trapping column with the uPickUp injection method at a loading flow rate of 4 μ l min⁻¹ for 10 min. For peptide separation, an EASY-Spray analytical column 75 μ m $\times$ 25 cm, C18, 2 μ m, 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 pre-filled 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 μ l) was performed using 200 μ l 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 μ l HPLC-grade water, and 1 μ l 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 $\times$ 100 mm, 2.5- μ m 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.

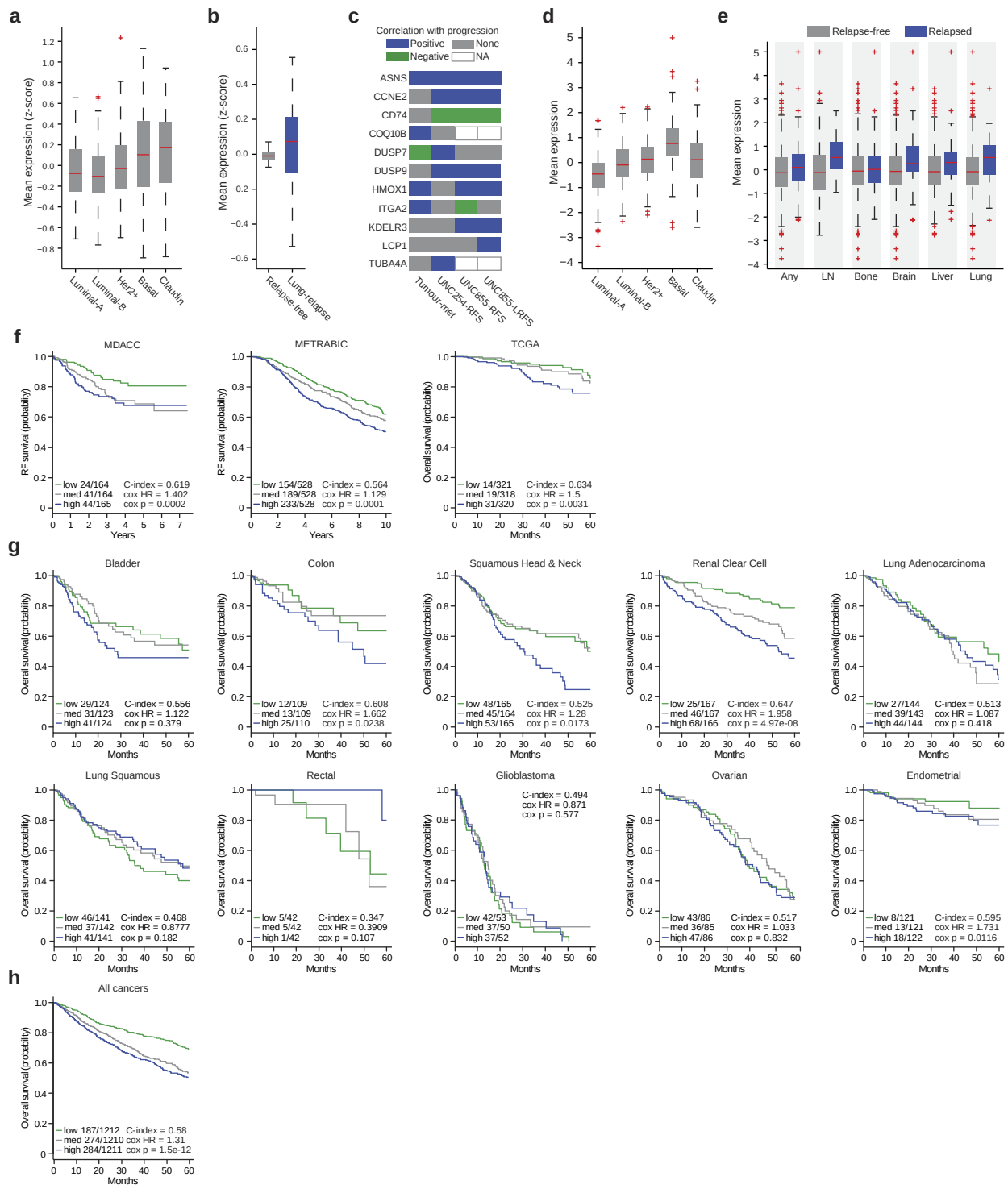
Ribosome profiling analysis. For analysis, quality trimming and sequencing adapter removal was performed using Cutadapt²². Bowtie2 was used to remove reads that mapped to contaminating RNAs (for example, rRNA and tRNA sequences)¹⁸. STAR was subsequently used to map reads of length 29–33 to the human 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.

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* **10**, 48 (2009).
18. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
19. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
20. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
21. Ruifrok, A. C. & Johnston, D. A. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**, 291–299 (2001).
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).
24. Vizcaino, *et al.* 2016 update of the PRIDE database and related tools. *Nucleic Acids Res* **44**, D447–D456 (2016).

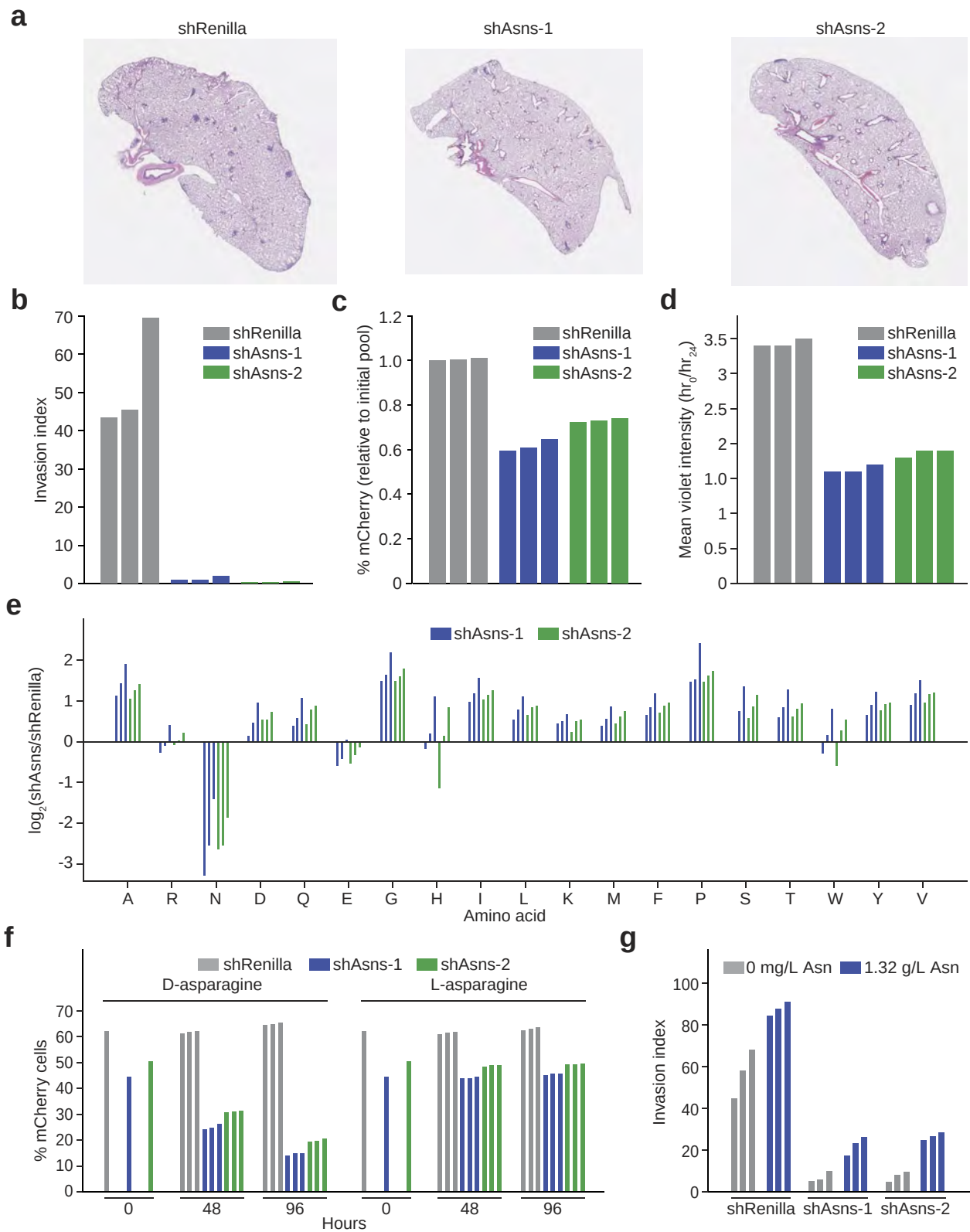


Extended Data Figure 1 | See next page for caption.

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 $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 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

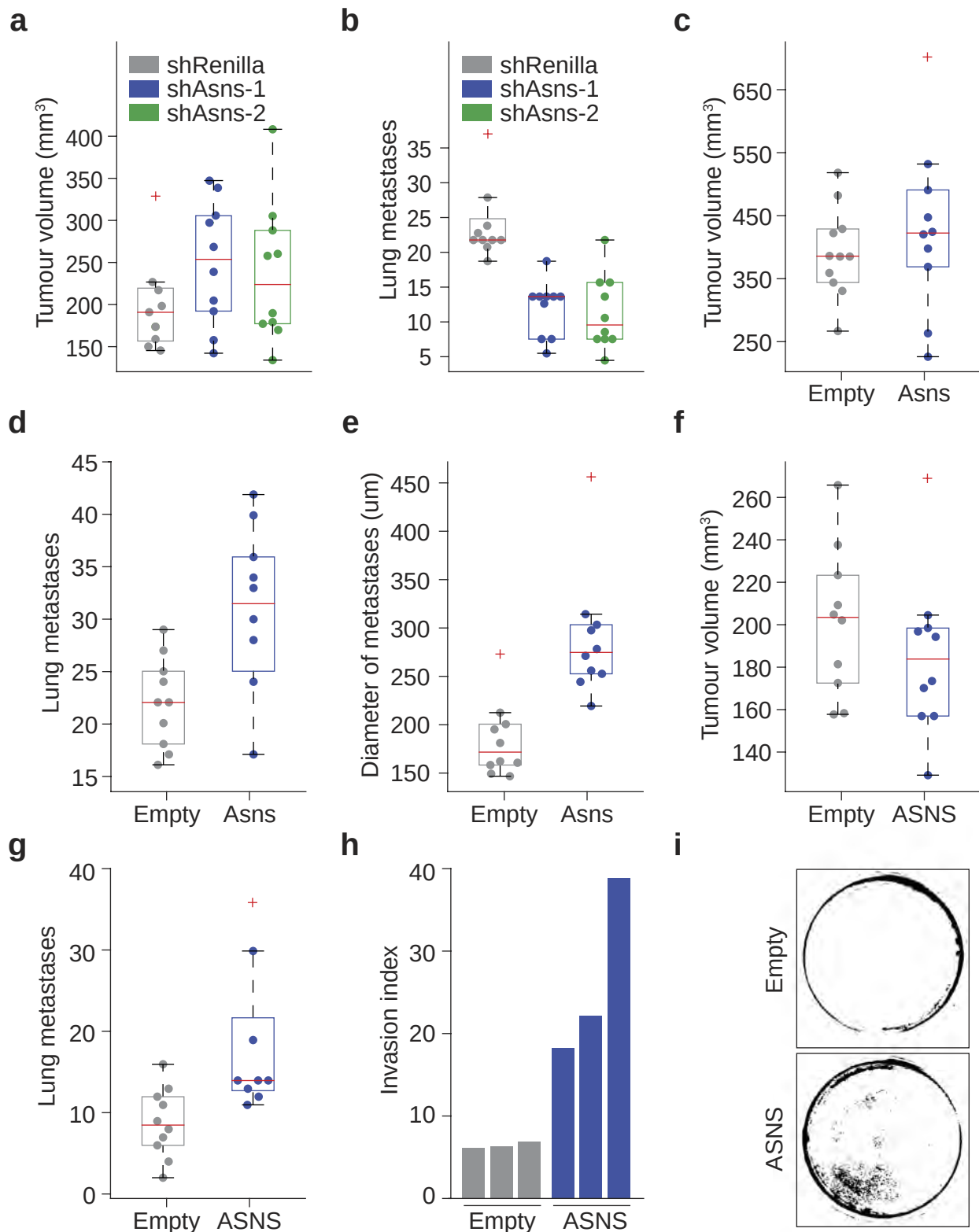
negatively correlated if these ratios were significant (Cox $P < 0.05$) and less than 1. As the UNC855 dataset also had site of relapse information, here both relapse-free and lung relapse-free survival (RFS and LRFS) hazard ratios were used to classify genes as positively or negatively correlated with 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 | See next page for caption.

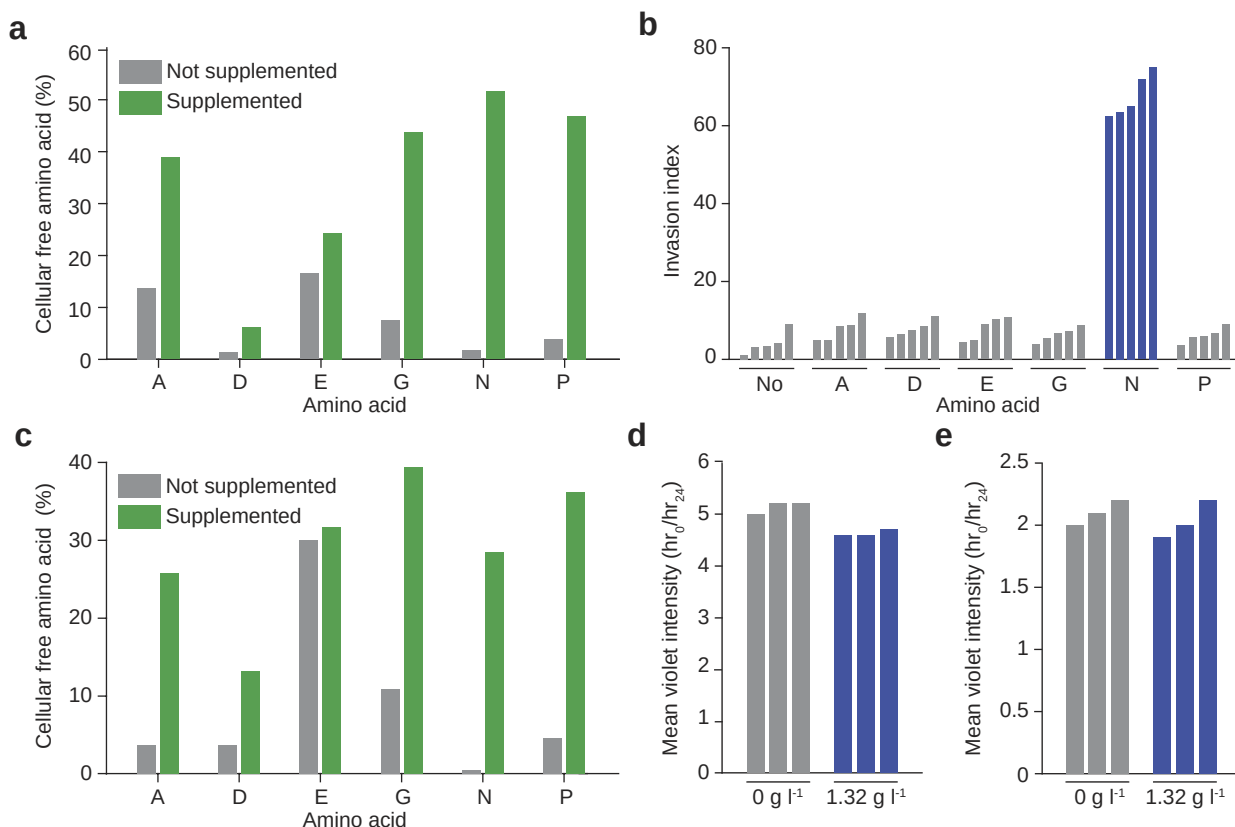
Extended Data Figure 2 | Primary validation of Asns as a driver of invasion and metastasis. **a**, Representative images of the lungs of mice that were intravenously injected with Asns-silenced or -expressing 4T1-T cells as described in Fig. 2a. **b**, Quantification of Matrigel invasion capacity for 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).



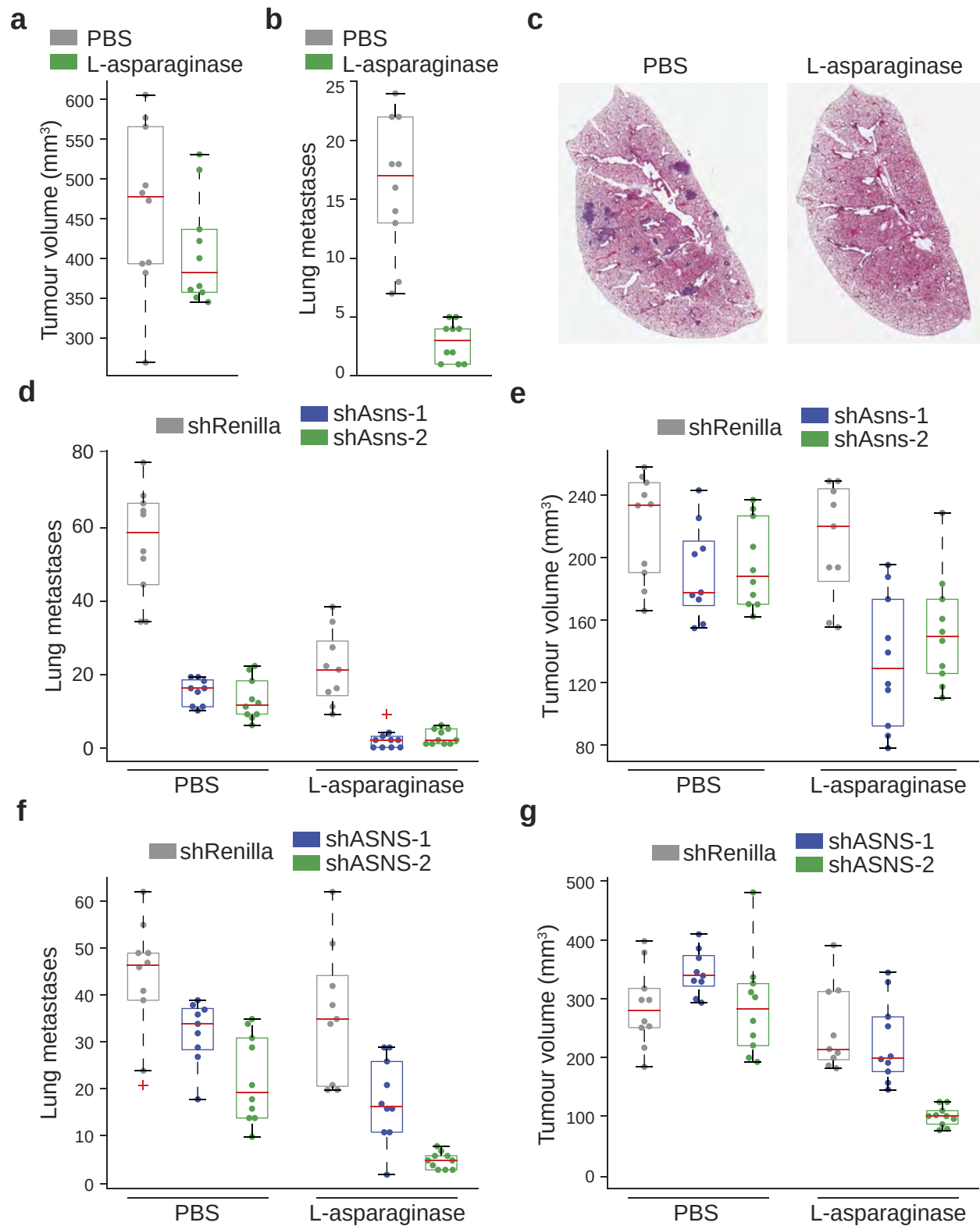
Extended Data Figure 3 | Secondary validation of *Asns* as a driver of invasion and metastasis. **a**, Volume measurements of tumours resulting from orthotopic injection of *Asns*-silenced and -expressing parental 4T1 cells ($n = 10$ mice per cell line, edges of the box are the 25th and 75th percentiles and error bars extend to the values $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 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 amino-acid percentages for MDA-MB-231 cells when cultured in the medium

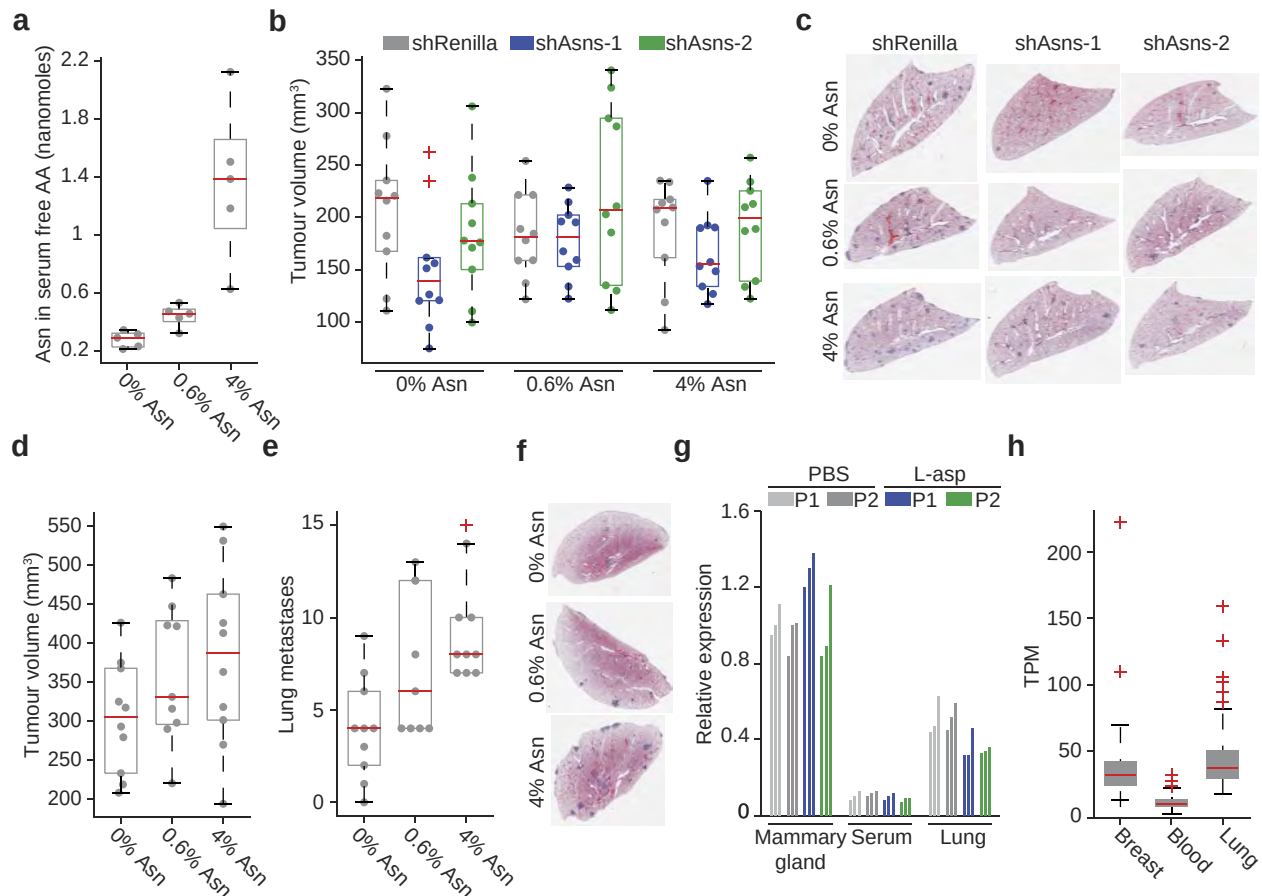
conditions described in **a** ($n = 3$ replicates per cell line). **d**, Violet cell-labelling intensity of parental 4T1 cells when grown in asparagine-lacking or -supplemented medium for the same period that the Matrigel invasion assay described in Fig. 3a was being performed ($n = 3$ replicates per 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 | See next page for caption.

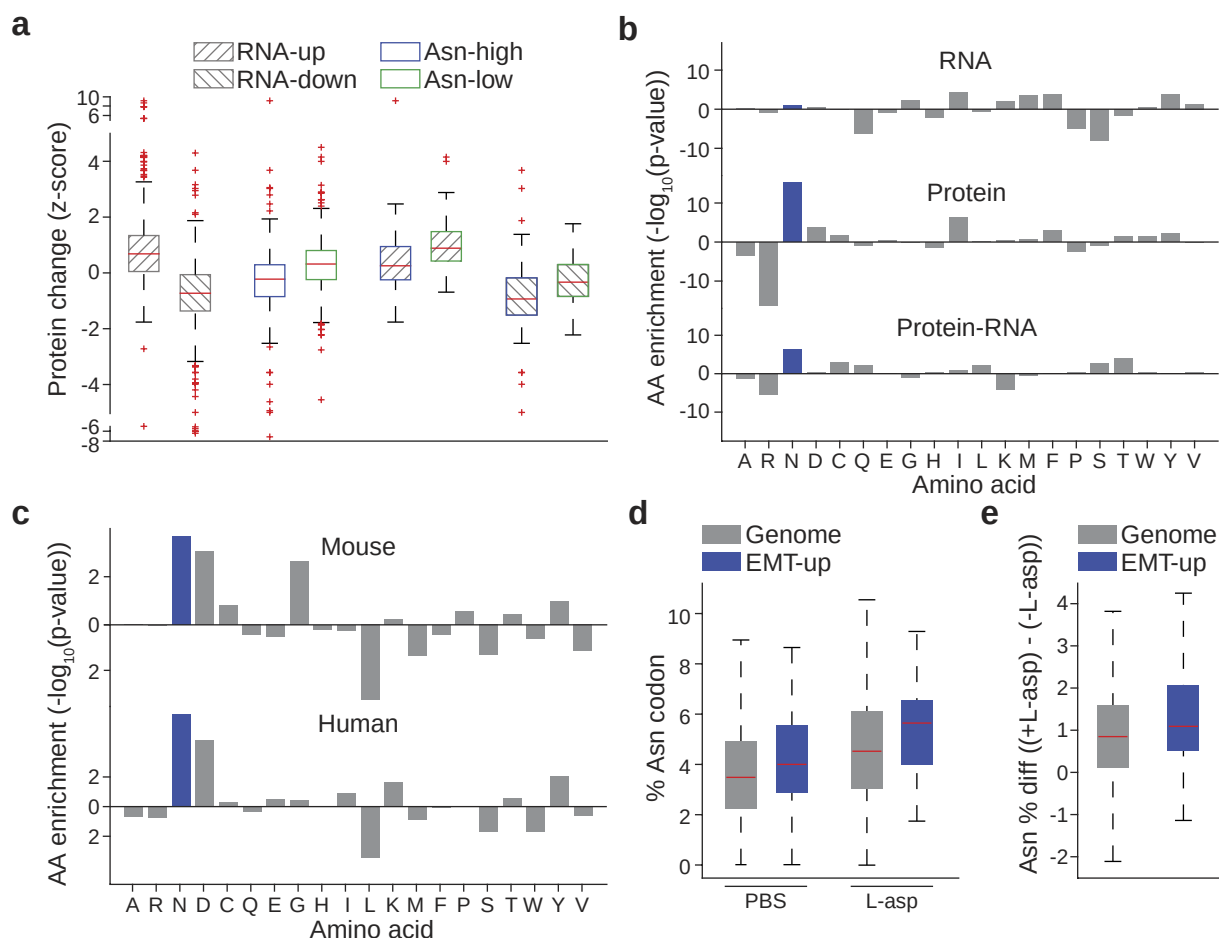
Extended Data Figure 5 | Secondary validation that extracellular asparagine availability affects invasion and metastasis. **a**, Tumour volumes resulting from the orthotopic injection of parental 4T1 cells. Half of the mice received L-asparaginase and the other half received an equivalent volume of PBS 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 $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 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 $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 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 orthotopic tumours corresponding to the lung metastases described in Fig. 3c, in which 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 in 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 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 $q_1 - w(q_3 - q_1)$, in which 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, whereas positive correlations indicate the amino acid is enriched.

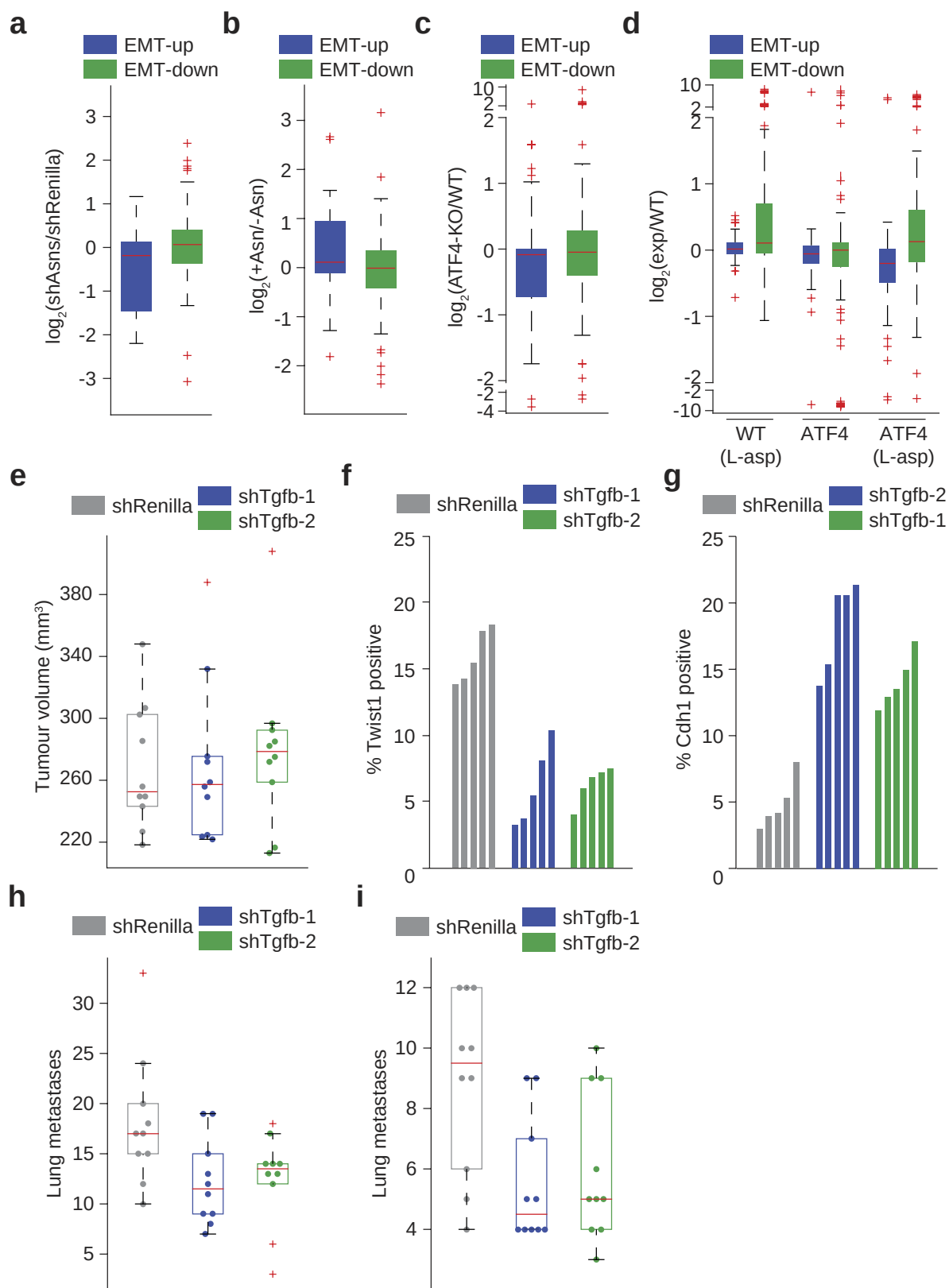
For protein minus RNA level expression changes, amino acids with positive correlations are enriched in proteins in which depletion levels exceed what is predicted by corresponding RNA changes. Negative correlations indicate the amino acid is enriched in proteins in which depletion levels are less than 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.

Knott et al. 2017, Extended Data Figure 8



Extended Data Figure 8 | Conservation of asparagine enrichment in EMT promoting proteins. Asparagine enrichment analysis of EMT-promoting 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).

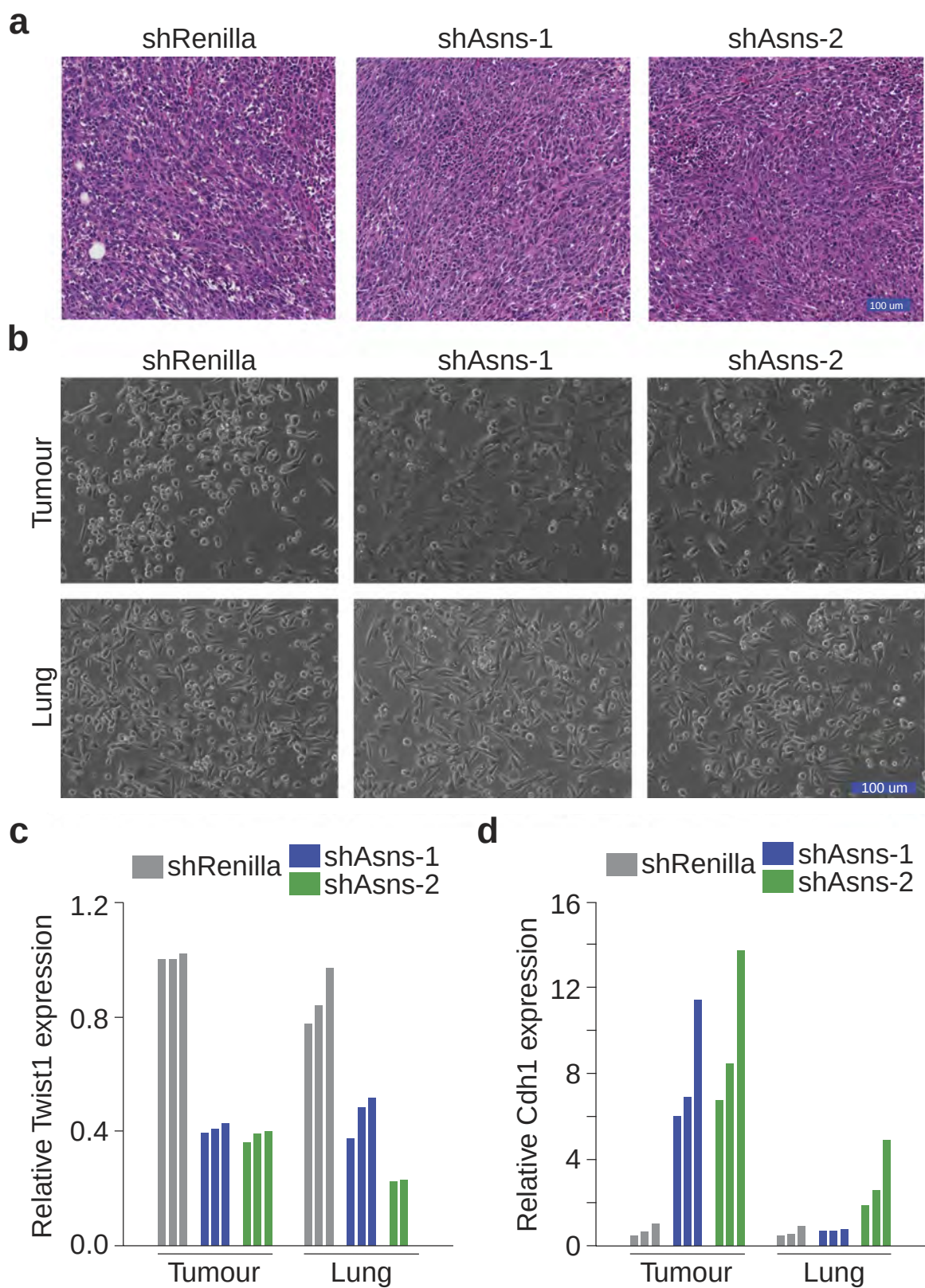


Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Secondary validation that asparagine availability regulates EMT.

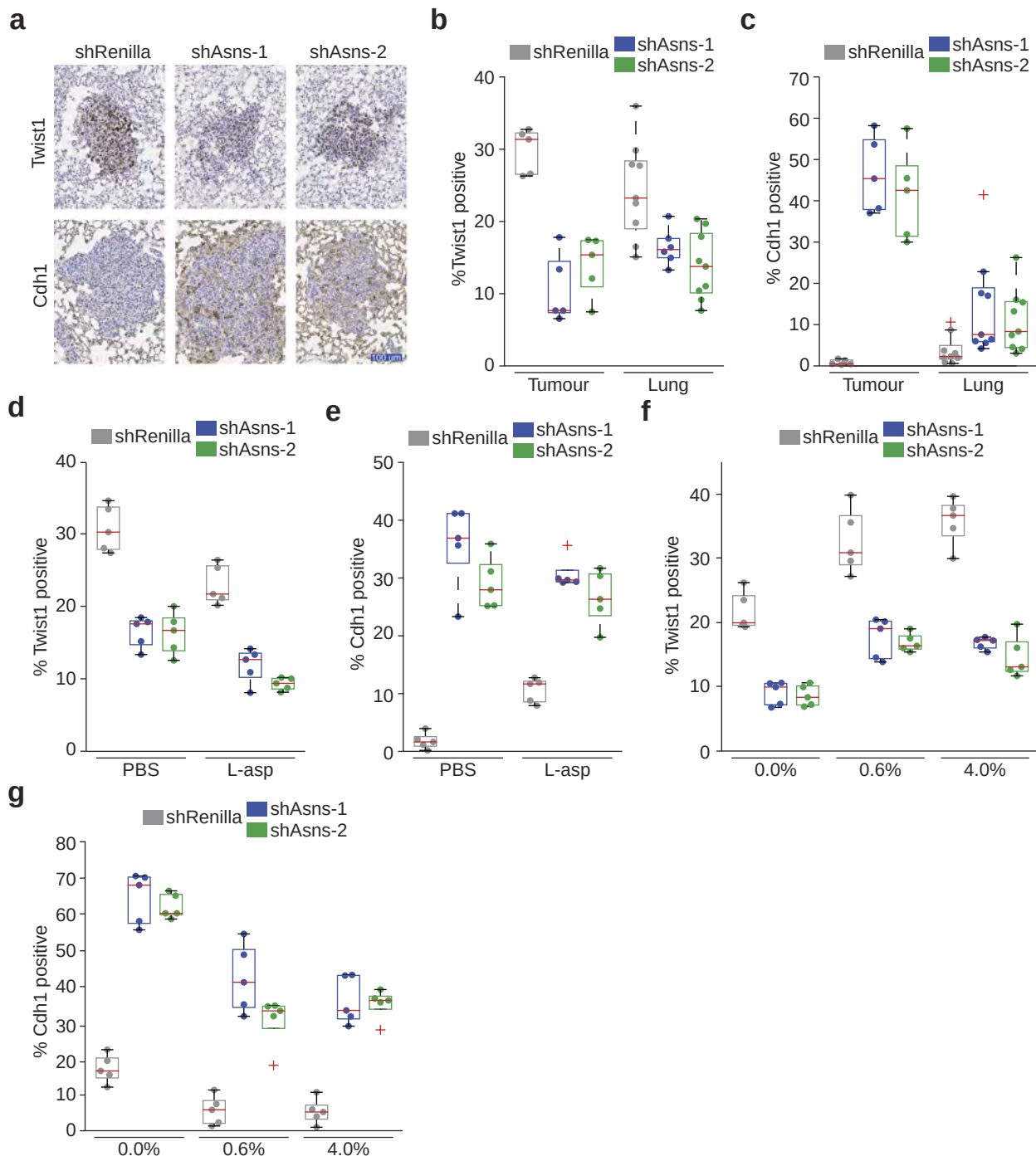
a, Transcription-level changes in EMT-up and -down genes that occur in response to Asns silencing in 4T1-T cells ($n = 2$ replicates per condition, edges of the box are the 25th and 75th percentiles and error bars extend to the values $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 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 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, 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 $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$, in which w is 1.5 and q_1 and q_3 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 Asns-expressing and -silenced cells into mice fed a 0%, 0.6%, or 4% asparagine diet ($n = 5$ tumour sections per condition, rank-sum $P < 0.01$ 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).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

Sample size was chosen to give sufficient power for calling significance with standard statistical tests. Mouse experiments were performed with 10 replicates to account for the variability that is observed in such in vivo experiments. In vitro experiments were performed with 5 replicates. In both cases standard rank-sum tests were able to identify significantly different observations.

2. Data exclusions

Describe any data exclusions.

No data was excluded.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All experiments have been performed several times in multiple cell lines and using multiple methods to perturb the in vivo environment in a specific manner.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

In vitro samples were split into different groups based on well locations in the culture dishes and matrigel chambers. In vivo groups were based on cage location, to simplify drug/food dosing by animal staff.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

All image quantification was performed in a blinded manner. Data analysis required prior knowledge of the sample annotation.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ A statement indicating how many times each experiment was replicated
- ☐ ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- ☐ ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- ☐ ☒ The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Standard high-throughput data pipelines were used for the initial processing of the high-throughput sequencing and proteomic data. All downstream analysis was performed using matlab. These custom codes will be made available to readers by the corresponding author on reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

There are no restrictions on availability.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

E-Cadherin (24E10) Rabbit mAb (3195, Cell Signaling) and Twist1 (Twist2C1a) Mouse mAb (ab50887, Abcam) were used for IHC analysis. These antibodies have been used previously in mouse tissues (E-cadherin: Nathan et al. The Wilms tumor protein Wt1 contributes to female fertility by regulating oviductal proteostasis; Twist1: Xu et al. Inducible Knockout of Twist1 in Young and Adult Mice Prolongs Hair Growth Cycle and Has Mild Effects on General Health, Supporting Twist1 as a Preferential Cancer Target).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

The 4T1 and MDA-MB-231 cell lines were obtained from ATCC.

b. Describe the method of cell line authentication used.

As cell lines were received from ATCC they will have been authenticated there.

c. Report whether the cell lines were tested for mycoplasma contamination.

Cell lines were tested for mycoplasma intermittently throughout the study.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

The cell lines used in this study are not listed on the ICLAC list of commonly misidentified cell lines.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Only NOD scid Gamma (NSG) mice were used in this study. Mice were housed in a biohazard (contamination free) room.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No human subjects were used in this study.

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

► Data presentation

For all flow cytometry data, confirm that:

- ☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ 3. All plots are contour plots with outliers or pseudocolor plots.
- ☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

► Methodological details

- | | |
|--|---|
| 5. Describe the sample preparation. | For mouse and human cell lines (4T1 and MDA-MB-231), cells were trypsinized and washed in PBS, prior to flow cytometry analysis. Mouse primary tumors and lungs were minced and digested into single cells with 1× collagenase/hyaluronidase buffer (StemCell) and 10 U DNase I (Sigma) in DMEM high glucose media for 1 h at 37 °C. Cells were then washed in PBS twice before sorting. |
| 6. Identify the instrument used for data collection. | BD LSRII and Sony SH800 for analysis and BD FACSAria III for sorting. |
| 7. Describe the software used to collect and analyze the flow cytometry data. | Data was collected with BD FACSDiva and Sony SH800 software. Flow cytometry data was analyzed using FlowJo (V.10) |
| 8. Describe the abundance of the relevant cell populations within post-sort fractions. | Post-sort fractions were >90% as determined by a post-sort purity check of one representative sample. |
| 9. Describe the gating strategy used. | For analysis, preliminary FSC/SSC gates were drawn using uninfected or unstained 4T1 or MDA-MB-231 cell lines. mCherry positive gates were set at least one log fold away from the unstained/uninfected gates. For CellTrace violet, the intensity was measured among all viable cells. For tumors and lungs, preliminary FSC/SSC gates were drawn using uninfected and mCherry-infected 4T1 cells. mCherry positive gates were set at least one log fold away from the unstained/uninfected gates. |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☐

ERRATUM

doi:10.1038/nature26162

Erratum: Asparagine bioavailability governs metastasis in a model of breast cancer

Simon R. V. Knott, Elvin Wagenblast, Showkhin Khan, Sun Y. Kim, Mar Soto, Michel Wagner, Marc-Olivier Turgeon, Lisa Fish, Nicolas Erard, Annika L. Gable, Ashley R. Maceli, Steffen Dickopf, Evangelia K. Papachristou, Clive S. D'Santos, Lisa A. Carey, John E. Wilkinson, J. Chuck Harrell, Charles M. Perou, Hani Goodarzi, George Poulogiannis & Gregory J. Hannon

Nature **554**, 378–381 (2018); doi:10.1038/nature25465

In this Letter, several errors were inadvertently introduced during the production process. In Fig. 3d, the blue bars should be labelled 'L-asparaginase' rather than 'L-asparagine'. In the main text, "These were collected onto Matrigel..." should be "These were placed onto Matrigel...", and in the Methods, 'corresponding proteins' should be 'corresponding peptides' in the text: "...proteins were removed from subsequent analysis if fewer than 5 corresponding peptides were identified in any sample." The legend to Extended Data Fig. 6b should state 'Volumes of orthotopic..' rather 'Volumes of orthotropic...'. Finally, in the HTML the legend to Supplementary Table 4 was repeated for Supplementary Table 5. The legend to Supplementary Table 5 should have stated: 'Amino acid composition of serum with and without L-asparaginase treatment. 4T1-T cells harbouring the non-targeting *Renilla* shRNA were injected into immunocompromised mice. Five mice each were either injected with 60 U L-asparaginase or PBS 5 days per week. After blood collection and serum isolation, free amino acids were quantified using High Performance Liquid Chromatography (HPLC) and a fluorometric detector.' These errors have all been corrected online.