nature research

Corresponding author(s): Gregory Hannon

☐ Initial submission ☐ Revised version ☒ Final submission

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 recieved 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.

nature research

Corresponding author(s): Gregory Hannon

☐ Initial submission ☐ Revised version ☒ Final submission

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. ☐