nature research

Corresponding author(s): Trey Ideker

☒ 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.

We chose to do triplicates for the flow cytometry experiments (Fig 4c, d. and Sup. Fig. 3). Each triplicate consists of 50,000 events counted through the flow counter and the median of these events is reported. The sample size is sufficient to observe a very significant difference in GFP intensity (two-tailed t-test $p < 0.001$).
In the cell migration assay, 74 and 83 cells were imaged in the two groups (Fig. 4h). The sample size is sufficient to observe a significant difference in invasion distance (Mann–Whitney U test $p = 0.01$).

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analysis

3. Replication

Describe whether the experimental findings were reliably reproduced.

Flow cytometry analyses for U2OS cells were performed three times in duplicates, as well as one experiment in triplicates as shown in Fig. 4c, d. All four attempts showed the same trend. Flow cytometry analyses for MDA-MB-231 cells were performed twice in duplicates, and once in triplicates as shown in Sup. Fig. 3. Two out of the three attempts showed the same trend as Sup. Fig. 3.
Cell migration assays (Fig. 4f-h) were performed two times and both attempts showed the same trend.

4. Randomization

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

Not applicable as no human or animal participants were involved in this study.

5. Blinding

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

Not applicable as no human or animal participants were involved in this study.

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.

Custom code are deposited at https://github.com/wzhang1984/Noncoding-tumor-mutation-paper, and are available to editors and reviewers upon request. They will be publicly available before publication. Other than that, we used bedtools v2.23.0 to cluster the mutations, R v3.2.5 for statistical learning, probabilistic estimation of expression residuals (PEER) v1.3 to identify hidden factors, HOMER v4.8.2 for motif analysis, FlowJo v10.2 for flow cytometry data analysis, and Cytoscape v3.3.0 for network visualization.

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 the availability of the materials used for this project except for the GFP reporter constructs used for the DAAM1 promoter activity assays, which will be available upon publication.

9. Antibodies

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

DAAM1 overexpression was verified by extracting total protein and quantitating using the Wes electropherogram (Proteinsimple) with anti-DAAM1 antibody (Santa Cruz, WW-3, 1:250 dilution, mouse) and anti-tubulin antibody (Millipore, 1:250 dilution, rat).

10. Eukaryotic cell lines

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

U2OS were acquired from ATCC atcc htb-96) osteosarcoma.
HT1080 cells were acquired from ATCC and the MDA-MB-231 cells were acquired from the PSOC network.

b. Describe the method of cell line authentication used.

U2OS cell genomic DNA was submitted for STR characterization to IDEXX BioResearch.

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

All cell lines are tested for mycoplasma when received or 48-72 hours post thawing from cryostorage.

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.

None of the cell lines used are listed in ICLAC's v8 records.

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.

No animals were used

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.

The study does not involve human research participants

nature research

Corresponding author(s): Trey Ideker

☒ 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.	U2OS (Fig 4) or MDA-MB-231 (Sup Fig 3) cells were transfected and then harvested 2 days later for FACS.
6. Identify the instrument used for data collection.	BD FACSCalibur
7. Describe the software used to collect and analyze the flow cytometry data.	FlowJo v10.2
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Not applicable (no sorting performed)
9. Describe the gating strategy used.	A "live" gate based on FSC v SSC was applied. GFP+ cells in the "live" gate were determined based on GFP v SSC with the gates shown in the figures.

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