

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

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

Data collection

For targeted metabolomics peak areas from the total ion current for each metabolite SRM transition were integrated using MultiQuant v2.0 software (AB/SCIEX). For qPCR, QuantStudio Real Time PCR software v1.3 (Life Technologies) was utilized. For xenografted cells in mice luminescence was measured and quantified using the Living Image Software v4.5 (Perkin-Elmer). For incucyte proliferation assays, Incucyte S3 live cell imaging system was utilized to take images of the cells on a schedule and then the Incucyte software (v. 2021A) was used to analyze the images to determine confluency.

Data analysis

All data analyses are explained in detail in the methods. General data analyses were performed using Microsoft Excel (2013 or 365) and GraphPad Prism (7 or 9). For targeted metabolomics peak areas from the total ion current for each metabolite SRM transition were integrated using MultiQuant v2.0 software (AB/SCIEX) and statistical analysis of the data was carried out using MetaboAnalyst v4.0, a free online software for the analysis of metabolomic experiments (www.metaboanalyst.ca). For quantification of invasion/migration assays image quantifications were carried out in an automated way on Fiji/ImageJ (v1.52 or v1.53n) using a custom macro script. This macro is a basic automation script and cannot be used as a stand alone code but it is available from the corresponding authors upon reasonable requests. Briefly, binary images of the area covered by crystal violet-positive cells was generated using thresholding and settings that were appropriate for control samples, and these settings were used throughout the analysis. The percentage area covered by crystal violet-positive cells was quantified for each condition, using a minimum of two technical replicates. For RNA-seq analysis: The TopHat algorithm v2.1.0 (Bowtie v2.2.6.0) was used to align the reads to the hg38 genome, which was obtained through iGenomes (https://support.illumina.com/sequencing/sequencing_software/igenome.html). The alignments (37.5-51.4 M aligned pairs) in the BAM files were further analyzed using the Cufflinks suite of programs v2.2.1 (running consecutively: Cufflinks v2.2.1.Linux_x86_64, Cuffcompare v2.2.1, Cuffdiff v2.2.2). Gene set enrichment analysis was performed on GSEA v4.0.3 using Gene Ontology (GO)-Biological Processes (BP) gene set database from the Molecular Signatures Database (MSigDB) collection v6.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Source data information for the metabolomics experiment can be found on Supplementary Table 1. RNA sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under the accession code GSE161108, as well as summary information in Supplementary Table 2. The raw data supporting each figure and the raw images for the western blots can be found in the corresponding Source Data files.

For RNA-seq analysis, the hg38 reference genome database was obtained from iGenomes (https://support.illumina.com/sequencing/sequencing_software/igenome.html) and the GSEA analysis was done with gene sets derived from the GO biological processes gene sets in the Molecular Signatures Database (MSigDB) collection v6.2, which can be accessed through <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were done. For in vivo studies, sample sizes were determined based on our previous experiments. In our experience, n = 7-10 mice per group is sufficient to detect meaningful biological differences with good reproducibility. For the mouse experiments n refers to biologically independent samples as individual mice. The vast majority of the experiments outlined in this manuscript utilize cell-based systems, where conditions are standard and extreme variability is decreased as a result. Thus, we utilized at least 3 independent samples for each nominal variable. This number of independent experiments is a standard sample size to accurately detect differences in cell biology and molecular biology experiments (such as Napolitano et al. Nature 2020. https://doi.org/10.1038/s41586-020-2444-0.) Western blots are shown as representative images, however they were reproduced with at least 3 independent samples.
Data exclusions	No data were excluded.
Replication	Experiments were repeated multiple times with similar results as indicated in the figure legends. All attempts to replicate the data were successful. The core experiments were replicated by at least two people independently and in multiple cell lines to insure reproducibility and validity of the experiments.
Randomization	Mice obtained from the vendor were randomly split into experimental groups before administering xenografts. For cell culture experiments plates/wells for all treatment groups were set up from a single pool of cells. The plates/wells were randomly assigned to treatments. In case of genetic modification and treatments, a single pool of cells were utilized and randomly assigned for the transductions. Once infected, the genetically modified cells (e.g: control vs shRNA's) were maintained in parallel (same culture conditions/timing, same selection conditions, same trypsinization schedule, same seeding numbers, etc). For further treatments each pool of genetically modified cells were used as a single pool for setting up plates/wells for treatments and treatments were randomly assigned to these plates/wells.
Blinding	Blinding was not performed for the in vitro and in vivo experiments. The investigators needed to know the treatment groups in order to perform the study, and the data analyses were based on objectively measurable data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

The antibodies used to detect the proteins of interest were: E-Cadherin (610181-BD Biosciences), Fibronectin (ab2413-Abcam), Vimentin (5741S-Cell Signaling), Serpine1 (612024-BD Biosciences), CTGF (ab6992-Abcam), CXCR4 (ab124824-Abcam), MMP2 (4022S-Cell Signaling), HA (sc-7392-Santa Cruz), EGR1 (ab133695-Abcam), SP1 pS739 (ab195733-Abcam), SP1 (39058-Active Motif), ERK1/2 (9102L-Cell signaling), ERK1/2 pT202/pY204 (M9692-Cell Signaling), MMP9 (3852S-Cell Signaling), MCEE (19499-1-AP-Proteintech), MMAB (ab174831-Abcam), MUT (ab134956-Abcam), PCCA (ab18768-Abcam), PCCB (HPA036940-Sigma-Aldrich) and Actin (sc1615-Santa Cruz). As secondary antibodies horseradish peroxidase-conjugated (HRP) anti-rabbit (NA934-Cytiva), anti-mouse (NA931-Cytiva), or anti-goat (AP180P-Millipore) antibodies were used.

Validation

All antibodies used are commercially available and were tested by the manufacturer. They are standard tools used in the field, and have been previously validated and characterized by multiple labs. More information about each antibody can be found on manufacturers' websites.

E-Cadherin (610181 - BD Biosciences) : <https://www.bdbiosciences.com/us/applications/research/stem-cell-research/cancer-research/human/purified-mouse-anti-e-cadherin-36e-cadherin/p/610181>
 Fibronectin (ab2413 - Abcam): <https://www.abcam.com/fibronectin-antibody-ab2413.html>
 Vimentin (5741S - Cell Signaling): <https://www.cellsignal.com/products/primary-antibodies/vimentin-d21h3-xp-rabbit-mab/5741>
 Serpine1 (612024 - BD Biosciences): <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/cell-biology-reagents/cell-biology-antibodies/purified-mouse-anti-pai-1-41pai-1/p/612024>
 CTGF (ab6992 - Abcam): <https://www.abcam.com/ctgf-antibody-ab6992.html>
 CXCR4 (ab124824-Abcam): <https://www.abcam.com/cxcr4-antibody-umb2-ab124824.html>
 MMP2 (4022S - Cell Signaling): <https://www.cellsignal.com/products/primary-antibodies/mmp-2-antibody/4022>
 HA (sc-7392-Santa Cruz): <https://www.scbt.com/p/ha-probe-antibody-f-7>
 EGR1 (ab133695-Abcam): <https://www.abcam.com/egr1-antibody-epr50142-ab133695.html>
 SP1 pS739 (ab195733-Abcam): <https://www.abcam.com/ab195733.pdf?>
 SP1 (39058-Active Motif): <https://www.activemotif.com/catalog/details/39058/sp1-antibody-pab>
 ERK1/2 (9102L-Cell signaling): <https://www.cellsignal.com/products/primary-antibodies/p44-42-mapk-erk1-2-antibody/9102>
 ERK1/2 pT202/pY204 (4370-Cell Signaling): <https://www.cellsignal.com/products/primary-antibodies/phospho-p44-42-mapk-erk1-2-thr202-tyr204-d13-14-4e-xp-rabbit-mab/4370>
 MMP9 (3852S-Cell Signaling): <https://www.cellsignal.com/products/primary-antibodies/mmp-9-antibody/3852>
 MCEE (19499-1-AP-Proteintech): <https://www.ptglab.com/products/MCEE-Antibody-16499-1-AP.htm>
 MMAB (ab174831-Abcam): <https://www.abcam.com/mmab-antibody-epr8695-ab174831.html>
 MUT (ab134956-Abcam): <https://www.abcam.com/methylmalonyl-coenzyme-a-mutase-antibody-epr7738-ab134956.html>
 PCCA (ab187686-Abcam): <https://www.abcam.com/pcca-antibody-c-terminal-ab187686.html>
 PCCB (HPA036940-Sigma-Aldrich): <https://www.sigmaaldrich.com/catalog/product/sigma/hpa036940?lang=en®ion=US>
 Actin (sc1615-Santa Cruz): <https://www.scbt.com/p/actin-antibody-c-11>
 anti-rabbit HRP secondary (NA934-Cytiva): <https://www.cytivalifesciences.com/en/us/shop/protein-analysis/blotting-and-detection/blotting-standards-and-reagents/amersham-ecl-hrp-conjugated-antibodies-p-06260>
 anti-mouse HRP secondary (NA931-Cytiva): <https://www.cytivalifesciences.com/en/us/shop/protein-analysis/blotting-and-detection/blotting-standards-and-reagents/amersham-ecl-hrp-conjugated-antibodies-p-06260>
 anti-goat HRP secondary (AP180P-Millipore): https://www.emdmillipore.com/US/en/product/Donkey-Anti-Goat-IgG-Antibody-HRP-conjugate-Species-Adsorbed,MM_NF-AP180P

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

MCF-10A, BT-20, MDA-MB-436, HCC38, MDA-MB-231 SKBR3, T47D, BT-549, MCF-7, A549, HCC1806, Hs578T were bought from ATCC. Hek293T was bought from GenHunter. MDA-MB-231-luciferase cells described previously (Minn et al., 2005) were generated from the MDA-MB-231 parental cell line (ATCC) in Dr. Massague's lab and obtained from Memorial Sloan Kettering Cancer Center Monoclonal Antibody Core facility. Mouse breast cancer cell line clones 4T1, the metastatic one, and the locally invasive clone 4T07 were originally derived by Dr. F. Miller (Miller et al., 1983) and obtained from Dr. William Schiemann.

Authentication

ATCC utilizes multiple methods - morphology, karyotyping, and PCR based approaches - to authenticate the cell lines it maintains. No further authentication was performed by the authors of this manuscript.

Mycoplasma contamination

All cell lines were routinely tested for mycoplasma and were at all times mycoplasma negative.

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified lines were used in this study.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Female nu/nu athymic mice were purchased from Envigo at the age of 4-6 weeks. Xenograft experiments were started 7-10 days after the mice were received. The mice were maintained at Weill Cornell Medicine in compliance to Weill Cornell Medicine Institutional Animal Care and Use Committee protocols. 6-week-old female BALB/c mice were obtained from Envigo and used for the syngeneic tumor model. The BALB/c mice were maintained at KU Leuven/VIB in compliance with local ethical regulations and was approved by the KU Leuven ethics committee.

Wild animals

No wild animals were used in this study.

Field-collected samples

No field collected samples were used in this study.

Ethics oversight

The nu/nu mice were maintained at Weill Cornell Medicine in compliance to Weill Cornell Medicine Institutional Animal Care and Use Committee protocols. BALB/c mice were maintained at KU Leuven/VIB in compliance with local ethical regulations and was approved by the KU Leuven ethics committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.