nature research

Corresponding author(s): Yibin Kang

Last updated by author(s): Mar 4, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	Data was collected using standard software programs including: Nikon Elements for microscopy, FlowJo v10 and FACSDiva for flow cytometry, Licor Odyssey Clx for western blot, ImageJ for bone metastasis x-rays, LivingImage for bioluminescence, Skyline for Mass spectrometry, Genespring v13 for microarray, and GSEA-BroadInstitute for GSEA analysis. These have mentioned in the text where its reporting is customarily used. Stata and Excel were used for basic data display and statistical tests. No code was used to collect any data.
Data analysis	Excel functions ttest and stdev used for Student's T-test and standard deviation calculations. Stata ranksum used for Mann-Whitney U test and stcox used for Cox's Proportional Hazards model. Vioplot add-in used for stata graphing of violin plots. GSEA pre-ranked GSEA used for GSEA analysis.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Microarray data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession codes GSE96754, and the mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012942; mass spectrometry data is further available in Supplementary Table 2.
The human clinical breast cancer data are not available in GEO; these were derived from the TCGA Research Network and the data-set derived from this resource that supports the findings of this study is available in https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Breast%20Cancer%20(BRCA)
The human breast cancer data were also derived from the NKI-295 dataset which is available at https://xenabrowser.net/datapages/?cohort=Breast%20Cancer%

20(Vijver%202002)
Finally, human breast cancer data from the EMC-MSK dataset is available in Bos et. al, 2009. Source data relevant for the clinical data analyses performed in this study are available in Supplementary Table 9.

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

Life sciences study design

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

Sample size	For most experiments, biological replicates and/or technical repeats were included and most experiments were repeated at least 3 times. Such sample sizes are typical of the in vitro experiments (e.g. qRT-PCR) used in the study. For in vivo experiments, sample size of n = 8-10 are typical of experiments and are used in the study, but this number varies according to the number of wild-type and knockout mice per litter in the E-selectin knockout experiment or when the number of mice in a single experiment was >40 animals. No sample size calculation was performed to predetermine sample size. All sample sizes are listed in the corresponding figure legend.
Data exclusions	Animals were excluded only if they died or had to be sacrificed because of moribund conditions specified in the IACUC protocol. Such events occurred in less than 5% of animals during the experimental time frame. This exclusion criteria was pre-determined by the protocol.
Replication	Immunofluorescence, western blot, mass spectrometry controls, flow cytometry and E-selectin plate-binding assays were conducted ≥3 times to ensure reproducibility unless otherwise noted. All in vivo bone metastasis experiments were performed at least 2 times to ensure reproducibility and significance with the exception of the BM2-CD44 KO injection, catalytic Fut3 mutants and the GMI-1271 treatment, which was performed once. All lung metastasis and primary tumor growth experiments were performed once with appropriate group size and statistics. All attempts at replication were successful.
Randomization	Standard mice (Nu/Nu or NSG) for xenograft experiments were randomized, with each group receiving an equal number of mice from each litter. E-selectin knockout compared to wild-type mice were not randomized, rather, every available mouse within 20 days of the same date of birth was used.
Blinding	No blinding was performed with the exception of x-ray analysis and quantification. In this case, the investigator was blinded to the experimental group of each x-ray. For all other cases, blinding was not relevant to the experiment as the results were directly collected by instrumentation.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*

Randomization | *If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used

Glg1 (Sigma HPA010815, 1:1000-IB or 1:100 IF), CD44 (Cell Signaling Technologies clone 156-3C11, 1:1000), Beta-actin (Santa Cruz sc47778, 1:1000), N-cadherin (BD Pharmigen 610920, clone 32, 1:1000-IB, 1:100 IF), FLAG (Sigma F7425, 1:1000), Fut3 (Abcam ab110082, 1:1000), E-selectin (BD Pharmigen clone 109.E6, 1:100), EpCam (CST VU1D9, 1:100), E-cadherin (BD Pharmigen 610181, clone 10E9.6, 1:100), Ki67 (Dako M7240, clone MIB-1, 1:40), ZO-1 (CST 5406, 1:100), EpCAM (CST VU1D9, 1:100), K14 (Biolegend 905301, 1:100), Slug (Santa Cruz clone A-7, 1:50), E-selectin (R&D systems BBA18, 1:50), Glg1 (Personal gift, Charles Dimitroff, 1:1000). This data is available in Supplementary Table 8.
All secondary antibodies used were Licor IRDye 680/800 anti-rabbit or anti-mouse for western blot (925-322#, 1:10000), Dako polyclonal anti-rabbit or anti-mouse-HRP for western blot (P0448, 1:2000) or Invitrogen AL555 or AL647 anti-rabbit or anti-mouse (e.g. A-21244, 1:1000) for immunofluorescence.

Validation

All commercial antibodies were validated by the manufacturer for the species and application used in this study.
Glg1, CD44, Fut3, and FLAG antibodies were validated for western blot via specific knockout or overexpression followed by western blotting- data for each is displayed in the manuscript. All bands appeared at the correct apparent molecular weight.
Beta-actin, Slug, CD31, EpCam, E-cadherin, K-14, Ki-67 antibodies were previously validated in our lab for the intended species and application used here.
E-selectin and Glg1 antibodies were validated for IF by overexpression in HEK-293T cells followed by indirect immunofluorescence.
In addition, all antibodies used here have been validated by reviews on Biocompare.
anti-Glg1 from Charles Dimitroff has been previously validated by Robert Sackstein's laboratory and was furthermore validated in this study by specific knockout and overexpression of Glg1 followed by western blotting in Supplementary Fig. 5.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Key characteristics of each cell line, including source, are included in Supplementary table 1. The MDA-MB-231 cells and its derivatives (SCP6, SCP32, SCP28, SCP2, SCP25, LM2, BM2) were from Joan Massague's lab. SUM159 was from Stephen Ethier's lab at the Karmanos Institute and the derivatives were made by this study's lead author. DU145 was obtained from ATCC and the ob1 bone-metastatic derivatives were generated by this study's lead author. L Cells were obtained from ATCC. PCR1 cell line was obtained from Steven Barthel.

Authentication

All cell lines were authenticated by STR profiling with a similarity score >90%.

Mycoplasma contamination

All cells were tested for mycoplasma infection either upon derivation and freezing or after 1 month of culture via PCR testing. All cultures were confirmed negative and no instances of mycoplasma contamination occurred throughout the study period.

Commonly misidentified lines
(See ICLAC register)

No cell lines used in this study were found in the database of commonly misidentified cell lines that is maintained by ICLAC and NCBI Biosample.

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

E-selectin knockout mice were ordered from the Jackson Laboratory (B6.129S4-Seletm1/DmiL) and were backcrossed for 5 generations into the NOD/SCID strain. All xenograft experiments were conducted on 8 week old female mice (athymic Nu/Nu, NOD/SCID, NOD/SCID Gamma) except in prostate cancer xenografts, which were performed on male mice of the same age. All mice were originally ordered from the Jackson Laboratory and breeding was conducted in an SPF barrier facility.

Wild animals

The study did not use wild animals.

Field-collected samples

The study did not use field-collected samples.

Ethics oversight

All procedures involving mice and experimental protocols were approved by the University Institutional Animal Care and Use Committee (IACUC).

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 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).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For E-selectin binding experiments, RFP- and GFP-labeled cells were co-cultured at equal ratios for 48 h, then cells were harvested at 80% confluence by non-enzymatic dissociation buffer (Life Sciences) at 37°C, washed once with PBS, and suspended in flow buffer (10% FBS in PBS supplemented with 1 mM Ca2+/Mg2+) at 1 million cells/mL. Recombinant mouse E-selectin/ IgG Fc or Isotype IgG (RD systems) was added at 10 ug/mL for 1.5 h with vortexing every 15 min. Cells were washed with PBS and incubated with anti-human IgG-AL647-conjugated antibody (Biolegend) at 0.1ug/mL for 45 min. Cells were washed with PBS, resuspended in flow buffer with DAPI (1 ug/mL), and analyzed with the BD LSRII flow cytometer. Internal control cell lines (RFP-labeled) corresponded to the parental population of each cell line – either MDA-MB-231-RFP or SUM159-RFP. E-selectin binding ratios were quantified using the formula (SELEGFP/IgGGFP)/(SELERFP/IgGRFP). Gating for DAPI, RFP, and GFP were performed using negative controls of the corresponding cell line using FlowJo version X. For pharmacological, cells were grown in Tunicamycin, Deoxymannojirimycin, or PDMP for 48h. Following differential treatment, internal control RFP-labeled cells were added to the cell mixture. For FACS sorting, MDA-MB-231 cells were labeled using this protocol and were sorted into the top and bottom 10% of binding intensities. E-selectin-sorted MDA-MB-231 cells were passaged 5 times prior to assessing E-selectin binding levels to generate a large enough pool both for flow cytometry and sub-culturing. Negative control treatments (EDTA or GMI-1271) were tested in every cell line to ensure that E-selectin binding was specific. For Wnt signaling activation analysis, BM2 cells stably transduced with the 7x-TCF-GFP-SV40-mCherry (TGC) reporter were trypsinized, washed once with PBS, and suspended in PBS + 1 ug/mL DAPI. Cells were analyzed with the LSRII instrument, and gating analysis was performed using unlabeled BM2 cells using the FlowJo version X software. All flow cytometry experiments were repeated a minimum of three times.
Instrument	BD LSRII
Software	FACSDiva and FlowJo Version X were used to set gates and collect data.
Cell population abundance	No cell sorting was performed beyond isolating the top and bottom deciles for the E-selectin-sorted cell lines.
Gating strategy	FSC and SSC gates were set to include at least 90% of events and to exclude very small events. SSC-A and -H or FSC-A and -H were used to discriminate for only single cells. DAPI or PI were used to discriminate live cells from dead using appropriate negative and positive controls. GFP and RFP gating was performed using negative controls from the respective cell lines tested, or from using appropriate internal controls (I.e. the Sore6 reporter system which used a control plasmid to control for endogenous mCherry transcription) such that the negative control sample showed no more than 1% positivity. All binding analyses were performed using an internal control to control for differences in cell culture, harvesting or probing conditions. This internal control was used to normalize all experiments as described.

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

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) — *Specify: functional, structural, diffusion, perfusion.*

Field strength — *Specify in Tesla*

Sequence & imaging parameters — *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition — *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software — *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization — *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template — *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal — *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring — *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings — *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested — *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) — *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction — *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity — *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis — *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis — *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*