nature portfolio

Corresponding author(s): Ana Rita Nobre and Julio A. Aguirre-Ghiso

Last updated by author(s): Jun 27, 2022

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	No software was used for data collection.
Data analysis	Basepair, Enrichr and GSEA were used for data analysis. Detailed information about the analysis can be found in the Methods section.

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

The datasets generated during and/or analyzed during the current study are available within the paper, extended data and Source data files. All sequencing data is available in a public data repository. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.
GSE165431, RNA sequencing of MMTV-Neu early lesion (EL) and primary tumor (PT) spheres
GSE165444, ZFP281 ChIP-sequencing of MMTV-Neu early lesion (EL) and primary tumor (PT) spheres

GSE165456, scRNAseq of MMTV-Neu primary site and lung cancer cells in early and late stage
GSE165459, scRNAseq of MMTV-Neu lung cancer cells in early and late stage.
GSE81044, RNAseq data from primed versus naïve mouse pluripotent stem cells
GSE93042, ZFP281 ChIPseq from primed mEpiSCs

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Samples were collected from the Cancer Biorepository at Icahn School of Medicine at Mount Sinai, New York, New York. 28 samples were analyzed, 14 carcinoma in situ (DCIS), 14 invasive breast cancer (IBC).
Population characteristics	Women with DCIS or IBC diagnosed, treated and with samples available in the Cancer Biorepository at Icahn School of Medicine at Mount Sinai, New York.
Recruitment	No recruitment of participants was needed. Samples were collected from a biorepository.
Ethics oversight	Samples were de-identified and obtained with Institutional Review Board approval (Program for the Protection of Human Subjects / Institutional Review Board of Icahn School of Medicine at Mount Sinai), which indicated that this work does not meet the definition of human subject research according to the 45 CFR 46 and the Office of Human Subject Research.

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

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	No statistical methods were used to pre-determine sample sizes, sample sizes were chosen empirically.
Data exclusions	No exclusion criteria were applied.
Replication	Unless otherwise specified, 3 or more independent experiments were performed, in vitro with at least 2 technical replicates per condition; and 2 or more independent experiments were performed in vivo with at least 5 mice per condition.
Randomization	Block randomization was used for all mice assignments.
Blinding	The investigators were not blinded to allocation during experiments but all quantifications were done in coded samples to reduce operator bias.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used

PE anti-mouse HER2, Novus Biologicals, NBP2-34641PE, FACS: 1:100, 15min at 4°C;
APC anti-mouse CD45, Biolegend, 103112, FACS: 1:100, 15min at 4°C;
PerCP/Cyanine5.5 anti-mouse CD326 (Epcam) Biolegend, 118220, FACS: 1:100, 15min at 4°C;
PE/Cyanine7 anti-mouse CD105 (Eng) Biolegend, 120410, FACS: 1:100, 15min at 4°C;
HER2, Millipore, OP15L, IF: 1:100, ON at 4°C;
HER2, Abcam, ab214275, IF: 1:100, ON at 4°C;
ZFP281 Santa Cruz, sc-166933, IF: 1:50, ON at 4°C;
Ki67, Invitrogen, 14-5698-82, IF: 1:100, ON at 4°C;
E-cadherin, BD Biosciences, 610182, IF: 1:100, ON at 4°C;
Twist1, Millipore, ABD29, IF: 1:50, ON at 4°C;
Cdh11, Invitrogen, 71-7600, IF: 1:100, ON at 4°C;
GFP, Aves, GFP-1020, IF: 1:500, ON at 4°C.

Validation

All antibodies used are commercially available and tested by manufactures. All antibody staining were done in parallel with negative controls to guarantee specificity.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

Only primary cell cultures were used in this study. Cells were isolated from mice for each experiment and kept in culture no more than 2 weeks.

Authentication

No cell line authentication was performed.

Mycoplasma contamination

No mycoplasma tests were performed.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell line was used.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

MMTV-HER2/Neu mice were maintained on FvB background and bred and crossed in our facilities. 14 to 18-week-old female mice were used as early ('pre-malignant') stage mice and 20-week-old or older females with palpable tumor(s) were used as late stage of cancer progression. MMTV-PyMT mice on C57/BL6 background were purchased from the Jackson Laboratory. 4 to 6-week-old female mice were used as early stage mice and 8-week-old or older females with palpable tumor(s) were used as late stage. BALB/c nu/nu mice were purchased from Charles River. No randomization or blinding was used to allocate experimental groups. Tumors were not allowed to grow beyond the IACUC allowed limit of 1 cm3 per animal.

Wild animals

No wild animals were used in this study.

Reporting on sex

Only female mice were used in this study.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

All experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Icahn School of Medicine at Mount Sinai.

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

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.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165444

Files in database submission

Raw (fastq) and processed (bed files) were submitted for all samples. GSM5033687 Ch IP Early lesion (EL) spheres - input
GSM5033688 Ch IP Early lesion (EL) spheres - replicate 1
GSM5033689 Ch IP Early lesion (EL) spheres - replicate 2
GSM5033690 Ch IP Early lesion (EL) spheres - replicate 3
GSM5033691 Ch IP Primary tumor (PT) spheres – input

	GSM5033692 Ch IP Primary tumor (PT) spheres - replicate 1 GSM5033693 Ch IP Primary tumor (PT) spheres - replicate 2 GSM5033694 Ch IP Primary tumor (PT) spheres - replicate 3
Genome browser session (e.g. UCSC)	No longer applicable.

Methodology

Replicates	One EL input, one PT input, three biological EL and three biological PT replicates ChIPed DNA were used to prepare ZFP281 ChIP-seq libraries.
Sequencing depth	Massively parallel sequencing was performed with the lllumina HiSeq4000 according to the manufactory protocol, and paired-end 150 bp-length reads were produced.
Antibodies	Ch IP was performed using EZ Ch IP protocol (Millipore, Massachusetts, USA) with Zfp281 antibody (ab101318, Abeam) for EL and PT samples (after 7 days iin cultures).
Peak calling parameters	ChIP-seq reads were aligned to the mmlO genome using bowtie2 (v2.3.4.3), followed by removing PCR duplicates using Picard with the parameter REMOVE_DUPLICATES=true. Ch IP-seq peaks were determined by the MACS program (v.2.1.2) using input Ch IP-seq as the control data, and all other parameters followed the default setting.
Data quality	Binding difference around the transcription start sites [-Skb, +Skb] between the EL and PT samples are analyzed using the DiffBind (v2.l.6).
Software	Mentioned above.

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	Mammary glands of 'early stage' mice, primary tumors from 'late stage' mice and lungs from 'early' and 'late stage' mice were dissected and digested (see 'Animal experiments' section). In case of cells in culture, single cell suspensions were obtained by incubating the cells in accutase (Sigma) for 20 minutes, at 37C. Cells were stained using antibodies and conditions described in the Antibodies section.
Instrument	All experiments were performed using BD FACSAria II sorter equipped with FACS Diva software (BD Biosciences) or analyzed using Aurora analyzer (Cytek Biosciences) equipped with SpectroFlo software.
Software	Data were analysed with FACS Diva (BD Biosciences) or FCS Express Cytometry 7 (De Novo) softwares.
Cell population abundance	HER2+ cells vary from 0.5-2% in lungs to 80-90% in primary tumors. Biological negative (corresponding FvB tissues) and positive (primary tumor) were used as fluorescent control and set the sorting gates. 4-way purity was used in all sorting experiments.
Gating strategy	Dead cells and debris were excluded by FCS, SSC and DAPI (4',6- diamino-2-phenylindole) (Fisher Scientific) staining profiles.

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