nature portfolio

Corresponding author(s): Dr. Eva González-Suárez

Last updated by author(s): 09/Jan/2026

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	RNA-seq: sequencing libraries were processed with the "QuantSeq 3'mRNA-Seq Library Prep Kit (FWD) for Illumina" (Lexogen, #015) with a UMI Second Strand Synthesis module. ATAC-seq: Sequenced reads obtained from NGS were demultiplexed and quality filtered using FastQC (Andrews, 2010). Subsequently, Bowtie 2 (Langmead et al., 2012) was used for paired-end read alignment against mouse (mm9) genome.
Data analysis	RNA seq Analysis: For the analysis of the RNA-seq reads, Nextpresso v2 was used, as follows: UMI sequences within reads were extracted with UMItools. The resulting FASTQ files were processed with bbduk (https://sourceforge.net/projects/bbmap/) for the extraction of adapters, polyA and rRNA sequences. Read quality was verified with FastQC v0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The derived FASTQ files were aligned to the mouse genome (GRCm39) with Hisat2. The obtained BAM files were sorted with SAMtools, and the UMItools were employed once again for read deduplication. Htseq-count was used to count the number of reads that aligned within genes, using the Gencode vM29 gene annotation. ComBat-seq was then used to adjust the batch effect. Differential expression was performed with DESeq2, keeping only those genes with more than 10 counts across the samples. Finally, those genes that had an adjusted p-value below 0.05 were selected. GSEAStandard was used to perform gene set enrichment analysis (GSEA) for the selected gene signatures, setting 1000 gene set permutations. Only those gene sets with significant enrichment levels (FDR < 0.25) were considered. A minimal and maximum number of genes per reference were defined as 10 and 500 respectively. GSEA was performed with gene signatures identified by previous literature at a single-cell level 18. Summary GSEA bubble heatmaps were generated with ggseabubble. ATAC-seq data Analysis:

Sequenced reads obtained from NGS were demultiplexed and quality filtered using FastQC (Andrews, 2010). Subsequently, Bowtie 2 was used for paired-end read alignment against mouse (mm9) genome. Signal tracks were generated with RPKM normalization, using a bin size of 50 bp. For quantification of ATAC-seq data, differential signal analysis was performed at ENSEMBL annotated genes through the integration of R libraries, including 'csaw', 'edgeR', 'rtracklayer', and 'statmod'. Bam files were used as input with the specific parameters: window width = 150, spacing = 50 and minq = 40. 'normFactors', 'asDGEList', and 'estimateDisp' functions were used for normalization and differential binding analysis using CPM values. Subsequently, the 'glmQLFit' and 'glmLRT' functions were used for comparative tests.
Visualization of signal tracks was performed using the UCSC Genome Browser, with mean values and a smoothing window between 6-12 pixels. MACS2 was used for calling broad peaks from ATAC-seq experiments. Peaks were result of common peaks between replicates. Differentially enriched motifs were identified using the Sequence Enrichment Analysis (SEA) tool based on the unique peaks in each comparison. The SEA tool (v5.5.5) was run using default parameters comparing sequences from ATAC-seq peaks against control sequences (shuffled input sequences). Known motifs from the JASPAR CORE 2022 vertebrate database were scanned. Enrich TF criteria (enrichment > 1, pval ≤ 0.05).

Proteomic and Phosphoproteomic Data Analysis
Raw files were processed using MaxQuant (v2.1.4.0) with TMTpro 18-plex quantification enabled. Database searches were performed against the UniProtKB/TrEMBL mouse reference proteome (21,990 sequences; downloaded March 1, 2023) supplemented with a curated contaminant list routinely used in the CNIO Proteomics Unit. Carbamidomethylation of cysteine and TMTpro (+304.2071 Da) on peptide N-termini and lysine residues were set as fixed modifications. Oxidation of methionine, N-terminal acetylation and deamidation (N/Q) were set as variable modifications. A minimum peptide length of seven amino acids and up to two missed cleavages were allowed. Peptide-spectrum matches, proteins and phosphosites were filtered at a 1% false discovery rate (FDR). TMTpro labelling efficiency and potential over-labelling were assessed via dedicated MaxQuant and Proteome Discoverer searches of pooled TMT samples. Labelling efficiency exceeded 98%, with <1% of total precursor intensity corresponding to unlabelled peptides. Over-labelling at serine, threonine and tyrosine residues accounted for 4.99% of total precursor intensity (6.46% of PSMs), well below reported values for TMTpro workflows, indicating minimal impact on quantitative accuracy. Reporter ion intensities were $\log_2$-transformed and processed using Prostar (v1.30.0) and DAPAR (v1.22.3). LOESS normalization was applied independently to the proteome and phosphoproteome datasets. Entries with fewer than 18 valid reporter values were excluded. Together, these metrics confirm that TMTpro labelling was highly efficient with minimal unintended STY modifications, supporting the robustness of our quantitative proteomics workflow.
Because TMT-based quantification at the MS2 level (without FAIMS) is prone to reporter ion interference and ratio compression, and therefore not optimized for estimating small effect sizes at the individual-protein level 68, we did not apply fixed fold-change cutoffs or multiple-testing–adjusted significance thresholds to define sets of differentially expressed proteins. Although per-protein statistical scores (e.g., p-values) are reported for visualization (e.g., volcano plots), downstream analyses were based exclusively on rank-based approaches that capture coordinated pathway-level trends. For the proteome, proteins were ranked solely by $\log_2$ fold change (K14-Rank vs K14-Ctrl) and used as input for gene set enrichment analysis (GSEA; v4.2.3) with the mouse msigdb_v2022.1.Mm collection (GMT format). For phosphoproteome-level analysis, GSEA was performed independently of protein abundance changes. Quantified phosphosites were first collapsed to a single value per gene to enable gene-centric enrichment analysis. For each gene, the phosphosite exhibiting the maximum absolute $\log_2$ fold change across conditions was selected, and the signed $\log_2$ fold change of that site was used to rank genes. This strategy captures the dominant phosphorylation change per protein while avoiding dilution or cancellation effects arising from phosphosites with opposing regulation on the same protein.

Protein-level fold changes were not used to normalize, weight, or adjust phosphoproteomic rankings, even when the corresponding protein was detected in the global proteome dataset. Proteome and phosphoproteome GSEA analyses were therefore conducted as independent analyses reflecting changes in protein abundance and phosphorylation state, respectively. All GSEA analyses were performed using the full ranked lists and the complete set of quantified proteins or phosphoproteins as background, consistent with a rank-based framework designed to detect coordinated pathway-level shifts rather than individual differentially expressed entities.

Flow cytometry analyses:
FlowJo v10.10.0

Statistical analyses:
GraphPad Prism (v10)

Image Processing: F
IJI- ImageJ2 (version 2.14.0/1.54f) and Leica Application Suite X (LASX) v3.7.4.23463

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 raw data generated for RNA-seq analysis in this study have been deposited in the Gene Expression Omnibus (GEO) database under the ID GSE272902. The mouse genome GRCm39 was used for the transcriptomic analyses.

The ATAQ-seq raw data has been deposited in GEO under the ID: GSE272968 and alignment to the genome can be found here (http://genome.ucsc.edu/s/JoseTerron/mm9_jaime).

The proteomic and phospho-proteomics data have been deposited to the ProteomeXchange Consortium with the dataset identifier PXD054102 and PXD058134.

Biological materials can be obtained upon request. Source Data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This study does not involve the recruitment of human participants. All data sets include only female data (as were based on cohorts of female breast cancer patients). Selection criteria of the cohorts has been described in the published studies (Curtis et al., 2012; Bergholtz et al., 2020; Strand et al., 2022; Reed et al.; 2024).
Reporting on race, ethnicity, or other socially relevant groupings	See above.
Population characteristics	See above.
Recruitment	This study does not involve the recruitment of human participants.
Ethics oversight	No ethical approval is required for the data used in this study.

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	Analyses of mouse samples were designed to have at least n ≥ 3. Based on previous experience, this sample size is considered suitable for detecting changes in related processes of mammary gland development and biology.
Data exclusions	Data was excluded if identified as outlier. Outliers were identified using ROUT test in GraphPad PRISM (v10).
Replication	The number of biological replicates is indicated in each experiment (figure legends).
Randomization	If randomization was required, block randomization was performed to maintain treatment groups even.
Blinding	All investigators were blind to the time-point and mice genotype, and during data collection and analysis.

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
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Immunofluorescence primary antibodies:

(catalog no. , supplier name , clone name)
Ck8 (Troma-I, DSHB, Troma-I)
Ck14 (AF-64, Covance, polyclonal)
Ki-67 (556003, Biolegend, B56-RUO)
Ck5 (905901, Biolegend, polyclonal)
Sma-1 (A2547-100UL, Merk, 1A4)
Progesterone receptor (Pr) (12683667, Fisher Scientific, SP2)
Gfp (ab13970, Abcam, polyclonal)
Ck6a (PRB-169P, Covance, polyclonal)

Immunofluorecence secondary antibodies:
(catalog no. , supplier name, clone name)
Alexa Fluor 488 Donkey anti rabbit (A-21206, ThermoFisher Scientific, polyclonal)
Alexa Fluor 555 Goat anti-rat (A21434, ThermoFisher Scientific, polyclonal)
Alexa Fluor 647 Goat anti-rabbit (A21245, ThermoFisher Scientic, polyclonal)
Alexa Fluor 647-conjugated AffiniPure Donkey anti-chicken IgY (H+L) (A78948, ThermoFisher Scientific, polyclonal)
Alexa Fluor 488 Goat anti-mouse (A11001, ThermoFisher Scientific, polyclonal)

Immunohistochemistry antibodies:
(catalog no. , supplier name , clone name)
Rankl (AF462, RnD, polyclonal)
Rank (AF692, RnD, polyclonal)
Gfp (ab13970, Abcam, polyclonal)
Estrogen receptor (Er) (AM (CRET94D/H9), CNIO Monoclonal Antibodies Core Unit, CRET94D/H9)

Flowcytometry antibodies:
(catalog no. , supplier name , clone name)
CD24-AlexaFluor700 (564237, BD Pharmingen™, M1/69)
CD49f-AlexaFluor647 (562473, BD Pharmingen™, GoH3)
CD31-PECy7 (102417, Biolegend, 390)
CD45-PECy7 (103113, Biolegend, 30-F11)

Validation

Validation of IF and IHC primary antibodies are available in the provider website:
Ck8 (Troma-I, DSHB, Troma-I): https://dshb.biology.uiowa.edu/TROMA-I

Ck14 (AF-64, Covance, polyclonal): https://www.informatics.jax.org/antibody/MGI:3033405

Ki-67 (556003, Biolegend, B56-RUO): https://www.bdbiosciences.com/en-pt/products/reagents/microscopy-imaging-reagents/immunofluorescence-reagents/purified-mouse-anti-ki-67.556003?tab=product_details

Ck5 (905901, Biolegend, polyclonal): https://www.biolegend.com/Files/Images/media_assets/pro_detail/datasheets/905901-keratin-5-polyclonal-chicken-antibody-purified-10957-IFU-Rev7.pdf?v=20230602063026

Sma-1 (A2547-100UL, Merk, 1A4): https://www.sigmaaldrich.com/DE/de/product/sigma/a2547?srsltid=AfmBOoqhQoqKPxZJbfOZaZ69GNcO9MX-x0cpsNqWznHsVppxUvG31FY8

Progesterone receptor (Pr) (12683667, Fisher Scientific, SP2): https://www.fishersci.dk/shop/products/lab-vision-progesterone-receptor-rabbit-monoclonal-antibody/12683667

Gfp (ab13970, Abcam, polyclonal): https://www.abcam.com/en-us/products/primary-antibodies/gfp-antibody-ab13970

Ck6a (PRB-169P, Covance, polyclonal): https://www.biolegend.com/en-gb/sean-tuckers-tests/purified-anti-mouse-keratin-6a-antibody-11459?GroupID=GROUP26

Rankl (AF462, RnD, polyclonal): https://www.rndsystems.com/products/mouse-trance-tnfsf11-rank-l-antibody_af462

Rank (AF492, RnD, polyclonal): https://www.rndsystems.com/products/mouse-rank-tnfrsf11a-antibody_af692?_gl=1*97xww2*_up*MQ..*_ga*ODY1NzE2NTUuMTc2ODEyMDkzMw..*_ga_2T0XXM39JP*czE3NjgxMjA5MzEkbzEkZzAkdDE3NjgxMjEwMjUkajYwJGwwJGgxNzM5NTYzMDcy

Estrogen receptor (Er) (AM (CRET94D/H9), CNIO Monoclonal Antibodies Core Unit, CRET94D/H9): https://www.cnio.es/investigacion-e-innovacion/servicios/histopatologia/tecnicas-disponbles/catalogo-de-anticuerpos-muestras-raton/

Flow cytometry antibodies were tested and titrated following manufacturer's indications regarding concentration for flow cytometry and under the supervision of flow cytometry technicians at the Spanish National Cancer Centre (CNIO).

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	Control (K14-Ctrl) mice were: Tg.KRT14-rtTA: Tg.tetO-CMV-Cre (provided by Dr. Cedric Blanpain) and Rosa26_ACTB-tdTomato_EGFP (mTmG, MGI: 3716464). Experimental mice (K14-Rank) carried Tg.KRT14-rtTA: Tg.tetO-CMV-Cre (provided by Dr. Cedric Blanpain) and Rosa26_ACTB-tdTomato_EGFP (mTmG, MGI: 3716464) combined with the R26Sortm5(CAG-Rank,-GFP)Ics/N allele generated by C Mueller. K14-RankΔ/Δ mice described in (Rocha et al., 2023) carried g.KRT14-rtTA: Tg.tetO-CMV-Cre (provided by Dr. Cedric Blanpain) and Rosa26_ACTB-tdTomato_EGFP (mTmG, MGI: 3716464) combined with Rank flox/flox (Rankfl/fl) (Hanada et al., 2009). All mice used were females in C57BL/6JOlaHsd or mixed background. The age of each the mice rage from 8 weeks to 16 weeks (exact age is indicated in the figure legends of each experiment). For aging and tumorigenesis experiments mice were sacrificed either when tumor reached 10 mm diameter or at humanitarian end point (age range: 50-100 weeks).
Wild animals	NA
Reporting on sex	We have only used female experimental mice as our study focus on the study of mammary gland development in females, including the reproductive cycle (puberty, pregnancy, lactation), controlled by sex hormones. Our study is also aimed at studying underlying mechanisms promoting breast cancer, which mainly affects women (99%), therefore, as a model organism we have only used female mice.
Field-collected samples	NA
Ethics oversight	Animal husbandry, ethical handling of mice, and all animal work were carried at Spanish National Cancer Research Center (CNIO) and IDIBELL animal facility according to guidelines approved by the Committee on Animal Care and following Spanish and European Union regulations. All animal experiments were approved by the ethical Committee for Animal Experimentation (#CBA07_2021) from the Autonomous Community of Madrid (Spain). All mice used in this study were maintained in pathogen free cages under controlled condition of temperature (20-26ºC), humidity of (30-70%) 12-hour light/dark cycles, food and water ab libitum.

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

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA

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	Fresh mamary glands were mechanically dissected by scissor chopping after draining lymph nodes were removed and enzymatically digested with digestium medium (Dulbecco's modified Eagle's medium (DMEM)/F-12 (Biowest, #L0093-500ml), 0.6 U/ml collagenase A (Sigma-Aldrich, #1108879300), 2.85 U/mL dispase II (ThermoFisher Scientific, #17105041), 2% HEPES (ThermoFisher Scientific, #H8357-100 ml) and 1% penicillin–streptomycin (ThermoFisher Scientific, #15140-122)) for 1h

	rocking (50 rpm) at 37°C. Samples were washed with CO2 independent media (ThermoFisher Scientific, # 18045088) supplemented with 2mM glutamax (Gibco, #35050-061). Single cells were isolated by sequential treatments with 0.25% trypsin/0.1% EDTA solution (Sigma-Aldrich, #T3924-100ml) for 1min at RT and dispaseII/DNAseI (ThermoFisher Scientific, #17105041 / Roche, #10104159001) at 37°C for 5 min. Enzymatic digestions were blocked by adding CO2 independent media (ThermoFisher Scientific, #18045088) supplemented with 2mM glutamax (Gibco, #35050-061) and 5% FBS (ThermoFisher Scientific, #F7524-500ml). Erythrocytes were eliminated by treating samples with cold (4°C) ammonium chloride solution (STEMCELL Technologies, #7008) for 1min at RT. Cell aggregates were removed by filtering the cell suspension with a 40µm strainer and single epithelial cells were counted manually using a Neubauer chamber. Single cells were blocked with blocking buffer (PBS + 0.5% Bovine Serum Albumin (Sigma-Aldrich, #13391+ 2mM EDTA) for 20min on ice. Single cells were incubated for 40 min on ice with the corresponding surface antibodies (CD24-AlexaFluor700 (564237, BD Pharmingen™, 1ug/ml), CD49f-AlexaFluor647 (562473, BD Pharmingen™, 1.25 ug/ml), CD31-PECy7 (102417, Biolegend, 0.25 ug/ml), and CD45-PECy7 (103113, Biolegend, 0.125 ug/ml)). Prior sorting, single cells solution was filtered by a 70µm strainer.
Instrument	BD FACSAria™ Cell Sorter
Software	FlowJo v10.10.0
Cell population abundance	In control females within the epithelial compartment (CD45-CD31-), luminal cells accounted for 16% while basal cells comprised 53%. In experimental females: luminal cells 6%, basal cells 60%.
Gating strategy	First, scattering assists in distinguishing the population of interest by assessing its volume (forward scatter) and morphological complexity (side scatter). Subsequently, doublets and dead cells are removed from the analysis. Then, the CD31-CD45- events are collected and classified into two populations: the luminal population (CD49flowCD24hi) and the basal population (CD49fhi/CD24low). Both populations are further examined for GFP expression.

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