

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
<input checked="" type="checkbox"/>	<input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of all covariates tested
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> 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)
<input checked="" type="checkbox"/>	<input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), 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
Data analysis	Illumina BCL Convert v4 Bedtools v2.26.0 Trimmomatic v0.39 Bowtie2 v2.3.4.1 Picard v2.18.26 SAMtools v1.7 deepTools v3.4.3 Macs2 v2.1.1 Genrich v0.6.2 STAR v020101 ChIPseeker DESeq2 v1.38.3 GenomicRanges RepeatMasker IGV ggplot2 EpiProfile2.0

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](#)

We used the *S. rosetta* genome (GenBank assembly accession: GCA_000188695.1) available from Ensembl protists (https://protists.ensembl.org/Salpingoeca_rosetta_gca_000188695/Info/Index?db=core) for all analyses. RNA-seq data was published previously and have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE267344 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE267344>). ATAC-seq and ChIP-seq data generated in this study have been deposited to the NCBI Short Read Archive. ATAC-seq is bioproject PRJNA1107385 and ChIPseq data is bioproject PRJNA1112805. Analysed BigWig files are available on figshare (https://figshare.com/articles/dataset/Salpingoeca_rosetta_ATACseq_and_ChIPseq_bigwig_files/_26028751). Mass spectrometry data has been deposited in the MassIVE database under dataset ID: MSV000094416.

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

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, 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.

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

n/a

Data exclusions

n/a

Replication

All sequencing experiences were performed in duplicate.

Randomization	n/a
Blinding	Not done as it was not relevant in this study to do so

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<p>H3K27me3 (Rose et al. 2016)</p> <p>H3K4me3 (Farcas et al. 2012)</p> <p>H3K4me1 Cell Signalling Technologies D1A9</p> <p>H3K27ac Cell Signalling Technologies D5E4</p>
Validation	Antibodies used were either validated in the paper specified or validation information is available on the company website!

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Salpingoeca Rosetta (ATCC PRA390)
Authentication	n/a
Mycoplasma contamination	n/a
Commonly misidentified lines (See ICLAC register)	n/a

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

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.

ChIP-seq data generated in this study have been deposited to the NCBI Short Read Archive, bioproject PRJNA1112805.
<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1112805>
 For BigWigs: https://figshare.com/articles/dataset/Salpingoeca_rosetta_ATACseq_and_ChIPseq_bigwig_files_/26028751

Files in database submission

All raw data files are available in the SRA. Final Bigwig files are available on FigShare.

Genome browser session

(e.g. [UCSC](#))

We do not use UCSC genome browser, all of the bigwig files have been deposited on figshare (https://figshare.com/articles/dataset/Salpingoeca_rosetta_ATACseq_and_ChIPseq_bigwig_files_/26028751) and can be easily visualized in IGV or similar free browsers

Methodology

Replicates

We performed 2 replicates for each ChIPseq experiment

Sequencing depth

All ChIPseq samples were sequenced 150bp paired-end to a depth of between 19.2 and 35.3 million reads per sample.

Antibodies

H3K27me3 (Rose et al. 2016)
 H3K4me3 (Farcas et al. 2012)
 H3K4me1 Cell Signalling Technologies D1A9
 H3K27ac Cell Signalling Technologies D5E4

Peak calling parameters

There is no peak calling of ChIP in the manuscript

Data quality

No peaks were called

Software

Reads were first processed using Trimmomatic to remove adaptors and they were trimmed to 100 bps. They were then mapped to the genome using Bowtie2102 (with the “-very-sensitive” option) and converted to BAM files and sorted using SAMtools103. PCR duplicates were removed using Picard and low-quality reads were removed using SAMtools view (“-q 30” option). SAMtools was used to extract sub-nucleosome sizes reads, i.e. reads with insert sizes less than 100 bps. DeepTools104 was used to generate PCA and Pearson correlation plots. Replicates were merged together for visualization and BigWig files were generated using bamCoverage (--binsize 10 --effectiveGenomesize 55000000 --normalizeUsing RPGC) and visualized using the Integrative Genomics Viewer (IGV) 105. Heatmaps and profile plots were generated using deepTools.