Corresponding author(s): Manuel Serrano
manuel.serrano@irbbarcelona.org

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	- qRT-PCR data were obtained using an ABI PRISM 7700 thermocycler (Applied Biosystem) and data were analyzed using Graph Pad Prism (version 9). - Cytometric data were collected using either of these two cytometers: FACSscalibur (BD Biosciences, Franklin Lakes, NJ) and FACS CANTO II (BD, San Jose, CA); and analyzed using FlowJo 9.6.2 software. - Immunofluorescence images were obtained using a Leica SP5 microscope and analyzed with Fiji (version 1.0). -RNAseq data were obtained in an Illumina HiSeq2500 sequencer; alignments were performed using with STAR software with default parameters. - Mass spectrometry data were collected using a Q Exactive HF mass spectrometer (ThermoFisher Scientific) and analysis was done using MaxQuant (v 1.6.0.16) software.
Data analysis	- Cytometric data were analyzed with FlowJo 9.6.2 software. - Immunofluorescence images: analyzed with Fiji (version 1.0). - Statistical analyses (excluding the sequencing data): analyzed with Graph Pad Prism (version 9). -RNAseq alignments also with STAR software with default parameters. RNAseq: number of reads per gene were calculated using the featureCounts function of the Rsubread package of the R statistical software. Gene annotations were performed using biomaRt with version "may2015.archive.ensembl.org". Perseus v1.5.5.2, Limma, Prostar package, Differential analysis was done with the R package DESeq2. RRHO available at: http://systems.crump.ucla.edu/rankrank/rankranksimple.php. Correlation matrix of ChIP-seq data was produced using Morpheus software, available from the Broad Institute: https://software.broadinstitute.org/morpheus/. -Analysis of deep-sequencing data: analyzed with R and with standard programs and packages as detailed step-by-step in the Methods

section. Briefly, ChIP-seq analysis was performed with the RUbioSeq pipeline (v3.8), using the following tools: FastQC v0.11.5, BWA v0.7.10, SAMtools v0.1.19, Picard tools vl.107, Bedtools v2.16.2 and MACS2 v2.1.1.20160309, SeqMiner 1.3.3e.

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 upon request. 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

DATA AVAILABILITY
RNA-seq and ChIP-seq data are available from the GEO database under accession numbers GSE112208 and GSE127186. The mass spectrometry proteomics data are available from the ProteomeXchange Consortium/PRIDE repository with the dataset identifier PXD009200. Published datasets included in this study: Mouse: GSE56138, GSE23943, GSE81285, GSE81045, GSE81045, E-MTAB-2600; Human/Primate: GSE76970, GSE76970, GSE76970, GSE87239, GSE87239, GSE60945, GSE59435, E-MTAB-4461, E-MTAB-2031, GSE44183, GSE36552, E-MTAB-3929; see publication References in Source Data Figure 4E and 4K. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

List of Figures with associated Raw data:
Fig1
Fig2
Fig3
Fig4
Fig5
Fig6
Fig7
Fig8

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Sample size was determined based on our previous experience and the work of other groups using embryos and stem cells as experimental model systems.
Data exclusions	FACS analysis: live/dead staining was used to exclude dead cells and debris from the analysis for the quantifiaction of Nanog-GFP, HERVH-GFP, Zscan4-GFP, MERVL-tomator-Red reproter PSC cell lines. FACS gating strategy is in Extended Data Fig.9G.
Replication	The experimental findings were reliably reproduced. Key inhibitor treatment experiments with CDK8/19-inhibitors with mouse and human PSCs were reproduced in independent laboratories. All experimental data was replicated at least in two independent experiments.
Randomization	Embryos and cell cultures were randomly allocated to control and experimental groups before experimental treatments (Methods section).
Blinding	The investigators were not blinded to group allocation. This report relies on studying the differences between specific cell types and culture conditions. Therefore blinding was not applied in these cases

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

Two of the five CDK8/19-inhibitors used in this study (see Table S1) were developed in-house at the CNIO Experimental Therapeutics program. Samples of these compounds (ETP-47799; ETP50586) are available upon reasonable request to the Corresponding Author.

Antibodies

Antibodies used

The following antibodies were used for Western blotting, Cytometry, cell culture immunofluorescence and ChIPseq: (Target; Provider; Clone; Supplier-product-code; dilutions used. Note: where possible, antibody clone and dilutions are listed):

Nanog (Western, IF) Chemicon/Millipore #AB5731 (1/5000)
Nanog (FACS, IF) eBiosciences #51-5761 (1/1000)
Pou5fl/Oct4 BD Biosciences/Pharmingen #611203 (1/5000)
Total RNA Pol II (RPBl) Santa Cruz #sc-899x (N-20) (1/5000)
RNA Pol II Ser-SP Abeam #ab5131 (1/5000)
RNA Pol II Ser-2P Abeam #ab5095 (1/5000)
SMCl Bethyl Laboratories #A300-055A (1/5000)
Gapdh Sigma #G8795 (1/10000)
Beta-Actin Sigma #A5441 (1/50000)
gamma-Tubulin Sigma #T6557, CLONE GTU-88 ascites fluid (1/50000)
Lamin A/C Santa Cruz #sc-6215 (N-18) (1/1000)
CDK8/19 Cell Signaling #4106 (P455) Atlas Antibodies #HPA007053 CDK8/19 (both) Santa Cruz #SC-1521 specificity for both. (1/1000)
Cyclin C Santa Cruz #sc-1061 (1/500)
STATl total Cell Signaling #9172 (1/1000)
STATl Ser727-phospho Cell Signaling #9177 (1/1000)
ERKl/2 total (p44/42) Cell Signaling #9102 (1/2000)
ERKl/2 phospho (Thr202/Tyr204) Cell Signaling #9101 (1/2000)
SSEA4 Stem Cell Technologies #60062 Clone MC-813-70 (1/500)
Tral-81 Millipore #MAB4381 cMyc Santa Cruz #sc-40 {9E10) (1/200)
Sox2 Chemicon/Millipore #AB5603 (1/5000)
Zscan4c Millipore #AB4340 GATA6 R+D Systems #AF1700 (1/1000)
Tfe3 Atlas Antibodies #HPA023881 (1/500)
ICAM1/CD54 eBiosciences #13-0541 (1/500)
H3K9me3 Upstate/Millipore #07-442 (1/5000)
Alpha-fetoprotein (AFP) Abeam #ab46799 (1/500)
Vimentin Santa Cruz #sc-6260 (1/5000)
Nestin ThermoFisher Scientific #MAl-110 (10C2) (1/500)
Antibodies used for embryo/embryoid immunofluorescence in this study: Target, Code, Company, Dilution
Rabbit pAb anti-Nanog #ab80892 Abeam 1:200
Goat pAb anti-OTX2 #AF1979 R&D Systems 1:200
Goat pAb anti-GATA6 #AF1700 R&D Systems 1:200
Mouse mAb anti-Oct-3/4 #sc-5279 Santa Cruz Biotechnology 1:200
Rat mAb anti-Podocalyxin Clone 192703 #MAB1556 R&D Systems 1:500
Alexa Fluor 488 Phalloidin (F-actin) #A12379 ThermoFisher Scientific 1:500
Alexa Fluor 568 Donkey anti-Rabbit #A10042 ThermoFisher Scientific 1:500
Alexa Fluor 488 Donkey anti-Mouse #A21202 ThermoFisher Scientific 1:500
Alexa Fluor 647 Goat anti-Rat #A21247 ThermoFisher Scientific 1:500
Alexa Fluor 647 Donkey anti-Goat #A21447 Thermo Fisher Scientific 1:500

Validation

The expression patterns and/or subcellular localization of all the proteins analyzed in this study has been previously reported. This was used to validate the specificity of the antibody. Also, CDK8, CDK19, and CyclinC knockout and/or shRNA-knockdown cell lines were used to validate the respective CDK8, CDK19 and CyclinC antibodies, in this study. See Extended Data Fig.1.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mouse Cells and culture conditions Mouse ES cells: E14Tg2a.4 (wild-type parental, 129/Ola background) were from BayGenomics/MMRRC resource, University of California; Wild-type ES cells were derived at the Transgenic Mouse Unit of CNIO from E3.5 C57BL6 blastocysts, or mixed background C57BL6/129 blastocysts; Rosa26-GFP and Tg.CAG-Katushka-red ES cell lines were derived at the Transgenic Mouse Unit of CNIO from 129- Gt(ROSA)26Sortm1(CAG-EGFP)Luo/J mice (Jackson 006053) and from Tg.CAG-Katushka mice (Diéguez-Hurtado et al., 2011), respectively. Nanog-GFP knock-in mouse ES cells (TNGA, TON) were previously described (Chambers et al., 2007) and were shared by the laboratory of Austin Smith; The MERVL-td:Tomato mouse ES line was a 2C-reporter were shared by the laboratory of Todd Macfarlan (Macfarlan et al., 2012); The ZS mouse ES line was a 2C-reporter shared by the laboratory of Minoru Ko (Zalzman et al., 2010). Primary mouse embryo fibroblasts (wild-type, MEFs, passage 2) were obtained at E13.5 from pure inbred C57BL6 background mice as described previously (Palmero and Serrano, 2001), or from CDK8 flox/flox RERT-Cre mice. Human 293T cells were from ATCC. Human PSC (HERVH-GFP reporter human iPS previously described: (Wang et al., 2014); and WIBR3 human ES cells: (Gafni et al., 2013). For References, please see Main Text of this manuscript.
Authentication	The self-renewal properties and pluripotency markers of mouse and human PSCs were confirmed by RT-PCR, FACS, immunofluorescence, RNAseq, proteomics, and chimera developmental assays (see Figs 1-4 and S1-S4). CDK8/19-double KO mouse PSCs were authenticated by PCR, sequencing, and/or Western blotting (primers provided in the Methods section).
Mycoplasma contamination	Cell lines were routinely tested for mycoplasma contamination by PCR and confirmed negative
Commonly misidentified lines (See ICLAC register)	No commonly mis-identified cell 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	Mice (Mus musculus) were used to obtain mouse embryos for this study. The following strains and genetically-modified models were used: F1 (C57Bl6xCBA), CD1, CAG-GFP (both males and females). Nude mice were used for teratoma developmental assays of PSC pluripotency. Superovulated females for chimera experiments were used at 6 +- 1 week of age. Females for natural matings were used at 3 +- 1 month of age.
Wild animals	No wild animals were used.
Field-collected samples	There are no field-collected samples.

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.*	GEO database: Series: GSE112208.
Files in database submission	Raw data files (fastq) and Processed data files (BED format) are included in the GEOaccession series: GSE112208. The list of pooled and processed data files is: GSM3061003 Input GSM3061004 Ser2P-2i GSM3061005 Ser2P-8i GSM3061006 Ser2P-Control GSM3061007 Ser5P-2i GSM3061008 Ser5P-8i GSM3061009 Ser5P-Control GSM3061010 Total-Pol-II-2i GSM3061011 Total-Pol-II-8i GSM3061012 Total-Pol-II-Control
Genome browser session (e.g. UCSC)	GEO database: Series: GSE112208.

Methodology

Replicates	Mouse ES cells were cultured in three conditions: "Serum/LIF" (control), "2i", or "CDK8/19i" (see text for details of these three conditions.

In each culture condition, ChIP-seq was performed separately with three antibodies: (i) anti-Total RNA Pol II (the central subunit, RPB1); (ii) anti-RNA Pol II phosphorylated at Serine5 of the C-Terminal Domain; (iii) anti-RNA Pol II phosphorylated at Serine2 of the C-Terminal Domain.

Three ChIP replicates were performed for each target and for each culture condition (27 replicates: 9 ChIP-seq targets with 3 technical replicates each). A pooled Input reference was included as a background control. The three ChIP replicates of 20-25 million reads each were pooled to make >60 million reads for each target and for each culture condition. From this, after quality control and read normalization between samples, >42 million reads per condition remained.

Sequencing depth

Reads were single-end and 50bp in length. Q20 sequencing quality was 95-99% for all samples. Three ChIP replicates of 20-25 million reads each were pooled to make >60 million total reads for each target and for each culture condition. From this, after quality control and read normalization between samples, >42 million reads per condition remained.

Table below summarizes the number of reads per 1 replicate representative of each condition:

Sample // Number sequenced reads // N. aligned reads // N. aligned reads without duplicates // N. reads after random sampling to balance read number
Input-2i 25737344 24575560 21876817 15281123
Input-8i 22303891 20402885 18128523 15284967
Input-Control 22755044 20889372 18198596 15286208
Ser2P-2i 23819011 22830756 16282619 15286714
Ser2P-8i 23950726 23154997 17037805 15287258
Ser2P-Control 23726068 22637170 16654134 15286369
Ser5P-2i 23766079 22130771 15285221 15285221
Ser5P-8i 23879973 22714294 16240914 15284796
Ser5P-Control 23943661 22708214 16396186 15285861
Total-Pol-II-2i 23662069 22923122 17729866 15284573
Total-Pol-II-8i 24008598 23260015 17853240 15287660
Total-Pol-II-Control 23901412 23196118 17298552 15285794

Antibodies

ChIP-seq was performed separately with three antibodies: (i) anti-Total RNA Pol II (the central subunit, RPB1) Santa Cruz, #sc-899x (N-20); (ii) anti-RNA Pol II phosphorylated at Serine5 of the C-Terminal Domain, Abcam #ab5131; (iii) anti-RNA Pol II phosphorylated at Serine2 of the C-Terminal Domain, Abcam #ab5095. All the antibodies were used at 1/1000 dilution

Peak calling parameters

Analysis of deep-sequecing data: analyzed with R and with standard programs and packages as detailed step-by-step in the Methods section, and briefly here:
BIOINFORMATIC METHODS: ChIP-seq analysis was performed with the RUbioSeq pipeline (v3.8), using the following tools: FastQC v0.11.5, BWA v0.7.10, SAMtools v0.1.19, Picard tools v1.107, Bedtools v2.16.2 and MACS2 v2.1.1.20160309. Sequencing quality for ChIP-seq samples was analyzed with FastQC (Andrews, 2011). Reads were aligned with Bwa 0.7.5a (Li and Durbin, 2009) to the mouse reference genome (NCBIm37/mm9) using the default seed length (32) and allowing 1 mismatch in the seed. SAMtools 0.1.16 (Li et al., 2009b) was used to convert the output alignment SAM files to the BAM file format, sort the alignments and eliminate duplicated reads. BEDTools 2.23.0 (Quinlan, 2014) was used to convert the resulting files to the BED format. All ChIP and input samples were randomly normalized to the same number of reads. Peak calling was performed with MACS 2.0.10.20130712 (Feng et al., 2012) using the input sample as control for each one of the ChIP samples. BED files containing aligned reads for each sample, with duplicates removed and with a balanced number of reads, used as input for MACS2. BigWig files were obtained with bedGraphToBigWig (Kent et al., 2010) from the BedGraph files generated with MACS. Resulting peaks were annotated with PeakAnalyzer 1.4 (Salmon-Divon et al., 2010) and the distribution of peaks was plotted with SeqMiner 1.3.3e (Ye et al., 2014). Genome build NCBIm37/mm9.

Data quality

Sequencing quality was assessed by Fastqc quality control checks.
MACS settings for peak calling:
tag size = 25
band width = 300
model fold = 32
pvalue cutoff = 1.00e-05
Ranges for calculating regional lambda are : peak_region,1000,5000,10000

Software

ChIP-seq analysis was performed with the RUbioSeq pipeline (v3.8), using the following tools: FastQC v0.11.5, BWA v0.7.10, SAMtools v0.1.19, Picard tools v1.107, Bedtools v2.16.2 and MACS2 v2.1.1.20160309. Sequencing quality for ChIP-seq samples was analyzed with FastQC (Andrews, 2011). Reads were aligned with Bwa 0.7.5a (Li and Durbin, 2009) to the mouse reference genome (NCBIm37/mm9) using the default seed length (32) and allowing 1 mismatch in the seed. SAMtools 0.1.16 (Li et al., 2009b) was used to convert the output alignment SAM files to the BAM file format, sort the alignments and eliminate duplicated reads. BEDTools 2.23.0 (Quinlan, 2014) was used to convert the resulting files to the BED format. All ChIP and input samples were randomly normalized to the same number of reads. Peak calling was performed with MACS 2.0.10.20130712 (Feng et al., 2012) using the input sample as control for each one of the ChIP samples. BED files containing aligned reads for each sample, with duplicates removed and with a balanced number of reads, used as input for MACS2. BigWig files were obtained with bedGraphToBigWig (Kent et al., 2010) from the BedGraph files generated with MACS. Resulting peaks were annotated with PeakAnalyzer 1.4 (Salmon-Divon et al., 2010) and the distribution of peaks was plotted with SeqMiner 1.3.3e (Ye et al., 2014). Genome build NCBIm37/mm9.

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	Cell cultures were harvested by trypsinization to a single cell suspension (ensured by filtration through 0,7 micro-metre pore filters) and then either analyzed by FACS directly (for for fluorescent reporter lines), alternatively labelled with antibodies for pluripotency markers (as indicated in each Figure panel).
Instrument	Becton-Dickinson FACScalibur -Cytometric analyses. FACS-Aria -for live-cell sorting.
Software	Data were analyzed with FlowJo 9.6.2 software.
Cell population abundance	The abundance of the selected cell populations is reported in Figure panels where relevant
Gating strategy	A live/dead stain was used to exclude dead cells and debris from the analysis of live-cell fluorescent reporter cell lines. See Extended Data Figure 9G for FACS cytometry images of gating strategy employed here.

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