nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-020-0573-1

Global hyperactivation of enhancers stabilizes human and mouse naive pluripotency through inhibition of CDK8/19 Mediator kinases

Cian J. Lynch[1,2], Raquel Bernad[1,2], Ana Martínez-Val[3], Marta N. Shahbazi[4,5], Sandrina Nóbrega-Pereira[6], Isabel Calvo[1,2], Carmen Blanco-Aparicio[7], Carolina Tarantino[8], Elena Garreta[8], Laia Richart-Ginés[9], Noelia Alcazar[1,2], Osvaldo Graña-Castro[10], Gonzalo Gómez-Lopez[10], Irene Aksoy[11], Maribel Muñoz-Martín[1,2], Sonia Martinez[7], Sagrario Ortega[12], Susana Prieto[13], Elisabeth Simboeck[13], Alain Camasses[13], Camille Stephan-Otto Attolini[14], Agustin F. Fernandez[15], Marta I. Sierra[15], Mario F. Fraga[15], Joaquin Pastor[7], Daniel Fisher[13], Nuria Montserrat[8,16,17], Pierre Savatier[11], Javier Muñoz[3], Magdalena Zernicka-Goetz[4,18] and Manuel Serrano[1,2,17] ✉

Pluripotent stem cells (PSCs) transition between cell states in vitro, reflecting developmental changes in the early embryo. PSCs can be stabilized in the naive state by blocking extracellular differentiation stimuli, particularly FGF–MEK signalling. Here, we report that multiple features of the naive state in human and mouse PSCs can be recapitulated without affecting FGF–MEK signalling or global DNA methylation. Mechanistically, chemical inhibition of CDK8 and CDK19 (hereafter CDK8/19) kinases removes their ability to repress the Mediator complex at enhancers. CDK8/19 inhibition therefore increases Mediator-driven recruitment of RNA polymerase II (RNA Pol II) to promoters and enhancers. This efficiently stabilizes the naive transcriptional program and confers resistance to enhancer perturbation by BRD4 inhibition. Moreover, naive pluripotency during embryonic development coincides with a reduction in CDK8/19. We conclude that global hyperactivation of enhancers drives naive pluripotency, and this can be achieved in vitro by inhibiting CDK8/19 kinase activity. These principles may apply to other contexts of cellular plasticity.

Each cell type contains a unique repertoire of active enhancer complexes at specific DNA regions that arise due to a high concentration of lineage-specific transcription factors and signalling pathways[1–3]. The Mediator complex is enriched at enhancers, where it integrates multiple upstream signals and recruits RNA Pol II to nearby and distant promoters[4–6]. A large fraction of Mediator clusters within a small number of unusually long multipartite enhancers, which are known as super enhancers (SEs)[2,3,7]. SEs drive high expression of master transcription factors that maintain cell identity, yet SEs are also vulnerable to perturbation through sharp transitions in their phase separation[2,3,7]. Given the central role of Mediator in enhancer-driven transcription, modulation of its activity may influence cellular identity and plasticity[7,8]. Indeed, global enhancer activation was identified across multiple types of cancer in humans[9]. The kinase CDK8 and its similar, albeit poorly studied, paralogue CDK19 represent the only enzymatic activity of the thirty-subunit Mediator complex[4–6]. This CDK8/19 submodule negatively regulates the recruitment of RNA Pol II through its kinase activity, and it may also sterically hinder the association between Mediator and RNA Pol II[5,6,10–13]. Accordingly, chemical inhibition of CDK8/19 results in global hyperactivation of enhancer function within cancer cells[14]. There are additional layers of

---

[1]Tumour Suppression Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [2]Cellular Plasticity and Disease Group, Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [3]ProteoRed-ISCIII Proteomics Unit, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [4]Mammalian Embryo and Stem Cell Group, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. [5]MRC Laboratory of Molecular Biology, Biomedical Campus, Cambridge, UK. [6]Department of Medical Sciences and Institute of Biomedicine (iBiMED), University of Aveiro, Aveiro, Portugal. [7]Experimental Therapeutics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [8]Pluripotency for Organ Regeneration, Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute for Science and Technology (BIST), Barcelona, Spain. [9]Maintenance of Transcriptional Repression by Polycomb Proteins, Institut Curie, Paris, France. [10]Bioinformatics Unit, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [11]Stem Cell and Brain Research Institute, Univ Lyon, Université Lyon 1, INSERM U1208, Bron, France. [12]Transgenic Mice Unit, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [13]IGMM, University of Montpellier, CNRS, Inserm, Montpellier, France. [14]Bioinformatics-Biostatistics Unit, Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [15]Cancer Epigenetics and Nanomedicine Laboratory, Nanomaterials and Nanotechnology Research Center (CINN-CSIC), Institute of Oncology of Asturias (IUOPA), ISPA-Hospital Universitario Central de Asturias (HUCA), Universidad de Oviedo, Oviedo, Spain. [16]Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Barcelona, Spain. [17]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain. [18]Division of Biology and Biological Engineering, Caltech, Pasadena, CA, USA. ✉e-mail: manuel.serrano@irbbarcelona.org

complexity—CDK8 can phosphorylate multiple Mediator subunits, the RNA Pol II C-terminal regulatory domain, chromatin regulators and transcription factors[4–6,11–17].

PSCs provide a prototypical model of cellular plasticity, the transcriptional program of which can be stabilized, extinguished or recaptured[18–23]. Although human PSCs offer great therapeutic promise, successful clinical applications remain limited, as human pluripotency is less characterized and less stable in vitro compared with in mice[21,22,24]. Chemical inhibition of MEK and GSK3 kinases with a two-inhibitor cocktail known as 2i shields mouse PSCs from extracellular differentiation in a state that is known as naive pluripotency[25]. Mouse PSCs cultured in 2i (referred to as 2i-naive cells) phenocopy the stable and homogenous state of undifferentiated naive pluripotency that exists transiently in embryonic day 4.5 (E4.5) preimplantation embryo epiblast[18,19,25]. By contrast, culture of PSCs without 2i shifts cell identity towards the postimplantation epiblast at ~E6.5, also known as primed pluripotency[18,19,23]. Enhancer destabilization by chemical blockade of BRD4, which is a key component of enhancers and SEs, triggers the loss of Mediator-driven gene expression in many cell types and induces differentiation in primed PSCs[26–28]. Notably, 2i-naive PSCs are highly resistant to enhancer destabilization[28], indicating that there is an association between naive pluripotency and enhancer stability/resilience. MEK inhibition has been implicated upstream of potent and rapid reconfiguration of the transcriptome, proteome and DNA methylome, within embryonic or 2i-naive pluripotency[18–20,23]. However, the molecular mediators of 2i that are responsible for enhancer stabilization remain unclear.

Here we assessed the effect of inhibiting the activity of the Mediator CDK8/19 kinases, in order to elucidate the transcriptional basis of PSC identity and their plasticity. In summary, stimulating Mediator through its kinase module represses differentiation, favours self-renewal and upregulates preimplantation naive epiblast gene expression in mouse and in human.

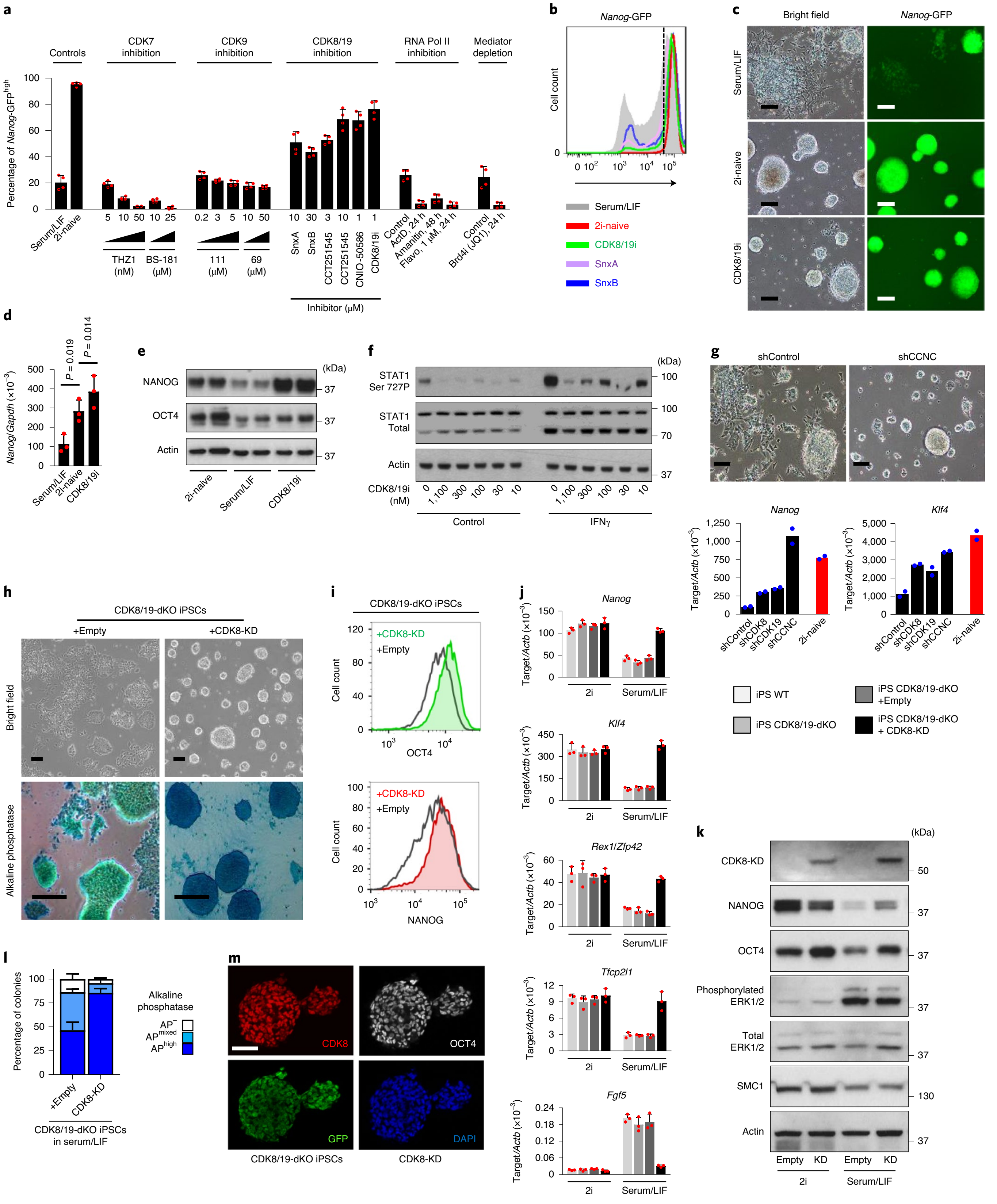
Results

Inhibition of Mediator kinase stabilizes mouse naive pluripotency. GFP knock-in reporters at key stem cell marker genes such as *Nanog* represent well-established and precise indicators of the naive (GFP^{high}) and primed states (GFP^{low})[18,22,29]. For example, in the 2i-naive state, *Nanog* promoter activity is enhanced, yielding a characteristically homogenous *Nanog*-GFP^{high} cell expression pattern and uniform dome-shaped colonies (Fig. 1a–c and Extended Data Fig. 1a). By contrast, the *Nanog* promoter is metastable in primed-state PSCs, reversibly oscillating between high and low activity, presenting a heterogeneous *Nanog*-GFP expression pattern and flattened diffuse colonies, indicative of a general underlying switch in the transcriptional program[18,20,23,29,30]. The BRD4 inhibitor JQ1 destabilizes enhancers, resulting in colony flattening and a GFP^{low} status (Fig. 1a), as reported previously[26–28]. In this experimental setting, we tested the effect of manipulating the transcriptional cyclin-dependent kinases (CDK7, CDK8/19 and CDK9) with a panel of small-molecule inhibitors. Several potent and structurally unrelated CDK8/19 inhibitors had a positive effect, inducing the formation of homogenous dome-shaped colonies, and upregulating both the *Nanog*-GFP reporter and endogenous *Nanog* expression, similar to PSCs in the 2i-naive state (Fig. 1a–e, Extended Data Fig. 1a and Supplementary Table 1), whereas inhibition of CDK7 or CDK9 did not. We assessed the potency and selectivity of CDK8/19 inhibitors, commercially available or developed in-house, using multiple methods: (1) selectivity was suggested using a KinomeScan panel of 456 kinases; (2) LanthaScreen assays demonstrated inhibitory activity at nanomolar concentrations against pure recombinant CDK8–CCNC and CDK19–CCNC; (3) luciferase reporter cell assays (TOP-FLASH); and (4) potent inhibition of phosphorylation of STAT1 at Ser 727 in human PSCs, which is a well-documented CDK8 target site[11,14,16,31] (Fig. 1f, Extended Data Fig. 1b, Supplementary Table 1 and Supplementary Information). On the basis of these data, we focused on the molecule that was generated at the Centro Nacional de Investigaciones Oncológicas (CNIO)—CDK8/19i-ETP-47799 (hereafter, CDK8/19i), which was the most effective at improving mouse PSCs (Fig.1a,b and Extended Data Fig. 1a; information about the structure and characterization of this inhibitor, as well as a comparison with other inhibitors used in this study, is provided in Supplementary Table 1 and Supplementary Information). In addition to improvements in the *Nanog*-GFP profile and colony morphology described above, the effect of CDK8/19i on mouse PSCs resembled 2i in three other ways: (1) it was observed in serum-containing and serum-free-based media (Fig. 1a and Extended Data Fig. 1a); (2) it was reversible after CDK8/19i withdrawal, with kinetics similar to that of 2i removal (Extended Data Fig. 1c); and (3) after removal of LIF or inhibition of LIF signalling using a JAK inhibitor, the presence of CDK8/19i delayed the downregulation of *Nanog*-GFP expression (Extended Data Fig. 1d,e). We conclude that inhibiting Mediator kinase CDK8/19 shifts mouse PSC morphology and *Nanog* expression towards their characteristic status in the naive state[18,23,29].

As a genetic validation, depletion of CDK8, CDK19 and, most successfully, their regulatory partner cyclin C (CCNC; which is

Fig. 1 | An inhibitor screen for factors that promote the PSC naive state identifies a distinct role for Mediator kinase activity. a, Fluorescence-activated cell sorting (FACS) analysis of the effect of the indicated treatments on *Nanog* expression (*Nanog*-GFP^{high}) per cell, using a mouse *Nanog*-GFP knockin reporter PSC line[29] in standard serum/LIF base medium. Data are mean ± s.d. of $n = 4$ independent experiments. ActD, Actinomycin D; Flavo, Flavopiridol; compounds 69 and 111 are specific CDK9 inhibitors. Details about inhibitor characterization are provided in Supplementary Table 1 and the Supplementary Information **b**, FACS analysis of *Nanog*-GFP expression with three different CDK8/19i inhibitor molecules. *Nanog*-GFP^{low} and *Nanog*-GFP^{high} cell populations in the serum/LIF population (grey). The dotted line indicates the threshold at which >95% cells are *Nanog*-GFP^{High} in 2i-naive culture conditions. Data are representative of three experiments. **c**, PSC colony morphology in the indicated treatments. Bright-field images and *Nanog*-GFP expression are shown. Data are representative of six experiments. **d**,**e**, Endogenous *Nanog* mRNA (**d**) or protein (**e**) expression levels in mouse PSCs adapted to the indicated conditions. Data are representative of three experiments. Data are mean ± s.d. Statistical analysis was performed using unpaired two-tailed *t*-tests; $^{*}P < 0.05$. **f**, The levels of the CDK8-target STAT1 phosphorylation at Ser 727 (Ser 727P). HERVH human iPSCs treated with CDK8/19i concentrations for 3 h with or without the simultaneous induction of STAT1-Ser 727P by interferon-γ for 3 h. Data are representative of two experiments. **g**, Cell morphology and qPCR with reverse transcription (RT–qPCR) analysis of mouse PSCs after 7 d of shRNA-mediated knockdown of *CDK8*, *CDK19* and *CCNC* (which encodes cyclin C). Data are the mean values from two experiments. **h**,**i**, Cell morphology and alkaline phosphatase staining (**h**) and FACS analysis of endogenous NANOG and OCT4 protein levels (**i**) in CDK8/19-dKO iPSCs stably expressing pMSCV-Empty (Empty) or pMSCV-CDK8-kinase dead (CDK8-KD). Data are representative of three independent clones. **j**,**k**, RT–qPCR analysis (**j**; data are mean ± s.d. from $n = 3$ independent clones) and western blot analysis of protein expression (**k**; data are representative of two experiments) in WT iPSCs or CDK8/19-dKO iPSCs stably expressing pMSCV-Empty or pMSCV-CDK8-KD, adapted to the indicated medium conditions. **l**, Alkaline phosphatase (AP) staining. Cells were fixed and stained 14 d after retroviral expression of pMSCV-Empty or pMSCV-CDK8-KD. Staining intensity was scored visually for each colony using ten fields of view. Data are mean ± s.d. from three experiments. **m**, Immunofluorescence in CDK8/19-dKO iPSCs expressing pMSCV-CDK8-KD-puro-IRES-GFP. Data are representative of four experiments. For **c**, **g**, **h** and **m**, scale bars, 100 μm.

essential for full kinase activity[8]) by short-hairpin RNA (shRNA) knockdown led to upregulation of *Nanog* expression and naive-like colony morphology (Fig. 1g and Extended Data Fig. 1f,g). In another genetic approach, we generated CDK8/19 double-knockout (dKO) mouse PSCs (Extended Data Fig. 1h–k). CDK8/19-dKO PSCs could self-renew indefinitely, but CDK8/19-dKO was insufficient to confer naive morphological features or *Nanog* upregulation. Importantly, CDK8/19-dKO PSCs no longer responded to CDK8/19 inhibitors (Extended Data Fig. 1l,m). Together, this suggested that the beneficial effects observed may require the physical



presence of the inactive kinase. In agreement, we found that reconstituting CDK8/19-dKO PSCs with exogenous CDK8 rescued the ability of these cells to respond to CDK8/19i, observed by naive morphological features and *Nanog*, *Klf4* and *Oct4* upregulation (Extended Data Fig. 1l,m). Moreover, CDK8/19-dKO PSCs that were reconstituted with a CDK8 kinase-dead mutant (CDK8-KD; D173A) displayed homogenous naive colony morphology, high expression of naive-state markers (Fig. 1h–m) and downregulation of *Fgf5*, which is a key marker of the primed state[18,19,23] (Fig. 1j), all without the need for any chemical inhibitor and despite maintaining active MEK–ERK signalling (Fig. 1k). Thus, CDK8/19-dKO cells expressing CDK8-KD phenocopy the effects of chemical inhibition of CDK8/19. Finally, postimplantation epiblast stem cells (EpiSCs; cultured with FGF2 and activin) are in a more developmentally advanced primed state than mouse PSCs in serum/LIF[18,19]. Interestingly, EpiSCs exogenously expressing CDK8-KD lost *Fgf5*, upregulated *Nanog*, *Rex1* and *Klf4*, and formed dome-shaped colonies with high alkaline phosphatase staining; together, these are characteristic of conversion to the naive state (Fig. 2a and Extended Data Fig. 2a). In summary, CDK8/19 kinase inhibition is sufficient to promote key characteristics of naive pluripotency, despite the continued presence of MEK–ERK signalling.

Long-term culture of mouse PSCs (>10 passages) in CDK8/19i maintained PSC naive features, including colony morphology, high alkaline phosphatase, *Nanog*-GFP$^{\mathrm{high}}$, high endogenous *Nanog*, high ICAM1 cell surface expression and nuclear localization of TFE3 (refs. [21,24,32–34]; Fig. 2b–d and Extended Data Fig. 2b,c). Long-term CDK8/19i-adapted PSCs displayed typical developmental capacity after inhibitor withdrawal, specifically, retinoic-acid-induced differentiation, embryoid body cardiac centre formation, spheroid polarization and lumenogenesis[35], generation of teratomas containing three germ layers, and robust chimaera contribution after morula aggregation and blastocyst microinjection assays (traced by constitutive GFP or RFP) evaluated at E4.5, E7.5, E14.5 and in fully developed adults that subsequently completed germline transmission (Fig. 2d–i and Extended Data Fig. 2d–f). Notably, the continued presence of CDK8/19i impaired the early developmental events[35] of polarization and lumenogenesis in vitro (Fig. 2e), an observation that is discussed below. Thus, PSCs that are long-term adapted to CDK8/19i maintain upregulation of naive features, self-renewal and developmental capacity.

CDK8/19i induces and stabilizes the naive state in human PSCs. We tested the effect of CDK8/19i on human stem cell identity. STAT3 overexpression plus 2i induces the human naive state[36], and we observed that CDK8/19i could replace 2i in this system (Fig. 2j). Even in the absence of STAT3 overexpression, other transgenes or chemicals, CDK8/19i treatment progressively converted human induced PSC (iPSC) colonies from flat and primed-like, to dome-shaped naive-like birefringent morphology. This was observed for a total of 7 human PSC lines treated with 0.4 μM or 1.1 μM CDK8/19i/LIF for 2–3 weeks (Fig. 2k and Extended Data Fig. 2g), including human iPSCs carrying a specific *HERVH*-GFP reporter insertion that marks human naive cell identity[37] (Fig. 2k and Extended Data Fig. 2h). A 2i-based chemical cocktail (hereafter 2i p38iJNKi) induced naive colony morphology, as expected[33,37], and combined with selection by cell sorting yielded cultures with homogeneous *HERVH*-GFP$^{\mathrm{high}}$ (Extended Data Fig. 2h). Interestingly, treatment with CDK8/19 inhibitors (CDK8/19i or SnxA) also produced morphological conversion and increased GFP, similar to 2i p38iJNKi (Fig. 2l and Extended Data Fig. 2h). The changes induced by CDK8/19 inhibition were gradual, required no selection after passage (sorting or manual picking), required no additional supplements except for rhLIF and were stable in the continuous presence of the inhibitor. By contrast, CDK7 inhibition failed to change colony morphology or GFP fluorescence, and produced cell death (Extended Data Fig. 2h). Culturing human PSCs in CDK8/19i, with or without p38iJNKi, increased their clonogenicity, alkaline phosphatase intensity and pluripotency markers[32–34,38] NANOG, OCT4, SSEA4, TRA1–81, TFCP2L1 and KLF17 (Figs. 2m and 3a, and Extended Data Figs. 2i and 3a–c). MYC, which is known to be reduced in naive cells[25,32], was also reduced in cells maintained in CDK8/19i (Fig. 3a). Thus, similar to the observations in mouse PSCs described above, treatment of human PSCs with CDK8/19i establishes features that are characteristic of the naive state.

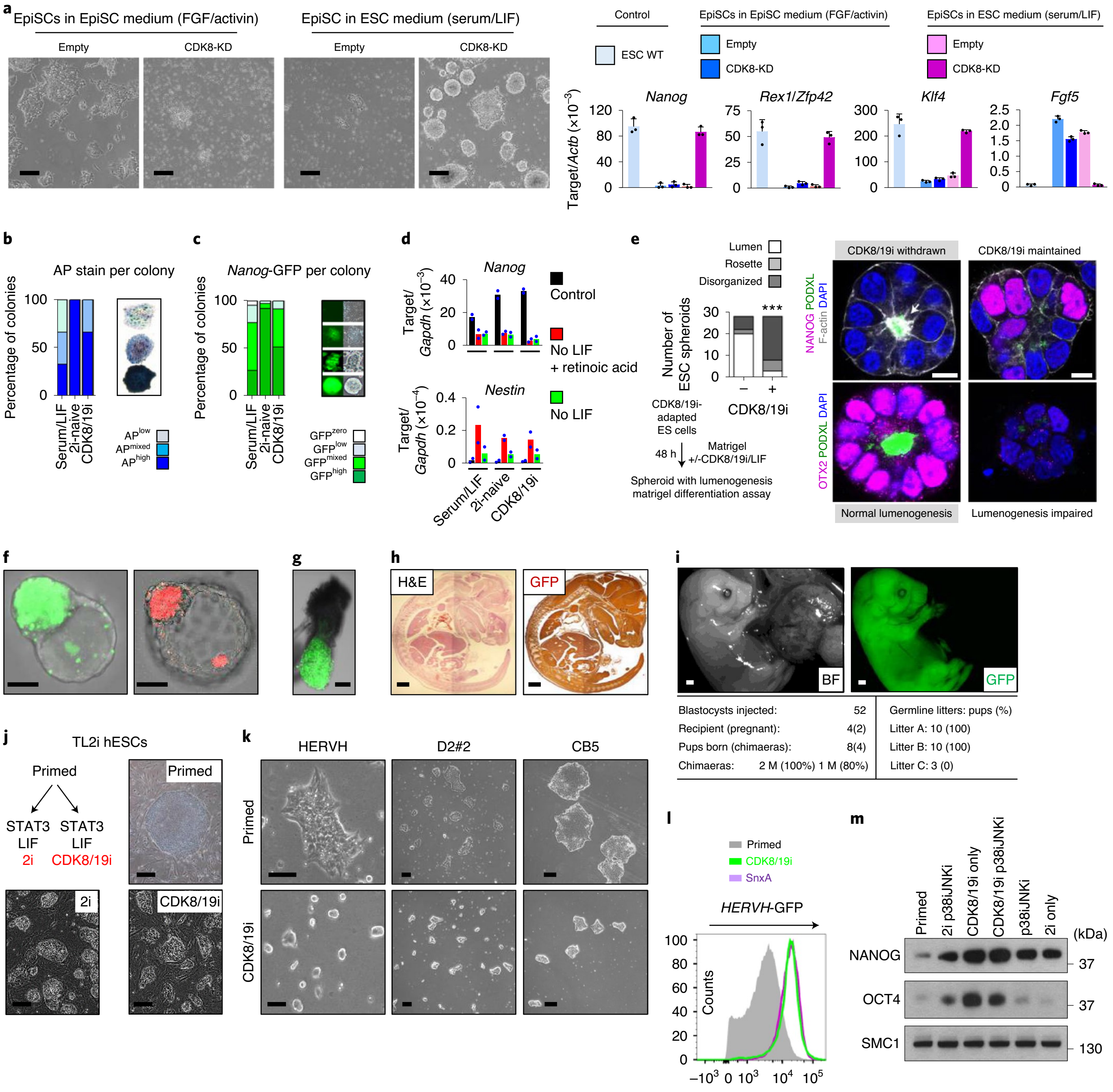
Developmental potential of CDK8/19i-adapted human PSCs. Chemical induction of the human naive state can trigger genomic instability, severely impairing developmental potential[24,39]. We found that CDK8/19i-adapted human PSCs (five lines) had a normal karyotype over >16 passages (Extended Data Fig. 3d) and, after inhibitor withdrawal, maintained the capacity to contribute towards all three embryonic germ layers by embryoid-body differentiation in vitro and by teratoma assay in vivo (Fig. 3b–d and Supplementary Table 1), comparable to control primed cells. Preimplantation interspecies chimerism tests for naive-specific properties, namely, capacity for clonal survival in a host embryo[40,41]. We tested CDK8/19i-adapted human iPSCs carrying a constitutive

Fig. 2 | Positive effect of long-term treatment of CDK8/19i on PSC self-renewal and pluripotency. a, Morphology and mRNA expression of mouse EpiSCs expressing pMSCV-Empty or pMSCV-CDK8-KD, then 7 d in EpiSC medium or standard embryonic stem cell (ESC) medium serum/LIF. Data are mean ± s.d. from $n=3$ experiments. Scale bars, 100 μm. **b**,**c**, Clonogenicity of mouse PSCs. *Nanog*-GFP PSCs were FACS-sorted one cell per well, cultured for 7 d and then stained for alkaline phosphatase (**b**) or scored for *Nanog*-GFP intensity (**c**) to assess the pluripotent status of each colony, in standard medium serum/LIF, 2i-naive or CDK8/19i conditions. Data are representative of three experiments. **d**,**e**, The differentiation capacity of mouse PSCs that were previously adapted to serum/LIF, 2i or CDK8/19i. **d**, PSCs differentiated as indicated in two-dimensional culture. Analysis of PSC exit from pluripotency (*Nanog* downregulation) and differentiation (*Nestin* upregulation) using RT–qPCR. Data are the mean values of two experiments. **e**, Pluripotency exit assessed using immunofluorescence after PSC culture in three-dimensional Matrigel with or without CDK8/19i/LIF to observe early epiblast development (rosette formation and lumenogenesis) in PSC spheroids[35]. 'Disorganized' indicates differentiation failure. Data are representative of three experiments, $n=30$ spheroids per condition. Statistical analysis was performed using unpaired two-tailed *t*-tests; $^{***}P=0.0097$. Scale bars, 10 μm. **f**–**i**, In vivo assays of developmental capacity. Mouse CDK8/19i-treated PSCs, constitutively labelled with ROSA26–GFP or Tg.CAG–Katushka, were aggregated with, or microinjected into, host E2.5 morulae. Embryo chimerism was assessed visually. E4.5 blastocyst, $n=10$ (**f**); E6.5 egg cylinder, $n=10$ (**g**); E14.5, $n=2$ (**h**); and perinatal E19.5, $n=4$ (**i**). In **i**, three male (M) adult chimaeras (bottom left, the percentage of chimerism on the basis of coat colour is indicated) displayed germline transmission, generating three litters (bottom right, coat colour confirmed germline transmission per litter). Scale bars, 25 μm (**f**), 100 μm (**g**), 1 mm (**h**), 1 mm (**i**). H&E, haematoxylin and eosin; BF, bright field. **j**, Induction of naive colony morphology in human OSCAR ESCs. Tamoxifen-inducible constitutively active STAT3/LIF/2i (TL2i)[36], or substituting CDK8/19i for 2i (TLCDK8/19i). Data are representative of three experiments. Scale bars, 100 μm. **k**, Induction of naive colony morphology in three human PSC lines; primed or cultured for 14 d with CDK8/19i. Scale bars, 100 μm. **l**, Cytometry analysis of *HERVH*-GFP intensity per cell in human PSCs; primed or cultured for 14 d with CDK8/19i. For **k** and **l**, data are representative of >5 experiments. **m**, Western blots of pluripotency markers in human PSCs; primed or cultured for 14 d with 2i-based or CDK8/19i-based medium, with or without p38iJNKi. SMC1 was used as a loading control.

Tomato-red marker for human–rabbit interspecies chimerism by microinjecting them into E2.5 rabbit morulae. Interestingly, the presence of human cells (Tomato$^+$) was detected 72 h later in up to 50% of the injected rabbit blastocysts (Fig. 3e). By contrast, human PSCs in the primed state were unable to integrate or survive in rabbit embryos (0 out of 24 rabbit embryos), similar to previous reports for primed state human PSCs within the embryos of mice, pigs and cattle[40,41]. In summary, long-term adaptation of human PSCs to CDK8/19i stabilizes naive pluripotency while preserving their developmental potential. We conclude that the role of CDK8/19 in pluripotency is conserved in mice and humans and, therefore, presumably across mammals.

CDK8/19i resets the transcriptome and proteome similar to 2i. Using RNA-seq, we compared global gene expression in mouse PSCs that were long-term adapted to CDK8/19i versus 2i. Overall, CDK8/19i altered gene expression with a magnitude similar to the magnitude in 2i conditions, and with a highly significant overlap in the identity and biological functions of genes that were up- or down-regulated in both serum-containing and serum-free media (Fig. 4a, Extended Data Fig. 3e,f and Supplementary Table 2). Compared with control primed conditions, naive pluripotency markers were enhanced in CDK8/19i and 2i (Fig. 4b and Extended Data Fig. 3g), whereas differentiation markers were globally downregulated in CDK8/19i and 2i conditions (Supplementary Table 2).

Endogenous retrovirus (ERV) expression is highly stage-specific during mammalian preimplantation and precisely defines naive and primed PSC identity[42–46]. The transcriptomic overlap between CDK8/19i or 2i treatments extended to ERVs; similar viral families were significantly up- or downregulated in mouse PSCs (Fig. 4c and Supplementary Table 2). In particular, LINE L1 families, each with thousands of copies across the genome, were regulated



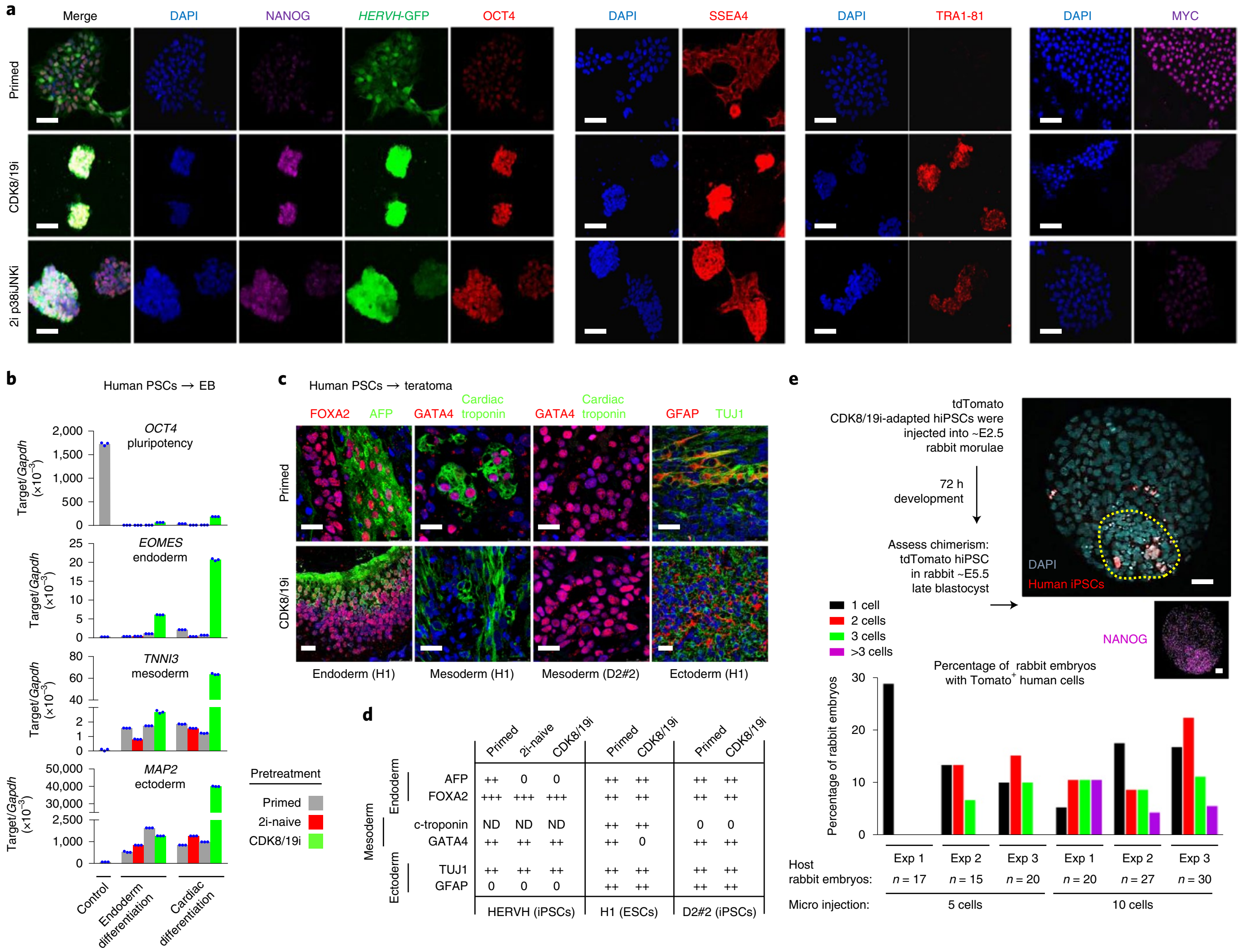


Fig. 3 | Positive effect of long-term treatment of CDK8/19i on the self-renewal and pluripotency of human PSCs. a, Immunofluorescence analysis of pluripotency markers. Human PSCs were cultured as indicated. Data are representative of two experiments. Scale bars, 100 µm. **b**, Embryoid body differentiation assay using endoderm-directed or cardiac-directed protocols with human PSCs. mRNA expression of pluripotency or embryonic germ layer markers on the basis of RT–qPCR analysis. Mean values of $n = 4$ technical replicates. Data for the H9 cell line are shown, representative of two experiments (human PSC lines H1 and H9: a summary of all tested lineage markers determined using RT–qPCR ($n = 17$) or immunofluorescence ($n = 6$) is provided in the Supplementary Information). EB, embryoid body. **c**,**d**, Human PSCs were adapted to primed or CDK8/19i conditions and then analysed using a teratoma differentiation assay. Data are representative of three experiments (human PSC lines: H1, D2#2 and HERVH). **c**, Immunofluorescence imaging shows markers for three embryonic germ layers in the H1 and D2#2 cell lines, as indicated. Scale bars, 50 µm. **d**, A summary of all of the tested lineage markers (6) determined on the basis of immunofluorescence in teratomas generated from the three human PSC lines described in **c**. The plus symbol (+) indicates detected; '0' indicates not detected. ND, not determined. A summary of all of the differentiation markers tested for all three cell lines in **c** and **d** is provided as source data. c-troponin, cardiac troponin. **e**, Interspecies chimaera assay in vivo to assess the developmental capacity of human PSCs that were adapted to primed or CDK8/19i conditions. Constitutively labelled human iPSCs (hiPSCs) (tdTomato (red); HERVH iPSC line) were introduced into host rabbit morulae of ~E2.5. Chimerism was assessed visually 72 h later in ~E5.5 rabbit blastocysts. The number of human cells introduced (5 or 10) and the number of embryos (n) in each of the three experiments (Exp 1–3) are indicated at the bottom. Quantification of the number of human cells observed in rabbit embryos is shown; data are from three independent experiments. A representative image (top) shows immunofluorescence of E5.5 rabbit blastocysts; the ICM is indicated by a dotted yellow line (determined by NANOG staining (inset)). Human PSCs that were adapted to CDK8/19i displayed a moderate contribution to human–rabbit chimaeras. Scale bars, 20 µm.

in close parallel, displaying highly similar pattern of expression in the CDK8/19i and 2i-naive states (Extended Data Fig. 3g,h). Another aspect of the plasticity of mouse PSCs is their ability to transition to a two-cell-like (2C) state, specifically marked by hyperactivation of the MERVL family of ERVs and by *Zscan4c* expression[46,47]. Stabilization of the naive state with 2i impairs the 2C-like fluctuation[46,47]. We also observed this in CDK8/19i-treated PSCs using multiple 2C markers, including *MERVL* and *Zscan4c*, and *MERVL*-Tomato and *Zscan4c*-eGFP 2C-reporter models (Fig. 4d, and Extended Data Figs. 3i–m and 4a–c). Finally, our CDK8/19i and 2i transcriptomic data correlated with published transcriptomes[48–51] from independent studies of 2i-naive mouse PSCs and the transcriptome of E4.5 epiblast single cells[52] (Fig. 4e and Extended Data Fig. 4d,e).

RNA-seq analyses of human PSCs that were adapted to CDK8/19i or a 2i-based naive cocktail overlapped significantly (Fig. 4f and

Supplementary Table 3). Markers of human and primate preimplantation epiblasts and in vitro naive human PSCs[32–39] were upregulated by CDK8/19i, whereas differentiation markers were repressed[53–59] (Fig. 4g, Extended Data Fig. 4f–j and Supplementary Table 3). Moreover, the global human ERV transcriptomes of CDK8/19i- and 2i-adapted cells overlapped extensively, including upregulation of the SVA, LTR7 and HERV families (Fig. 4h–j and Supplementary Table 3), consistent with reports of ERV expression in human and primate naive PSCs and preimplantation epiblast[43–45]. Finally, we observed that there is a high correlation between RNA expression data of CDK8/19i-adapted human PSCs and seven independent studies in human and primate PSCs from the in vitro naive state and from embryo naive epiblast single-cell analyses[39,53–60] (Fig. 4k).

Although PSC plasticity has been examined in terms of RNA expression, its proteome remains relatively poorly defined. We analysed the proteome of mouse PSCs in serum/LIF versus 2i-naive or CDK8/19i-adapted conditions. Across five mouse PSC lines, CDK8/19i altered the expression levels of 465 proteins, of which 159 (34%) changed in the same direction in 2i conditions (Fig. 4l,m, Extended Data Fig. 5a,b and Supplementary Table 4). Importantly, among the overlapping changes in both 2i-naive and CDK8/19i conditions, we noted that key pluripotency regulators, such as KLF4, and metabolic pathways, such as oxidative phosphorylation, featured among the most upregulated; by contrast, LIN28A, MYC-target genes and differentiation markers were downregulated (Fig. 4m, Extended Data Fig. 5c and Supplementary Table 4). Furthermore, proteomic changes in 2i and CDK8/19i conditions were significantly correlated with the transcriptomic changes observed (Extended Data Fig. 5d,e).

In summary, CDK8/19i upregulates pluripotency markers, reshapes the endogenous retroviral transcriptome and represses differentiation markers in a manner that is similar to the transcriptomic and proteomic resetting that was observed in previous studies of naive pluripotency, in vitro and in vivo, in mice and humans.

CDK8/19i does not reset global DNA methylation levels. Many 2i-based chemical cocktails induce global DNA hypomethylation, both in mouse and human PSCs[21]. This has been attributed to MEK-dependent stabilization of UHRF1, which is a critical factor for recruiting DNMT1 to the DNA[61]. Importantly, the pattern of demethylation induced by 2i diverges significantly compared with the preimplantation naive epiblast state, and is associated with PSCs exhibiting genomic instability, chromosomal defects and loss of pluripotency[24,39,62,63]. Recent 2i-variant cocktails (with partial MEK inhibition) offer the advantage of largely preserving global DNA methylation[62–64]. Importantly, neither mouse nor human CDK8/19i-adapted PSCs showed evidence of global DNA hypomethylation (Fig. 5a,b). Moreover, 2i or MEK inhibition alone induced demethylation of LINE L1 repeat regions (Fig. 5c) and major satellite regions (Extended Data Fig. 5f) but had no effect on the methylation of IAP repeats (Extended Data Fig. 5g), all as previously reported[65]. By contrast, CDK8/19i did not reduce methylation at any of these mouse repeat elements (Fig. 5c and Extended Data Fig. 5f,g) or UHRF1 protein levels (Supplementary Table 4). Thus, CDK8/19i induces naive features in the absence of global DNA hypomethylation, and this is probably due to its lack of MEK inhibition (Fig. 1k and see below) or UHRF1 downregulation. By not recapitulating the partial demethylation of the naive epiblast, CDK8/19i has the advantage of preserving chromosomal stability and pluripotency after cell expansion (see above), which is particularly relevant for naive human PSCs. This is consistent with variant medium cocktails that are based on minimizing MEK inhibition both in mouse and human naive PSCs[62–64].

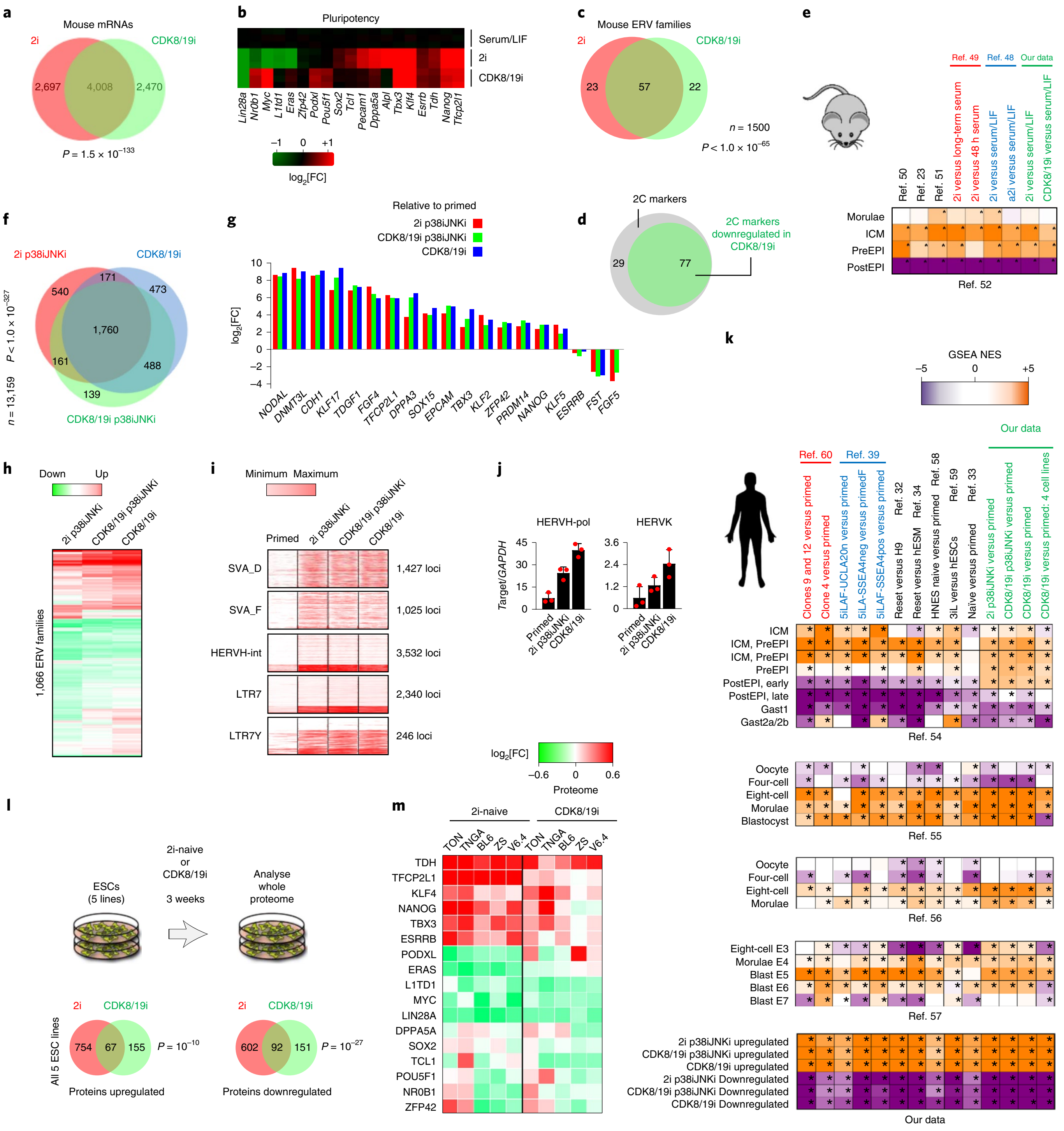
X-chromosome reactivation status is another molecular signature that has been reported in human naive pluripotency during MEK inhibition[21,66,67], which may be inferred by assessing XIST RNA expression in female cells. However, analysis using quantitative PCR (qPCR) revealed that there was very low XIST expression in our primed human PSCs (Extended Data Fig. 5h), suggesting that erosion of X-chromosome silencing may have already occurred in the parental cells under primed conditions, as observed previously[67]. Notably, some 2i-based cocktails reactivate XIST expression even in primed human PSCs displaying erosion of X-chromosome-silencing[66,67], but this was not the case for our CDK8/19i-adapted cells (Extended Data Fig. 5h). In summary, CDK8/19i treatment does not recapitulate the reactivation of XIST in X-chromosome-silencing-eroded primed cells, indicating another distinction between CDK8/19i and most human medium cocktails that are based on MEK inhibition.

CDK8/19i induces phosphorylation changes similar to 2i. We assessed the phosphoproteome of mouse PSCs after only 15 min exposure to CDK8/19i or 2i to explain their phenotypic similarity. Strikingly, out of 622 altered phosphorylation sites, 495 (79.6%) were similarly regulated by CDK8/19i and 2i (Fig. 5d,e). The co-regulated phosphorylation sites occurred on proteins that

Fig. 4 | Gene expression in mouse and human PSCs adapted to 2i or CDK8/19i. a,**b**, Overlap and hypergeometric significance (**a**) of differentially expressed mRNAs in mouse PSCs in 2i-naive or CDK8/19i conditions versus serum/LIF (RNA-seq; $n = 3$ biological replicates; false-discovery rate (FDR)-adjusted P (q) < 0.01). **b**, Heat map of changes in selected pluripotency regulators. **c**,**d**, Overlap and hypergeometric significance of differentially expressed ERV families (**c**) and overlap of 2C fluctuation markers (**d**) in mouse PSCs in 2i-naive or CDK8/19i versus serum/LIF. $n = 3$ biological replicates; $q < 0.05$. **e**, Heat map of normalized enrichment scores (NES) from a gene set enrichment analysis (GSEA) comparison of the mouse RNA-seq data shown in **a** ($n = 3$ biological replicates), or five other studies[23,48–51], versus stage-specific marker gene sets that were identified during mouse preimplantation development[52]. PreEPI/PostEPI, refer to pre- or post-implnatation epiblast cells. Significant comparisons are indicated by asterisks; $P < 0.05$ and $q < 0.05$. **f**,**g**, Overlap and hypergeometric significance (**f**) of differentially expressed mRNAs in human PSCs cultured as indicated versus control primed cells (RNA-seq; $n = 3$ biological replicates; $q < 0.05$, threshold >2× fold change). **g**, Selected human naive pluripotency markers (upregulated), or postimplantation primed epiblast markers (downregulated) in human PSCs cultured as indicated. **h**,**i**, Heat maps of the correlation in RNA expression changes for ERV families (**h**; $n = 1{,}066$ families, each row shows the mean value of the family) or individual loci of selected ERV families (**i**; genomic loci n values are indicated) in human PSCs cultured as indicated versus control primed cells. HERVH-int refers to the internal transcribed region of the HERVH genome. **j**, RT–qPCR analysis of RNA expression of the ERVs HERVH or HERVK, in human PSCs cultured as described in **f**. $n = 3$ biological replicates. Data are mean ± s.d. **k**, Heat map of NES scores from a GSEA comparison of our human PSC RNA-seq ($n = 3$ biological replicates), or seven other studies (indicated at the top)[32–34,39,58–60], versus stage-specific marker gene sets identified during human preimplantation development (below each heat map)[54–57]. Bottom, comparison between our data and the other studies. Significant comparisons are indicated by asterisks; $P < 0.05$ and $q < 0.05$. Colour scale as in **e** above. **l**,**m**, Proteomic overview. **l**, Overlap and hypergeometric significance of differentially expressed proteins averaged across five mouse PSC lines in 2i-naive or CDK8/19i versus standard serum/LIF culture. $n = 5$ biological replicates; $q < 0.05$. **m**, Heat map of protein changes in selected pluripotency regulators per cell line. Data for each cell line are provided in Extended Data Fig. 5a,b. Full gene lists, ERV lists, fold changes and statistical tests in **a**–**i** and **k** are provided in Supplementary Table 2. A list of differentially expressed proteins in **l** and **m** is provided in Supplementary Table 4.

were heavily enriched for functions in transcriptional regulation and key stem cell signalling pathways (Fig. 5e, Extended Data Fig. 5i and Supplementary Table 5). Note that CDK8/19i did not inhibit the kinase activity of purified recombinant GSK3 or MEK (Supplementary Table 1), and CDK8/19 inhibition did not reduce the relative levels of phosphorylated ERK (Figs. 1k and 5f,g and Extended Data Fig. 5j). However, 2i treatment reduced CDK8/19 kinase activity (Fig. 5f and Extended Data Fig. 5k) and moderately downregulated CDK8 protein levels (Extended Data Fig. 5l). These data suggest that CDK8/19 inhibition occurs downstream of 2i, such that both treatments result in highly overlapping changes to phosphorylation sites.

CDK8/19i resets global RNA Pol II loading similar to 2i. 2i and CDK8/19i induce similar phosphoproteomic changes (converging on transcriptional machinery) and similar transcriptomic changes. Thus, to understand how CDK8/19 inhibition phenocopies the transcriptome of 2i-induced naive pluripotency, we investigated the global regulation of RNA Pol II abundance on chromatin using chromatin immunoprecipitation with sequencing (ChIP–seq) in mouse PSCs in 2i or CDK8/19i, versus serum/LIF as control. Overall, total and Ser 5-phosphorylated RNA Pol II genomic distribution was consistent with published resources[23,68,69] (Encode, https://www.encodeproject.org/). We observed that 2i globally increases RNA Pol II binding to promoters (Fig. 5h,i and Extended



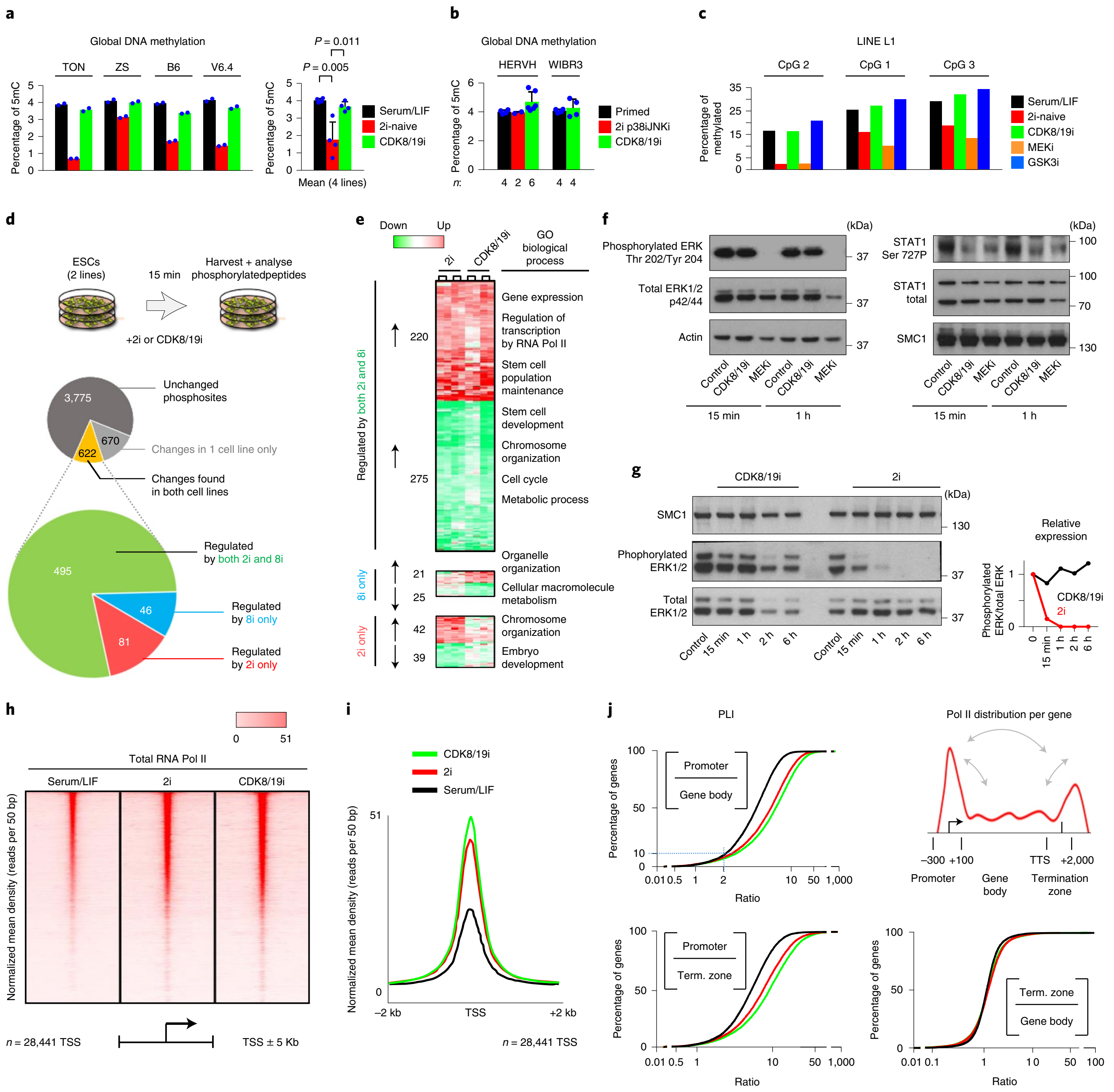


Fig. 5 | CDK8/19i regulates the phosphoproteome and RNA Pol II similar to 2i-naive pluripotency, but not DNA methylation. a, Global DNA methylation (5-methyl-cytosine (5mC)) changes in $n=4$ mouse PSC lines in 2i or CDK8/19i conditions. Left, data are the mean of $n=2$ technical replicates for each cell line. Right, data are mean ± s.d. of $n=4$ cell lines. Statistical analysis was performed using unpaired two-tailed *t*-tests; *P* values are indicated. **b**, Global DNA-methylation (5-mC) changes in two human PSC lines. Data are mean ± s.d. of $n=2$–6 biological replicates per cell line, as indicated. **c**, Pyrosequencing analysis of the CpG methylation status of three specific loci in the LINE L1 family of repeat elements in mouse PSCs in the indicated treatments in $n=1$ mouse PSC line (E14). **d**,**e**, Phosphoproteomic overview of mouse ESC lines in 2i or CDK8/19i (abbreviated as 8i) conditions for 15 min. **d**, Summary of total and differential phosphosites detected. **e**, Heat map and Gene Ontology (GO) summary for proteins displaying significantly differential phosphorylation. For **d** and **e**, $n=2$ biological replicates; $q<0.05$. Full protein lists are provided in Supplementary Table 5. **f**,**g**, Western blot analysis of protein phosphorylation during short-term exposure of PSCs to 2i or CDK8/19i, as indicated. **f**, ERK1/2 phosphorylation in mouse PSCs (left). Right, STAT1 Ser 727 phosphorylation in human iPSCs. **g**, ERK1/2 phosphorylation in mouse ESCs (left). Right, the relative ERK1/2 phosphorylation levels, normalized to total ERK1/2 levels. For **f** and **g**, data are representative of two biological experiments. **h**,**i**, ChIP–seq analysis of RNA Pol II abundance at all RefSeq transcription start sites (TSS; $n=28{,}441$), in mouse PSCs, treated as indicated. **h**, Heatmap of RNA Pol II abundance within +/−5kb of the TSS. **i**, Normalized mean density of metagenes within +/−2kb of the TSS. For the ChIP–seq analysis, $n=3$ pooled replicates. **j**, Cumulative plots of RNA Pol II loading ratios in mouse PSCs adapted to culture conditions as shown in **h** and **i**. RNA Pol II abundance was calculated in the indicated regions (top right) for each gene: promoter, gene body and transcription termination (Term.) zone (Methods). The ratios of RNA Pol II abundance between these regions were calculated for each gene, and are displayed cumulatively in the three plots ($n=12{,}072$ genes). For the promoter loading index (PLI; left), the dotted line indicates that 90% of genes have a PLI > 2.0. Colour scheme as in **i**. A full list of ChIP–seq data, genes and abundance ratios is provided in Supplementary Table 6.

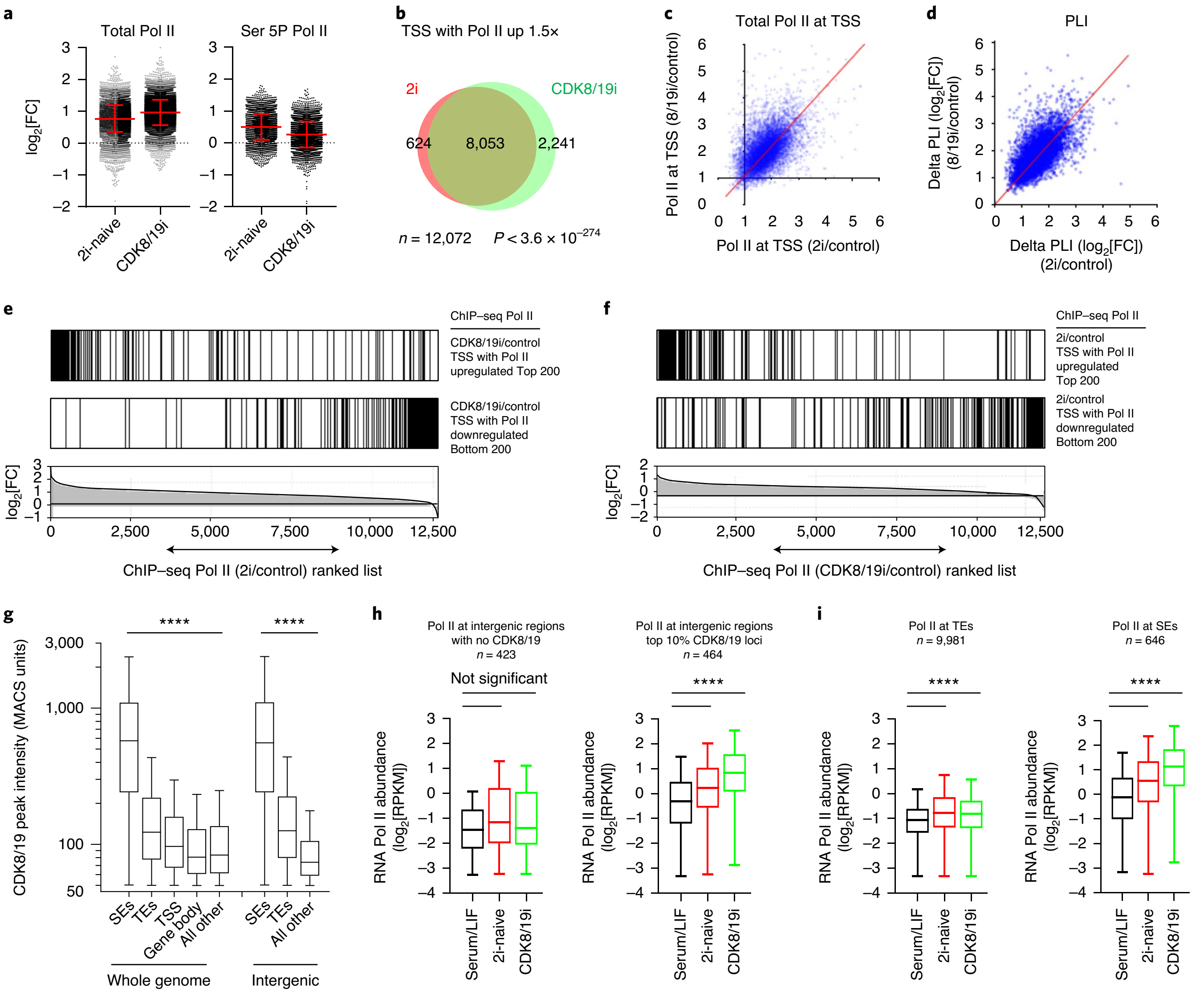


Fig. 6 | 2i and CDK8/19i exert similar effects on Mediator, RNA Pol II loading and enhancer activity. a, The change in total or phosphorylated Ser 5 (Ser 5P) RNA Pol II abundance at TSS, determined on the basis of ChIP-seq analysis in mouse PSCs in 2i or CDK8/19i conditions versus serum/LIF control. Data are mean ± s.d. Total Pol II, $n=12{,}693$ TSS; Ser 5P, $n=4{,}470$ TSS. **b**, Overlap and hypergeometric significance (the *P* value is indicated) of genes where RNA Pol II abundance at TSS increased by >1.5-fold in 2i and CDK8/19i (as described in **a**). TSS, $n=12{,}072$. **c**,**d**, The fold change in RNA Pol II abundance at TSS (**c**) and the fold change in RNA Pol II PLI (**d**) at genes ($n=12{,}693$) where RNA Pol II was detected in mouse PSCs in 2i or CDK8/19i versus control serum/LIF conditions. **e**,**f**, Genes with the largest changes in ChIP–seq RNA Pol II abundance are correlated between CDK8/19i and 2i-naive conditions. **e**, The individual promoter TSS with RNA Pol II loading altered in CDK8/19i by the greatest amount (top panel) or least amount (middle panel) versus serum/LIF control (top-200 TSS in each case) were compared against a ranked list (bottom panel) of differential Pol II loading on all TSS for 2i-naive versus serum/LIF control. **f**, Reverse comparison of the top-200 altered TSS in 2i-naive versus a ranked list of Pol II changes in CDK8/19i. **g**, CDK8/19 abundance in mouse PSCs, defined on the basis of ChIP–seq[2,3], peak-calling and grouped by genomic localization. The promoter TSS includes the TSS ±1 kb. Gene body includes exons, introns and the TTS ±1 kb. Statistical analysis was performed using unpaired two-tailed *t*-tests; $^{****}P<0.0001$. **h**, RNA Pol II abundance in mouse PSC genomic regions without CDK8/19 binding (left, $n=423$), or with the top-10% strongest CDK8/19-binding signals (right, $n=464$) as defined in Extended Data Fig. 8d. Statistical analysis was performed using unpaired two-tailed *t*-tests with Welch's correction; $^{****}P<0.0001$. RPKM, reads per kb of transcript per million mapped reads. **i**, RNA Pol II abundance in mouse PSCs at TEs (left, $n=9981$) or SEs (right, $n=646$), as defined previously[2,3]. Statistical analysis was performed using unpaired two-tailed *t*-tests with Welch's correction; $^{****}P<0.0001$. For **a**–**f**, **h** and **i**, RNA Pol II abundance was measured using ChIP–seq (three pooled replicates). For the Tukey box plots in **g**–**i**, the centre lines show the median values, the box limits represent the upper and lower quartiles, and the whiskers show 1.5× the interquartile range. The number of regions per group, *P* value calculations, a list of defined loci and HOMER functional annotations are provided in Supplementary Tables 3 and 7.

Data Fig. 6a–c), which was confirmed by reanalysing independent data[23,69]. Notably, this global effect of 2i was phenocopied by CDK8/19i, regarding both total and Ser 5-phosphorylated RNA Pol II (Fig. 5h,i and Extended Data Fig. 6a–c). We measured the abundance of RNA Pol II at the promoter, gene body and transcription termination site (TTS) for each gene (Fig. 5j, Extended Data Fig. 6d and Supplementary Table 6). Consistent with previous analyses in mouse PSCs[23,68], most genes (90%) possessed a ratio of promoter to gene body loading of >2.0 (Fig. 5j and Supplementary Table 6). A comparison of RNA Pol II ratios between the promoter, gene body

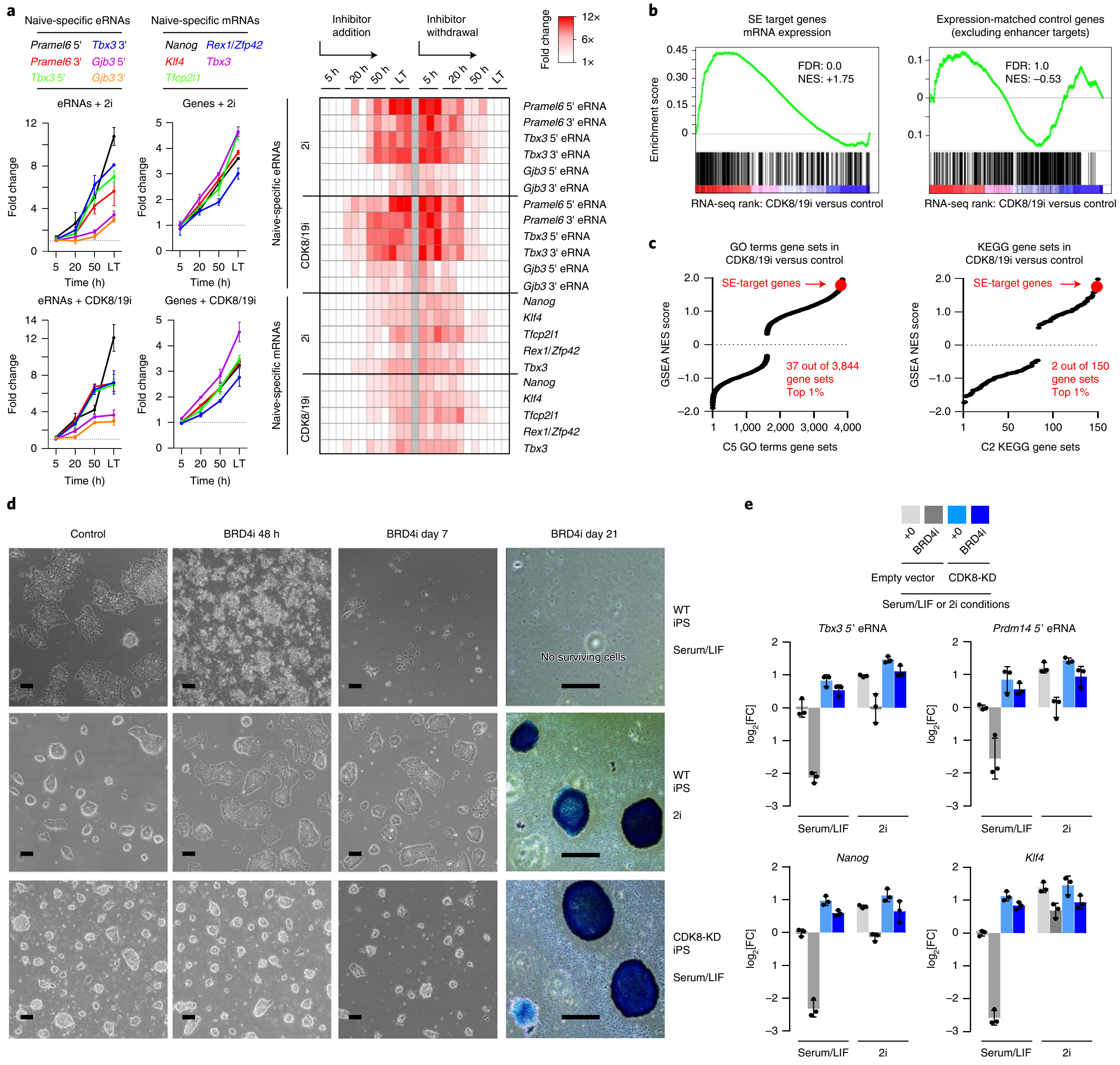


Fig. 7 | 2i and CDK8/19i hyperactivate naive-state enhancer activity, conferring resistance to enhancer destabilization. a, RT–qPCR analysis of pluripotency genes and naive-specific eRNA[71] abundance in mouse PSCs at time intervals after exposure to 2i or CDK8/19i, relative to standard serum/LIF control (left). LT, Long-term treatment. Data are mean ± s.e.m.; $n = 3$ independent experiments. Right, heat map summarizing the mean fold change in expression of the replicate experiments determined on the basis of RT–qPCR data shown on the left and from Extended Data Fig. 8g during inhibitor withdrawal. **b**,**c**, Selective upregulation of SE-target genes. **b**, GSEA of SE-target gene mRNAs in mouse PSCs adapted to CDK8/19i (left). $n = 3$ biological replicates. SE-target genes were defined as the single nearest gene on the basis of GREAT analysis (Methods) that were significantly upregulated (408 genes, $q < 0.001$). Right, expression-matched control genes (which were not predicted to be enhancer targets, but which are highly expressed similar to SE-target genes) show no significant change in levels (464 genes; $q = 1.0$). **c**, The relative specificity of SE-target gene upregulation by CDK8/19i was determined by comparison with databases of many other gene sets using GSEA. C5 GO terms ($n = 3,844$ gene sets) and C2 KEGG ($n = 150$ gene sets) as defined by the Broad Institute (Methods). SE-target genes lie within the top-1% most-significantly upregulated gene sets relative to these GO term or KEGG databases. For **b** and **c**, SE-target and expression-matched SE-non-target genes were defined as the single nearest gene on the basis of GREAT analysis (Methods). The SE-target gene list (408 genes) and the expression-matched control gene set (464 genes) are provided in Supplementary Table 7. **d**, Bright-field images of typical cell morphology after treatment with 500 nM BRD4i/JQ1 for 48 h (left) or 7 d (right) in WT iPSCs or in CDK8/19-dKO iPSCs stably expressing pMSCV-CDK8-kinase dead (CDK8-KD). Right, bright-field images of colonies that were fixed and stained for alkaline phosphatase at day 21/passage 5 of treatment with 500 nM BRD4i/JQ1. Data are representative of three independent cell experiments. **e**, RT–qPCR analysis of the expression of naive-specific eRNA[71] and marker genes after treatment with 500 nM BRD4i/JQ1 for 48 h. CDK8/19-dKO iPSCs with or without CDK8-KD were cultured with 2i or standard serum/LIF, as indicated. Data are mean ± s.d.; $n = 3$ clones.

or termination subregions of each gene indicated that 2i induces an increase in RNA Pol II binding selectively to the promoter region (Fig. 5j and Extended Data Fig. 6d). Importantly, this was recapitulated by CDK8/19i, increasing RNA Pol II binding to promoters at a similar magnitude to that observed in 2i-induced naive pluripotency, following a gene-specific pattern (Figs. 5j and 6a–f and Supplementary Table 6). Thus, 2i- and CDK8/19i-induced naive pluripotency is accompanied by widespread accumulation of RNA Pol II abundance at promoters. We also observed a correlation between changes in the abundance of RNA Pol II at promoters in 2i or in CDK8/19i conditions, as well as changes in mRNA expression for each gene (Extended Data Fig. 6e–i). In summary, gene-specific changes in RNA Pol II promoter loading may explain a significant proportion of the mRNA expression profile characteristic of 2i- or CDK8/19i-induced naive pluripotency.

CDK8/19i and 2i trigger activation of SEs. The primary role of Mediator is at enhancers, regulating the recruitment of RNA Pol II to promoters[4–6]. Using published ChIP–seq datasets[2,3] (Supplementary Table 7), we confirmed that CDK8/19 was enriched at promoter, typical enhancer (TE) and SE regions as previously defined in mouse PSCs[2] (Fig. 6g and Extended Data Fig. 7a–c). There is a strong correlation between the abundance of CDK8/19, Mediator subunits and other factors that are critical for enhancer activity[1,6] (such as, p300, CBP, Pol II or BRD4; Extended Data Fig. 7d); the highest levels of CDK8/19 occurred within SE regions (Fig. 6g and Supplementary Table 7); and, finally, putative target genes proximal to genomic CDK8/19-binding loci were highly enriched in preimplantation functions characteristic of pluripotent cell identity (Extended Data Fig. 8a–c). We therefore hypothesized that, in mouse PSCs, CDK8/19 inhibition might act through Mediator to trigger changes in enhancer activity, explaining the observed increase in RNA Pol II loading at promoters and regulation of pluripotent states. As CDK8/19 protein was particularly enriched at SE regions, we examined the impact of CDK8/19i on SE function. Enhancers contain RNA Pol II, which transcribes enhancer-derived RNAs (eRNAs)—a process that faithfully reflects enhancer activity[4,8,70]. We therefore measured the effect of CDK8/19i or 2i on the levels of RNA Pol II and eRNAs at SEs. Importantly, the abundance of RNA Pol II was selectively increased at CDK8/19-binding sites and, accordingly, RNA Pol II recruitment was also preferentially increased at SEs compared with TEs (Fig. 6h,i and Extended Data Fig. 8d). Consistent with this, mouse PSCs treated with 2i or CDK8/19i showed an increase in enhancer-derived eRNA levels, as well as RNA Pol II abundance, within enhancers specific for the naive state[71] (Fig. 7a and Extended Data Fig. 8e,f). Induction of naive-specific eRNAs and naive marker genes was an early event, occurring within 48 h after adding 2i or CDK8/19i, and it was rapidly reversible (Fig. 7a and Extended Data Fig. 8g). Finally, consistent with naive-specific enhancer activation, the expression levels of SE target genes were preferentially upregulated in both 2i and CDK8/19i conditions (Fig. 7b,c and Extended Data Fig. 8h). We conclude that, in PSCs, CDK8/19i and 2i hyperactivate existing SEs and upregulate SE-target genes in a manner that reinforces naive pluripotency.

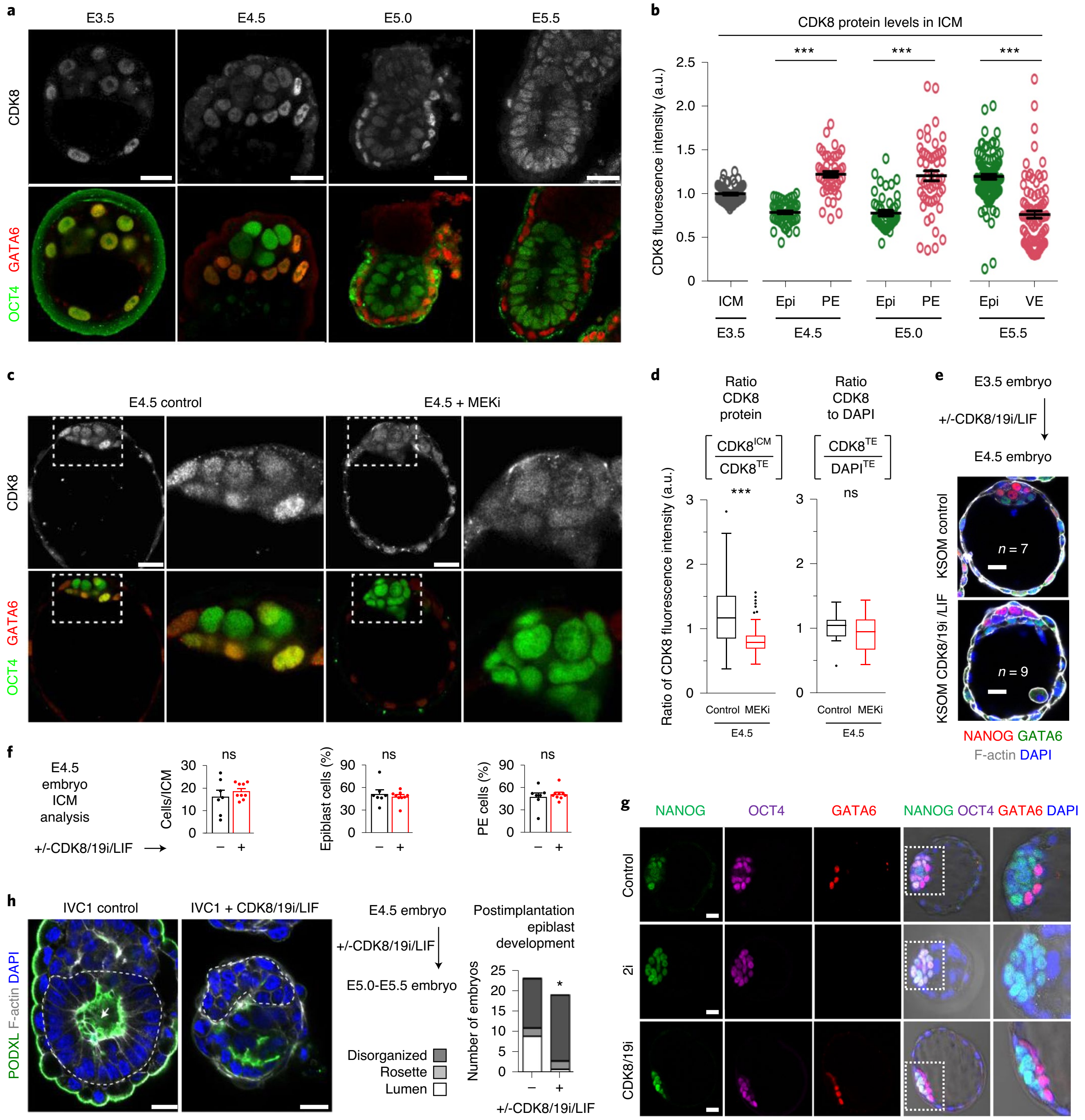
CDK8/19 inhibition compensates for BRD4 inhibition. Loss of Mediator function preferentially decreases the expression of enhancer target genes across multiple cell types[4–6,26–28]. In particular, BRD4 inhibition in primed state PSCs decreases the ability of Mediator to recruit RNA Pol II, and this results in the loss of Mediator-driven transcription, a collapse in pluripotency gene expression and differentiation[27,28] (Fig. 1a). Compared with primed PSCs, naive PSCs are highly resistant to the decreased Mediator activity and enhancer destabilization induced by BRD4 inhibition[28]. Interestingly, mouse PSCs lacking endogenous CDK8/19 and reconstituted with kinase-dead CDK8 were resistant to enhancer destabilization by BRD4 inhibition for ten passages (>3 weeks), maintaining naive morphology, and showed high expression of alkaline phosphatase, naive-specific pluripotency markers and naive-specific eRNAs, similar to 2i-naive PSCs (Fig. 7d,e and Extended Data Fig. 8i,j). Thus, PSCs expressing kinase-dead CDK8 phenocopy the robust resistance to enhancer destabilization that is characteristic of 2i-naive PSCs.

The roles of CDK8/19 during early embryonic development. Given our observations that CDK8/19 inhibition stabilizes naive pluripotency, we investigated CDK8/19 function during early embryonic development. We focused on CDK8, which we found is highly expressed compared with CDK19 in both mouse and human PSCs (Extended Data Figs. 1k and 9a). Using a CDK8-specific antibody (Extended Data Fig. 1j), we detected CDK8 protein from the mouse zygote to morula (Extended Data Fig. 9b). Consistent

Fig. 8 | CDK8 expression in vivo and the role of Mediator during mouse preimplantation development. a,**b**, Immunofluorescence imaging and quantification of CDK8 expression at the indicated time points during early mouse embryo development. **a**, A single *z* section. **b**, CDK8 protein levels per nucleus, quantified per time point, relative to the internal controls. OCT4/GATA6 coexpression marks all ICM cells at E3.5. OCT4/GATA6 segregation from E4.5 to E5.5 marks epiblast (Epi/OCT4$^+$) and PE (PE/GATA6$^+$), which later forms visceral endoderm (VE) at E5.5. Embryo staging, CDK8 quantification and normalization (Methods). Data represent two independent experiments (representative images); ICM, $n=6$ embryos, $n=64$ nuclei; E4.5, $n=5$ embryos, nuclei: $n=51$ (Epi), $n=48$ (PE); E5.0, $n=5$ embryos, nuclei: $n=48$ (Epi), $n=52$ (PE); E5.5, $n=6$ embryos, nuclei: $n=100$ (Epi), $n=84$ (PE/VE). Data are mean ± s.d. Statistical significance was assessed using Kruskal–Wallis tests with Dunn's multiple-comparison test. **c**,**d**, CDK8 expression is repressed by MEK inhibition in vivo. Embryos were incubated for 48 h with or without MEKi (8-cell stage to E4.5 blastocyst). **c**, Immunofluorescence analysis of CDK8 protein expression in E4.5 blastocysts. Data are representative of three experiments. **d**, CDK8 protein levels per cell quantified in ICM or trophectoderm (TE), relative to internal controls, per *z* slice/image. Left, control, $n=7$ embryos, $n=55$ images; MEKi, $n=12$ embryos, $n=44$ images. Right, control, $n=22$ embryos, $n=22$ images; MEKi, $n=27$ embryos, $n=27$ images. For the Tukey box plot, the centre lines show the median values, the box limits represent the upper and lower quartiles, and the whiskers show 1.5× the interquartile range. Statistical significance was assessed using unpaired two-tailed *t*-tests; $^{***}P<0.001$; ns, not significant. **e**,**f**, CDK8/19 inhibition does not prevent Epi/PE segregation. E3.5 embryos were incubated for 24 h with or without CDK8/19i/LIF during Epi/PE segregation, and then assessed using immunofluorescence (representative images; **e**). **f**, Quantification of ICM cell number, and lineage allocation in ICM, defined as: Epi/NANOG$^+$; PE/GATA6$^+$; ICM: NANOG$^+$ or GATA6$^+$. Data are mean ± s.e.m. from two experiments. Statistical significance was determined using unpaired two-tailed *t*-tests. Number of embryos: control, $n=7$; CDK8/19i, $n=9$. **g**, 2i prevents PE formation in the ICM, but CDK8/19i does not. E3.5 embryos were incubated for 24 h with or without 2i or CDK8/19i during Epi/PE segregation, and then assessed using immunofluorescence with brightfield overlap (representative images; $n=8$ embryos per condition, from two independent experiments). Lineage allocation defined as: ICM/OCT4+, Epi/NANOG+, PE/GATA6+. **h**, CDK8/19 inhibition interrupts pre- to postimplantation morphogenic events. Preimplantation E4.5 embryos were cultured until E5.0 in vitro[35] with or without CDK8/19i/LIF. PODXL and F-actin staining determines the emergence of the epiblast lumen (future pro-amniotic cavity). Epiblast indicated by the white dashed lines in the left and right image, respectively). Right, Lumenogenesis was quantified. Statistical significance was determined using the χ^2 test; $^*P<0.05$. Data are from two experiments. Number of embryos: control, $n=15$; CDK8/19i, $n=16$. For **a**, **c**, **e**, **g** and **h**, scale bars, 20 μm.

with this, CDK8-KO zygotes cannot progress beyond the 4–8-cell stage[72], and we observed that CDK8/19i impaired the progression of zygotes to the two-cell stage (Extended Data Fig. 9c). CDK8 activity is therefore essential for the zygote to morula transition.

We next investigated the role of CDK8 post-morula. CDK8 mRNA expression declines until the blastocyst stage, both in mouse and human preimplantation embryos (Extended Data Fig. 9d–f). CDK8 protein expression per cell was homogenous in the mouse inner cell mass (ICM) at E3.5 (Fig. 8a,b). Interestingly—at E4.5, when the ICM segregates into the naive epiblast and the primitive endoderm (PE)—CDK8 protein levels diverged, with lower levels in the epiblast compared to PE (Fig. 8a,b and Extended Data Fig. 10a). This pattern was transient, and it reversed in postimplantation epiblast at E5.5 (Fig. 8a,b and Extended Data Fig. 10a). To further document that CDK8 levels are reduced in the naive epiblast, embryos were cultured from E3.5–E4.5 with MEK inhibitor (MEKi), which blocks PE formation and permits only the development of naive epiblast[73,74]. As expected, in MEKi, the ICM contained only naive epiblast cells and not PE, this simplified the quantification of CDK8 in the ICM and trophectoderm and we confirmed reduced CDK8 expression in the ICM (Fig. 8c,d). The CDK8-binding partner and essential activating subunit cyclin C also altered its nuclear–cytoplasmic ratio during this developmental window. Specifically, E4.5 epiblast contained significantly less nuclear cyclin C compared



with E5.5 epiblast in vivo (Extended Data Fig. 10b,c), and a similar pattern was observed when comparing between 2i-naive and primed state PSCs in vitro (Extended Data Fig. 10d). In summary, the emergence of naive pluripotency during embryo development at E4.5 coincides with decreased CDK8 expression and decreased availability of its essential subunit cyclin C. This parallels the effect of MEKi on CDK8 expression and stabilization of naive epiblast identity in PSCs in vitro (Extended Data Fig. 5).

We wondered whether inhibition of CDK8/19 affects the emergence of naive pluripotency. Similar to MEKi, CDK8/19i treatment during E3.5–E4.5 did not interfere with embryo naive epiblast development (Fig. 8e–g) and enabled the derivation of PSC lines. In contrast to MEKi, CDK8/19i permitted PE formation (Fig. 8e–g). This suggests that the critical roles of MEK for PE segregation are independent of CDK8/19, and agrees with our observation that MEK activity is unaffected by CDK8/19i (Fig. 5f,g).

Finally, we examined the importance of CDK8/19 activity during preimplantation-to-postimplantation epiblast developmental progression. We focused on lumen formation within the epiblast, which marks the initiation of morphogenesis downstream of naive pluripotency exit[35]. We found that CDK8/19i treatment during E4.5–E5.5 impaired embryo epiblast lumenogenesis (Fig. 8h; for spheroids, Fig. 2e). This indicates that CDK8/19 activity is required to support epiblast development, from the naive preimplantation to primed postimplantation embryonic stages, consistent with the significant increase in CDK8 expression that we observed at this time (Fig. 8a,b and Extended Data Fig. 9e).

These data suggest that CDK8/19 expression in early embryonic development mirrors its function—the transition from zygote to morula, and the formation of the postimplantation epiblast require CDK8/19 activity; the intervening naive ICM has low CDK8 expression and reduced nuclear cyclin C (a summary is provided in Extended Data Fig. 10e).

Discussion

Here we uncovered a role for the Mediator kinases CDK8/19 in defining the equilibrium between naive and primed pluripotent states, in both mouse and human pluripotent cells. Collectively, our data indicate the following model: 2i and CDK8/19i rapidly induce a highly overlapping set of phosphorylation changes focused on the transcriptional machinery, triggering enhancer hyperactivation, a global increase in RNA Pol II recruitment to promoters and resetting gene expression. This includes upregulation of eRNAs, as well as resetting the expression of endogenous retroviral and repeat elements, as part of this cell identity conversion. Further evidence supporting transcriptional stabilization of naive pluripotency includes repression of 2C fluctuation in PSC identity, similar to in 2i conditions. Thus, the ability of 2i and CDK8/19i to induce naive features seems to originate from their common effect on Mediator and RNA Pol II transcriptional activity. In support of this, SEs interact with more target promoters[75], engage in more long-range interactions[75] and display increased H3K27ac[76] in the naive state versus primed. Our model agrees with the concept that transitions in cell identity are driven by early reconfiguration of the active enhancer network, which resets the transcriptional machinery to the new program[70,71,77].

The evidence presented here suggests a signalling hierarchy; in particular, MEK inhibition results in CDK8/19 inhibition, whereas inhibition of CDK8/19 does not affect MEK activity. Accordingly, we observed that (1) the ability of MEK and GSK inhibition (2i) to induce naive features in PSCs in vitro is recapitulated by CDK8/19i; (2) 2i and CDK8/19i exhibit a 79.6% overlap in downstream phosphorylation changes; (3) CDK8/19 downregulation coincides with the emergence of naive state in vivo, when MEK–ERK signalling is decreased[18,19]; and (4) both MEK–ERK activation[18,19,30,73,74] and CDK8/19 (refs. [78,79]) drive postimplantation epiblast differentiation, a process that we found is impaired by CDK8/19i, and a period during which CDK8 is upregulated. Thus, we propose that CDK8/19 inhibition is a common downstream feature of naive-inducing medium cocktails. Further studies will elucidate how MEK–ERK signalling regulates CDK8/19 Mediator activity in PSCs. Interestingly, Mediator hyperactivation through CDK8/19 inhibition triggers cancer cell death[14], while we find a similar approach reinforces naive pluripotent identity. Cancer cells commonly develop novel oncogenic SEs, becoming addicted to a defined range of enhancer-driven transcription that seems to be sensitive to perturbation[9,80]. This provides an interesting parallel with MEK inhibition, which is also detrimental to many cancer cells, but beneficial to pluripotency.

Stabilization of the human naive pluripotent state in vitro is challenging and remains to be optimized[21,24]. Our understanding of stem cell identity indicates a continuum of molecular changes along a spectrum from naive to primed states, which also reflects the developmental path in early embryos[18,19,21,22]. Where does CDK8/19i position PSCs along this gradient? We found that CDK8/19 inhibition recapitulates the majority of the molecular characteristics that are associated with the primed-to-naive transition. However, other molecular features that are associated with the more-naive end of this spectrum are not recapitulated by CDK8/19 inhibition, particularly, global DNA hypomethylation, X-chromosome reactivation[66,67] and SSEA4 downregulation[24,39]. Achieving these last features of naive pluripotency seems to come at a price. Naive-inducing medium cocktails that are dependent on MEK inhibition can generate harmful side effects, specifically acute chromosomal instability and imprinting erasure[24,39,62,63]. Interestingly, those other cocktails which do not downregulate SSEA4 and produce modest DNA demethylation, are not associated with genomic instability[24,33,64]. Similarly, CDK8/19i installs many naive features in human cells while maintaining SSEA4, DNA global methylation and genomic stability. CDK8/19i-treated cells retain a normal karyotype after prolonged culture. We suggest that these important differences are due to CDK8/19i not impinging directly on MEK signalling.

In summary, CDK8/19i stimulates the recruitment of RNA Pol II by Mediator. This hyperactivates enhancers and stabilizes the transcriptional program of naive pluripotent cell identity. Thus, chemical inhibition of CDK8/19 may help to solve remaining challenges in unstable human naive PSC culture. Similarly, these principles of stabilizing cellular identity may apply to other contexts of cellular plasticity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-020-0573-1.

Received: 2 May 2019; Accepted: 7 August 2020;

References

1. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
2. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
3. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
4. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
5. Jeronimo, C. & Robert, F. The Mediator complex: at the nexus of RNA Polymerase II transcription. *Trends Cell Biol.* **27**, 765–783 (2017).
6. Soutourina, J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **4**, 262–274 (2017).

7. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
8. Fant, C. B. & Taatjes, D. J. Regulatory functions of the Mediator kinases CDK8 and CDK19. *Transcription* **2**, 76–90 (2019).
9. Chen, H. et al. A pan-cancer analysis of enhancer expression in nearly 9000 patient samples. *Cell* **173**, 386–399 (2018).
10. Clark, A. D., Oldenbroek, M. & Boyer, T. G. Mediator kinase module and human tumorigenesis. *Crit. Rev. Biochem. Mol. Biol.* **50**, 393–426 (2015).
11. Poss, Z. C. et al. Identification of mediator kinase substrates in human cells using cortistatin a and quantitative phosphoproteomics. *Cell Rep.* **15**, 436–450 (2016).
12. Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev.* **23**, 439–451 (2009).
13. van de Peppel, J. et al. Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* **19**, 511–522 (2005).
14. Pelish, H. E. et al. Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **526**, 273–276 (2015).
15. Gonzalez, D. et al. Suppression of Mediator is regulated by Cdk8-dependent Grr1 turnover of the Med3 coactivator. *Proc. Natl Acad. Sci. USA* **111**, 2500–2505 (2014).
16. Bancerek, J. et al. CDK8 kinase phosphorylates transcription factor STAT1 to selectively regulate the interferon response. *Immunity* **38**, 250–262 (2013).
17. Galbraith, M. D., Donner, A. J. & Espinosa, J. M. CDK8: a positive regulator of transcription. *Transcription* **1**, 4–12 (2010).
18. Hackett, J. A. & Surani, M. A. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell* **15**, 416–430 (2014).
19. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
20. Galonska, C., Ziller, M. J., Karnik, R. & Meissner, A. Ground state conditions induce rapid reorganization of core pluripotency factor binding before global epigenetic reprogramming. *Cell Stem Cell* **17**, 462–470 (2015).
21. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* **17**, 155–169 (2016).
22. Wu, J. & Izpisua Belmonte, J. C. Dynamic pluripotent stem cell states and their applications. *Cell Stem Cell* **17**, 509–525 (2015).
23. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
24. Liu, X. et al. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods* **11**, 1055–1062 (2017).
25. Ying, Q.-L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
26. Bhagwat, A. S. et al. BET bromodomain inhibition releases the Mediator complex from select *cis*-regulatory elements. *Cell Rep.* **15**, 519–530 (2016).
27. Di Micco, R. et al. Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.* **9**, 234–247 (2014).
28. Finley, L. W. S. et al. Pluripotency transcription factors and Tet1/2 maintain Brd4-independent stem cell identity. *Nat. Cell Biol.* **5**, 565–574 (2018).
29. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
30. Ficz, G. et al. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
31. Dale, T. et al. A selective chemical probe for exploring the role of CDK8 and CDK19 in human disease. *Nat. Chem. Biol.* **11**, 973–980 (2015).
32. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
33. Gafni, O. et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
34. Theunissen, T. W. et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).
35. Shahbazi, M. N. et al. Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* **552**, 239–243 (2017).
36. Chen, H. et al. Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *Nat. Commun.* **6**, 7095 (2015).
37. Wang, J. et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409 (2014).
38. Stirparo, G. G. et al. Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human pre-implantation epiblast. *Development* **145**, dev158501 (2018).
39. Pastor, W. A. et al. Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* **18**, 323–329 (2016).
40. Wu, J. et al. Interspecies chimerism with mammalian pluripotent stem cells. *Cell* **168**, 473–486 (2017).
41. Wu, J. et al. Stem cells and interspecies chimaeras. *Nature* **540**, 51–59 (2016).
42. Gifford, W. D., Pfaff, S. L. & MacFarlan, T. S. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* **23**, 218–226 (2013).
43. Göke, J. et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135–141 (2015).
44. Grow, E. J. et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).
45. Theunissen, T. W. et al. Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515 (2016).
46. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
47. Eckersley-Maslin, M. A. et al. MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).
48. Kolodziejczyk, A. A. et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* **17**, 471–485 (2015).
49. Fidalgo, M. et al. Zfp281 coordinates opposing functions of Tet1 and Tet2 in pluripotent states. *Cell Stem Cell* **19**, 355–369 (2016).
50. Buecker, C. et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
51. Bulut-Karslioglu, A. et al. Inhibition of mTOR induces a paused pluripotent state. *Nature* **540**, 119–123 (2016).
52. Boroviak, T. et al. Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. *Dev. Cell* **35**, 366–382 (2015).
53. Blakeley, P. et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3613–3613 (2015).
54. Nakamura, T. et al. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016).
55. Yan, L. et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
56. Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
57. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).
58. Guo, G. et al. Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. *Stem Cell Rep.* **6**, 437–446 (2016).
59. Chan, Y. S. et al. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* **13**, 663–675 (2013).
60. Sahakyan, A. et al. Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87–101 (2017).
61. von Meyenn, F. et al. Impairment of DNA Methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* **62**, 848–861 (2016).
62. Yagi, M. et al. Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548**, 224–227 (2017).
63. Choi, J. et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219–223 (2017).
64. Di Stefano, B. et al. Reduced MEK inhibition preserves genomic stability in naive human embryonic stem cells. *Nat. Methods* **15**, 732–740 (2018).
65. Leitch, H. G. et al. Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
66. Sahakyan, A. et al. Human naive pluripotent stem cells model X chromosome dampening and X inactivation human naive pluripotent stem cells model. *Stem Cell* 87–101 (2017).
67. Khan, S. A., Audergon, P. N. C. B. & Payer, B. X-chromosome activity in naive human pluripotent stem cells—are we there yet? *Stem Cell Investig.* **4**, 54–54 (2017).
68. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
69. Williams, L. H. et al. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol. Cell* **58**, 311–322 (2015).
70. Arner, E. et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014 (2015).
71. Respuela, P. et al. Foxd3 Promotes exit from naive pluripotency through enhancer decommissioning and inhibits germline specification. *Cell Stem Cell* **18**, 118–133 (2016).
72. Westerling, T., Kuuluvainen, E. & Makela, T. P. Cdk8 is essential for preimplantation mouse development. *Mol. Cell. Biol.* **27**, 6177–6182 (2007).
73. Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724 (2010).

74. Nichols, J., Silva, J., Roode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215–3222 (2009).
75. Lopes Novo, C. et al. Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition. *Cell Rep.* **22**, 2615–2627 (2018).
76. Di Stefano, B. et al. The RNA helicase DDX6 controls cellular plasticity by modulating P-body homeostasis. *Cell Stem Cell* **25**, 622–638 (2019).
77. Factor, D. C. et al. Epigenomic comparison reveals activation of 'seed' enhancers during transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 854–863 (2014).
78. Postlmayr, A., Dumeau, C. E. & Wutz, A. Cdk8 is required for establishment of H3K27me3 and gene repression by Xist and mouse development. Development 147, dev175141 (2020).
79. Feldmann, A., Dimitrova, E., Kenney, A., Lastuvkova, A. & Klose, R. J. CDK-Mediator and FBXL19 prime developmental genes for activation by promoting atypical regulatory interactions. *Nucleic Acids Res.* **48**, 2942–2955 (2020).
80. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional addiction in cancer. *Cell* **168**, 629–643 (2017).

Methods

A list of primers, antibodies, shRNAs and CRISPR–Cas9 gRNAs is provided in Supplementary Table 8.

Pluripotency and differentiation assays. *Mouse and human research.* Animal experimentation at the CNIO was performed according to protocols that were approved by the CNIO-ISCIII Ethics Committee for Research and Animal Welfare (CEIyBA). Animal experimentation at the University of Cambridge was approved by the Home Office, performed according to Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 and reviewed by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB). Cdk8-flox/flox RERT-Cre mice were generated by D. Fisher (IGMM, Montpellier).

Human pluripotent stem cell studies were ethically approved at the CNIO, Madrid, by the Comisión de Garantías para la Donación y Utilización de Células y Tejidos Humanos and signed by Director of Instituto de Salud Carlos III (Nuevas fronteras en la Reprogramación Celular: Explotando la plasticidad cellular; 303). Studies at the IRB Barcelona, were approved by the Ethics Committee of the CMRB, and by Comisión de Seguimiento y Control de la Donación de Células y Tejidos Humanos del Instituto de Salud Carlos III and the Ministry of Health from Government of Catalonia (project numbers: 0336S/11730/2015; 0336S/11220/2016; 0336S/2473/2017; and 0336/747/2018).

Mouse cells and culture conditions. The following mouse PSCs were used: E14Tg2a.4 (wild-type parental, 129/Ola background) from BayGenomics/MMRRC resource, University of California; wild-type PSCs were derived at the Transgenic Mouse Unit of the CNIO from E3.5 C57BL6 blastocysts, or mixed background C57BL6/129 blastocysts; Rosa26–GFP and Tg.CAG–Katushka-red PSC lines were derived from 129-*Gt(ROSA)26Sor*$^{tm1(CAG\text{-}EGFP)Luo}$/J mice (Jackson, 006053) and from Tg.CAG-Katushka mice[81], respectively. *Nanog*-GFP knock-in mouse PSCs (TNGA, TON) were derived as described previously[29] and shared by the laboratory of A. Smith. MERVL-td:Tomato mouse 2C-reporter PSCs were shared by the laboratory of T. Macfarlan[46]. ZS mouse 2C-reporter PSCs were shared by the laboratory of M. Ko[82]. Mouse PSCs (ESCs and iPSCs) were routinely cultured on 0.1% gelatin-coated plates in a base medium of either serum/LIF (15% FBS) or knockout serum replacement (KSR, Invitrogen) 'KSR/LIF' (15% KSR) in DMEM (high glucose) basal medium with LIF (1,000 U ml^{-1}), non-essential amino acids, Glutamax and β-mercaptoethanol plus antibiotics. In cases in which the 2i two-inhibitor cocktail was used with mouse PSCs, 2i comprised 1 µM MEK inhibitor (PD0325901, Axon Medchem, 1408) plus 3 µM GSK3β inhibitor (CHIR 99021, Axon Medchem, 1386) as described previously[25]. Cultures were routinely tested for mycoplasma. Primary mouse embryo fibroblasts (MEFs; wild type, passage 2) were obtained at E13.5 from pure inbred C57BL6 background mice or from CDK8-flox/flox RERT-Cre mice. Human HEK293T cells were obtained from ATCC. All of the above-mentioned cells were maintained in DMEM and 10% FBS (Gibco), supplemented with antibiotics (penicillin–streptomycin, 100 U ml^{-1}). Reprogrammed iPSCs were initially derived and expanded on mitomycin-C-inactivated feeder cells on gelatin-coated plates, before transfer to gelatin only.

Human PSC resources. HERVH iPSCs were shared by the laboratory of Z. Izsvak (Max Delbruck Centre for Molecular Medicine)[37]. WIBR3 ESCs were shared by the laboratory of J. Hanna (Weizmann Institute of Science). OSCAR ESCs carrying inducible STAT3 were shared by the laboratory of P. Savatier (SBRI, Stem Cell and Brian Research Institute)[36]. H1 and H9 human ESCs, and CB5, D2#2 and D2#4 human iPSCs were shared by the laboratory of N. Montserrat (IBEC, Institute for Bioengineering).

Human PSC cell culture in the primed state. Human PSCs (H1, H9, WIBR3, HERVH, CB5, D2#2, D2#4 and OSCAR) were maintained in conventional primed conditions as described previously[33,36,37], specifically, by culture on growth-factor-reduced phenol-red-free Matrigel (BD Biosciences, 356231) with mTeSR1 medium (Stem Cell Technologies). Cultures were passaged every 5–7 d manually using either 2 mg ml^{-1} dispase (Gibco), 0.5 µM EDTA/1× PBS or Accutase (Gibco).

Resetting human PSCs from primed to naive state using 2i-based medium cocktail. The naive human pluripotent state was obtained using two methods. OSCAR PSCs were reset to the naive state with 2i (TL2i) or CDK8/19i (1.1 µM or 0.4 µM) plus rhLIF and STAT3 transgene induction, as described previously[36]. In a transgene-free approach, human PSCs were cultured in a 2i-based chemical cocktail[33] referred to here as 2i p38iJNKi. Cells were maintained on Matrigel (BD Biosciences, 356231) using mTeSR1 (Stem Cell Technologies). Medium was supplemented with 20 ng ml^{-1} of recombinant human LIF (Peprotech, as described previously[33]), 1 µM PD0325901 (MEKi, Axon Medchem), 1.5 µM CHIR 99021 (GSK3i, Axon Medchem), 10 µM SP600125 (JNKi, TOCRIS) plus 2 µM BIRB796 (p38i, Axon Medchem). To obtain and maintain the naive state using the 2i p38iJNKi medium cocktail, cells were selected at each passage by cytometry sorting for the top-10% *HERVH*-GFP levels or by repeated manual picking to select colonies with dome-shaped morphology. Conversion of human PSCs from primed to naive required three passages/rounds of selection over 14–18 d.

Resetting human PSCs from primed to naive pluripotent state using CDK8/19i. To adapt and maintain human PSCs to CDK8/19i culture (CDK8/19i-adapted), cells were maintained on Matrigel (BD Biosciences, 356231) using mTeSR1 (Stem Cell Technologies). Medium was supplemented with 20 ng ml^{-1} of recombinant human LIF (Peprotech), as described previously[33], plus 0.4 µM or 1.1 µM of CNIO-CDK8/19 inhibitor, or plus 10 µM of SenexinA-CDK8/19 inhibitor[83]. This adaption process was also successful with 10 µM SP600125 (JNKi, TOCRIS) plus 2 µM BIRB796 (p38i, Axon Medchem), but these additives were not required with CDK8/19i. Following background cell death in the first passage, colonies gradually became dome shaped within 10–14 d without additional selection, and could be expanded by passage, using 3–5 min treatment with Accutase (Gibco) or 0.5 µM EDTA/1× PBS to avoid confluency, usually every 5–7 d due to a slowdown in proliferation. The optimal CDK8/19i concentration was 1.1 µM for *HERVH*-GFP human iPSCs, whereas the optimal CDK8/19i concentration was 0.4 µM for all other human PSC lines.

Derivation of mouse ESCs. ESC line derivation was performed using standard methods. Eight-cell stage mouse embryos obtained from the oviducts of pregnant female mice were cultured in serum/LIF on mitomycin-C-inactivated MEF feeders plus 2i or CDK8/19i (added fresh every 2 d) until the emergence of colonies from hatched blastocysts. The feeders were not compatible with several days of CDK8/19i treatment; cells were therefore passed to fresh feeders every 2 d, then transferred to 0.1% gelatin only.

Derivation of mouse EpiSCs. PSCs in 2i/LIF cultured on gelatin were first induced to differentiate into epiblast-like cells over 48 h by seeding on fibronectin-coated plates (10 ng ml^{-1}) and switching to medium containing 1% KSR, N2B27, FGF2 (12 ng ml^{-1}) and activin A (20 ng ml^{-1})[50]. After 48 h, the cells were in a flat epiblast-like cell state, and the medium was switched to include 20% KSR, and expanded for five passages to stabilize the EpiSC primed state, which was confirmed by typical flat colony morphology and Fgf5 expression. EpiSC colonies were passaged as clumps.

Analysis of PSC self-renewal. Mouse or human PSC self-renewal and pluripotency was scored by colony morphology, cytometry (in mouse cells, *Nanog*-GFP heterogeneity and overall intensity, and costaining for ICAM1; in human cells, *HERVH*-GFP intensity, and assessing the expression of NANOG, OCT4, SSEA4 and TRA1-81), alkaline phosphatase staining (fixed cells; Promega, S3771), and using immunofluorescence and RT–qPCR (for pluripotency markers, indicated in each figure). The intensity of alkaline phosphatase staining was quantified by scoring colonies observed using bright-field microscopy in ten random fields of view per well.

Mouse PSC differentiation with LIF removal and retinoic acid. LIF was first removed for 24 h by culturing in LIF-free differentiation medium (as described for serum/LIF medium, except LIF was omitted). Next, retinoic acid was added (10 µM) from 24 h to 72 h, followed by LIF-free differentiation medium alone from 72 h to 96 h. Differentiation was also assessed using the same protocol of LIF-withdrawal except without adding retinoic acid.

Mouse PSC differentiation by hanging-drop culture and as embryoid bodies. PSCs were transferred to LIF-free differentiation medium (as described above) and suspended in hanging-drop culture at 1,000–5,000 cells per 20 µl for 48 h to form embryoid bodies (EBs), followed by transfer to suspension culture in low-adherence Petri dishes. Fresh medium was added every 3 d, and development of beating cells in cardiac centres was scored daily.

Morula aggregation and blastocyst microinjection in mouse chimaera assays. After ten passages in serum/LIF, 2i or CDK8/19i, mouse PSCs labelled constitutively with Rosa26–GFP or Tg.CAG–Katushka[81] underwent morula aggregation at E2.5 or blastocyst microinjection at E3.5 as described previously[25]. The extent of GFP$^+$ or Katushka-red$^+$ cell chimeric contribution was assessed on the basis of confocal fluorescence at E4.5 or embryos were introduced into CD1 pseudopregnant females for implantation, and collected at the following postimplantation time points: E6.5, E14.5 or E19.5. Chimaeras that developed to adulthood were assessed by coat colour contribution and capacity for germline transmission.

Cardiac-tissue- and endoderm-directed differentiation of EBs derived from hPSCs. Human PSC colonies were dissociated and cultured in suspension for 3 d to form EBs in DMEM/F12, 15% FBS, 2 mM L-glutamine, non-essential amino acids and penicillin–streptomycin. To generate endoderm, EBs were transferred to 0.1% gelatin-coated plates for 2 weeks in differentiation medium (DMEM, 20% FBS, 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol, non-essential amino acids and penicillin–streptomycin). To generate cardiac tissue, differentiation medium was supplemented with 100 µM ascorbic acid (Sigma-Aldrich). In all conditions, EBs spontaneously gave rise to neural cell clusters.

Teratoma assays. For mouse PSCs, 10^6 cells in 100 µl were injected subcutaneously in nude mice. For human PSCs, 2×10^6 cells in 30 µl were injected into the testis of male SCID beige mice.

Mouse embryo manipulation and analysis. Mouse embryo collection, culture for preimplantation embryo development in vitro and fixation for immunofluorescence analysis was performed as described previously[74,35]. Pre- to postimplantation embryo development in vitro, immunofluorescence analysis of CDK8 and cyclin C levels in preimplantation mouse embryos, and lumenogenesis by mouse PSC embryoid formation in Matrigel were performed as described previously[35]; further details are available from the corresponding author on request.

Viral production and iPSC reprogramming. Viral production and iPSC reprogramming were performed as described previously[84]. In brief, retroviral and lentiviral supernatants were produced in HEK293T cells. Filtered supernatants were collected after 48 h, and added to recipient cells in four infections. Retroviral supernatants delivered exogenous CDK8 expression constructs and iPSC reprogramming vectors. Lentivirus supernatants delivered shRNA knockdown vectors and CRISPR–Cas9 vectors. A list of plasmids is provided in Supplementary Table 8.

Interspecies chimaera developmental potency. Primed human iPSCs were precultured with ROCK inhibitor for 24 h, prepared as a unicellular suspension and electroporated (Neon Transfection System; Invitrogen; 1 pulse at 1,400 V for 20 ms) with 10 μg of DNA constructs for constitutive tdTomato expression (PB-Hygro-PGK-CAG-tdTomato and PBase pCMV-Transposase). Cells were subsequently plated on Matrigel in mTeSR1 medium supplemented with ROCK inhibitor for 24 h, then antibiotic selection with 20 μg ml^{-1} hygromycin was applied for 12 d, before cytometric sorting for tdTomato constitutively labelled cells.

Sexually mature NZW female rabbits (HyPharm) were superovulated. Then, 60 h after artificial insemination, fertilized embryos (eight-cell stage; E1.5) were flushed from explanted oviducts using Euroflush (IMV Technologies) and cultured in RDH medium (1/3 volume of DMEM–GlutaMAX, 1/3 volume of RPMI–GlutaMAX and 1/3 volume of Ham's F10–GlutaMAX; Life Technologies) at 38 °C and 5% CO_2.

Human PSCs were dissociated into single-cell suspension with trypsin, and 5–10 cells were microinjected under the mucus coat and zona pellucida of morula eight-cell stage rabbit embryos, the day after collection. After microinjection, embryos were sequentially cultured in CDK8/19i medium for 4 h, followed by 20 h incubation with a 1:1 mixture of RDH:CDK8i medium and finally in RDH medium for extended in vitro culturing. After 24 h of in vitro culture, early blastocyst stage embryos (E3.5) were rinsed three times in embryo-holding medium (IMV Technologies) and treated with 5 mg ml^{-1} protease E (Sigma-Aldrich) for 3 min at 37 °C to digest the mucus coat and weaken the zona pellucida. Embryos were then rinsed three times in 199 HEPES medium (Sigma-Aldrich) and cultured in RDH medium for 3 d until the late-blastocyst stage (E5.5). Rabbit embryos were fixed in 2% paraformaldehyde for 20 min at room temperature, washed in PBS + 0.1% Tween-20 and permeabilized in PBS + 1% Tween-20 overnight at 4 °C. After 1 h blocking with 5% donkey serum, embryo immunofluorescence was performed as described previously[36]. A list of the antibodies used is provided in Supplementary Table 8.

Molecular methods. *Transcriptional CDK inhibitors.* The structure and characterization of the CNIO CDK8/19 inhibitor (CDK8/19i-47799) as well as notes on all of the other transcriptional CDK inhibitors used in this study are provided in Supplementary Table 1.

Small-molecule inhibitor characterization assays. In vitro quantitative assays of enzyme inhibitors with recombinant proteins were performed using the LanthaScreen Eu-Kinase Binding Assay (Invitrogen) for CDK8–cyclin C, CDK9–cyclin T; CDKs, DYRK1A, GSK3β, mTOR, PI3K, PIM1/2, FLT3, KDR, KIT, PDGR-α and SRC. A summary of data from the small molecule inhibitor characterization assays is provided in Supplementary Table 1.

Generation of CDK8/19-dKO iPSCs. To target mouse CDK19, we designed sgRNA against CDK19 exon 1, targeting 76 bp downstream of the ATG translation start site to generate indels (a schematic of which is provided in Extended Data Fig. 1i). sgRNA sequences and plasmid details are provided in Supplementary Table 8. Primary CDK8-flox/flox RERT-Cre MEFs of passage 1–4 were infected with lenti-CRISPR-Cas9 containing the CDK19 sgRNA (pLenti-CRISPRV2; Addgene, 52961) followed by selection with puromycin (1 μg ml^{-1}). CDK19-KO was assessed using western blot. The MEFs were reprogrammed to iPSCs, single clones were picked and expanded, and CRISPR-induced indels were characterized by sequencing the CDK19 target region for frameshift mutations. CDK19-KO iPSC clones were compared versus iPSC clones that retained wild-type CDK19 expression, and no effect of CDK19-KO was observed in MEFs or in iPSCs. CDK8-KO was induced by 6 d of culture with 0.5 μM 4-hydroxy-tamoxifen to induce Cre-mediated deletion of CDK8 exon 2 (a schematic of which is provided in Extended Data Fig. 1j). CDK8-KO was confirmed using allele-specific PCR (to demonstrate deletion of exon 2; Extended Data Fig. 1h) and using western blot (to demonstrate complete loss of CDK8 protein; Extended Data Fig. 1j,k).

Stable exogenous expression of CDK8. Wild-type CDK8 (CDK8-WT) and catalytically inactivated kinase-dead CDK8 (CDK8-KD; D173A) were cloned into pMSCV-puro-IRES-GFP (Addgene, 21654) using the BglII and HpaI restriction enzymes, and confirmed by sequencing. Retroviral supernatants were generated in HEK293T cells with the packaging plasmid pCl-Eco (Addgene, 12371), followed by retroviral expression into CDK8/19-dKO iPSCs. Two rounds of FACS-selection by GFP expression were performed to enrich for expressing cells, and CDK8-WT or CDK8-KD protein expression was confirmed by western blot (a schematic of which in addition to western blot data are provided in Fig. 1k and Extended Data Fig. 1l).

FACS cytometry. FACS analysis of SSEA1 or ICAM1 was performed using FlowJo v.9.6.2 as described previously[84]. Live-cell analysis of the *Nanog*-GFP used 2i-adapted mouse PSCs to define the threshold (95% of cells) for the homogenous *Nanog*-GFP$^{\text{high}}$ population, against which other treatment groups were compared (Fig. 1a,b). Live-cell sorting for human PSCs carrying *HERVH*-GFP selected the top 10% GFP-expressing cells, as previously described[37]. The FACS gating strategy for live/dead cell discrimination is provided in Extended Data Fig. 9g.

Cell lysis, fractionation and western blot. Cell lysis, fractionation and western blots were performed as described previously[84]. A list of the antibodies used is provided in Supplementary Table 8. Nuclear/cytoplasmic fractionation was performed using the NE-PER kit (Thermo Fisher Scientific, 78833).

G-banding karyotype methodology. Subconfluent mouse and human PSC lines were arrested in metaphase by adding 0.02 μg ml^{-1} KaryoMax Colcemid (Gibco). Twenty metaphase spreads were analysed per condition.

Immunohistochemistry and immunofluorescence analysis of mouse embryos. Mouse tissues were fixed in formalin at 4 °C, embedded in paraffin block and sectioned at a thickness of 5 μm. Staining was performed using standard methods. A list of the antibodies is provided in Supplementary Table 8.

Mouse teratoma and embryoid body immunohistochemistry. Mouse teratoma and embryoid body immunohistochemistry analysis was performed as described previously[84]. A list of the antibodies used is provided in Supplementary Table 8.

Cell immunofluorescence. PSCs were grown on chamber slides using culture conditions indicated in each experiment. Confocal immunofluorescence staining and microscopy was performed as described previously[35,84] using a Leica SP5 microscope. A list of the antibodies is provided in Supplementary Table 8.

DNA methylation. Global DNA methylation status was quantified by mass spectrometry (MS). CpG methylation status at individual CpG sites of repeat DNA regions was assessed by DNA bisulphite conversion and pyrosequencing. A list of the primers used for PCR amplification and sequencing is provided in Supplementary Table 8.

Image analysis. All image analyses were performed using Fiji (http://fiji.sc).

Proteomics. *Full proteome quantitative analysis of five mouse ESC lines.* Five mouse ESC lines (ZS, TNGA, TON, BL6 and V6.4) were cultured in serum/LIF (as a control) or, additionally, with either 2i or CDK8/19i for >2 weeks. Cell pellets were collected by trypsinization, washed with cold 1× PBS and preserved immediately at −80 °C for further analysis. Protein sample preparation for MS, protein digestion, our scheme for isobaric labelling with iTRAQ8plex, detailed settings for high pH reverse-phase fractionation, detailed settings for the whole-proteome analysis using liquid chromatography coupled with tandem MS (LC–MS/MS) and bioinformatics analyses with the whole-proteome data were reported previously[84].

Phosphoproteome analysis of mouse PSC lines after 15 min of inhibitor treatment. Two mouse ES PSC lines (TON and ZS) were cultured in serum/LIF (as a control) or, additionally, with either 2i or CDK8/19i. Cells were treated with inhibitor for precisely 15 min, after which the cells were collected rapidly by scraping in ice-cold PBS, washed with ice-cold PBS, snap-frozen on dry ice and preserved at −80 °C for further analysis. Sample preparation for MS, protein digestion, isobaric labelling, phosphopeptide enrichment, micro high pH reverse-phase fractionation, settings used for phosphoproteome LC–MS/MS and bioinformatics analyses with phosphoproteomic data were performed as described previously[11,84].

Transcriptomics. *RNA isolation and RT–qPCR.* Total RNA was extracted (on-column; RNeasy kit with DNA digestion; Qiagen, 74104, 79254) and retrotranscribed into cDNA (Superscript Reverse Transcriptase; Biorad, 170–889). RT–qPCR was performed using the GoTaq qPCR Master Mix (Promega A6002) in an ABI PRISM 7700 thermocycler (Applied Biosystem). Input normalization of all RT–qPCR data was performed using the $2^{-\Delta\Delta C_t}$ method, using housekeeping genes *Actb* or *Gapdh* as indicated in each figure. A list of the primers used is provided in Supplementary Table 8.

RNA-seq transcriptomic analyses. The complete set of reads has been deposited in GEO (GSE112208 and GSE127186). A complete list of meta-analyses expression comparisons between this study and multiple mouse and human published datasets, in vitro and in vivo, is provided in Supplementary Table 3.

For RNA-seq analysis in mice, samples of 1 μg of total RNA (RIN numbers: 9.8–10; Agilent 2100 Bioanalyzer) were used. PolyA+ fractions were processed using the TruSeq Stranded mRNA Sample Preparation Kit (Agilent). Adapter-ligated library was completed by PCR with Illumina PE primers (8 cycles) and sequenced for 40 bases in a single-read format (Genome Analyzer IIx, Illumina).

For RNA-seq analysis in human cells, samples of total RNA (RIN numbers: 9.0–10; Agilent 2100 Bioanalyzer) were used. For library construction, 10 ng of total RNA samples were processed using the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech) according to the manufacturer's instructions. The resulting cDNA was processed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645). Adapter-ligated libraries were completed by PCR (8 cycles), and sequenced for 50 bases in a single-read format, (Illumina HiSeq2500).

Reads were aligned to the reference mouse genome (GRCm38/mm10) or the human genome (GRCh37/hg19) using TopHat-2.0.4 (using Bowtie v.0.12.7 and Samtools v.0.1.16, allowing for two mismatches and five multihits). Transcript assembly, estimation of abundance and differential expression were calculated using Cufflinks v.1.3.0. When comparing samples, total read numbers were normalized and visualized using SeqMiner v.1.3.3e or Integrated Genome Viewer from the Broad Institute (http://software.broadinstitute.org/software/igv/)

Functional analyses of differential gene expression. Lists of differentially expressed genes are provided in Supplementary Table 2 for mouse PSCs adapted to control serum/LIF, +2i or +CDK8/19i; and Supplementary Table 3, for human PSCs adapted to control/primed, +2i or +CDK8/19i. Genes were ranked using the FDR q-value statistic to identify significant genes ($q < 0.05$), and then by fold change in expression. Venn diagrams and hypergeometric testing were performed to assess any significant overlaps. GSEA (GSEA_Pre-ranked) was performed with MsigDB Hallmarks, C5 GO terms, C2 Curated, KEGG, Reactome and NCI databases, using the standard settings, and with 1,000 permutations for Kolmogorov–Smirnoff correction for multiple testing. GSEA enrichment data were obtained and ranked according to FDR q value (significance threshold, $q < 0.25$). Heat maps of expression data were generated using GenePattern. Rank–rank hypergeometric overlap (RRHO) analysis was performed using the ranked list of $\log_2$-transformed fold changes in gene expression or RNA Pol II abundance using the standard settings[85]. The colour intensity of the RRHO heat map indicates the $-\log_{10}$-transformed P value after Benjamini–Yekutieli correction of the hypergeometric overlap (http://systems.crump.ucla.edu/rankrank/rankranksimple.php)[85].

Analysis of repeat sequences and ERV expression was performed using Repbase datasets for rodent or human repeat elements and featureCounts. In Extended Data Fig. 3h, the total fragments per kb of transcript per million mapped reads for RNA expression of LINE L1 subtypes was calculated by grouping and summing by family, and was then arranged by evolutionary age[86]. A full list of three biological replicates for each viral subtype and the calculation for the summary of each viral LINE L1 family are provided in Supplementary Table 2.

Differential gene expression comparison of published mouse and human studies. Gene expression changes have been comprehensively characterized in mouse, primate and human PSCs in response to overexpression of transcription factors after culture in various medium cocktails or in vivo during the development of the mouse or human embryos[87,88]; a full list of datasets and references used here is provided in Supplementary Tables 2 and 3. We used the marker gene sets for each developmental stage to perform GSEA on the ranked list of genes that were up- or downregulated in the cellular studies of mice and humans. We also performed the analysis in reverse, comparing the gene sets of significantly differentially expressed mRNAs that were up- or downregulated in our cells versus the complete ranked list of differential gene expression in other studies. GSEA results are shown in Fig. 4e (mouse) and Fig. 4k (human). The readout is the NES. Data with $P < 0.05$ and $q < 0.05$ were considered to be significant and are indicated by asterisks in heat maps of GSEA NES scores.

ChIP–seq and genomic analyses. ChIP–qPCR was performed as described previously[84]; a list of the primers and antibodies used is provided in Supplementary Table 8. ChIP–seq was performed as described previously[2,3,68,84]. We performed six biological replicates for each condition (three conditions: serum/LIF, 2i and CDK8/19i) and for each antibody (three antibodies: anti-total RNA Pol II, anti-Ser 5P-RNA Pol II and control IgG). Three replicates were used for ChIP–qPCR validations, and the other three replicates were pooled for sequencing. Note that our RNA Pol II ChIP–seq data in this study for serum/LIF and 2i-naive cells very closely match previous ChIP–seq data involving the same comparison, that is, mouse PSCs in primed versus 2i-naive states[23,69] (compare Fig. 5h,i with Extended Data Fig. 6c).

Promoter and gene body regions were defined and RNA Pol II total and Ser 5P abundance along genes was calculated as described previously by Young and colleagues[68,84] (a schematic of which is provided in Fig. 5; Pol II abundance data are provided in Supplementary Table 6). RNA Pol II abundance was assessed by normalizing the total number of reads between treatments, and using featureCounts to calculate the background-subtracted $\log_2$-transformed RPKM of RNA Pol II abundance in the indicated regions. TSS and the transcription termination zone were identified using the Database of Transcriptional Start Sites (http://dbtss.hgc.jp). Metagenes were aligned to ±5 kb or ±2 kb around the TSS, and visualized using SeqMiner.

The promoter, gene body and transcription termination zone, as well as the ratios between these three regions for each gene (Fig. 5j, Extended Data Fig. 6d and Supplementary Table 6), were defined similar to that described previously[68,84]. Total and Ser 5P RNA Pol II abundance were quantified at the promoter, gene body and transcription termination zone for 31,167 RefSeq gene loci in which the transcription start and stop sites are known (Supplementary Table 6), in four steps, similar to previous reports[68]: (1) the number of reads per nucleotide was computed using BEDTools genomecov; (2) to extend this number to the number of reads per gene promoter or gene body, BEDTools map was used; (3) to correct for region size, the RNA Pol II abundance was calculated as follows: ((number of reads in region/region size) × scaling factor) × 10^5, where the scaling factor = (total number of reads in sample/genome length); (4) for the analysis of Pol II abundance according to inhibitor treatment, genes were first filtered for high-confidence Pol II detected at a threshold of >3,000 units at the promoter, and detected in all three conditions (serum/LIF, 2i or CDK8/19i), yielding 12,072 genes (Supplementary Table 6, for filtering and calculations). In Fig. 5h and Extended Data Fig. 6a,c, genes were arranged in rank by the abundance of RNA Pol II at the promoter region in the control serum/LIF condition.

CDK8/19 enrichment across the genome of wild-type mixed background V6.5 (C57BL/6–129) mouse PSCs was determined using a published dataset (GSE44286, GSM1082346) as previously described[2,3]; peak calling was performed with MACS v.1.4.1 using the standard settings and compared to the input negative control. Note that the ChIP antibody for this ChIP–seq analysis (Santa Cruz, sc-1521) is reported to bind to both CDK8 and CDK19 (ref. [89]). Peak annotation within local genomic features was performed using HOMER and the enhancer regions previously defined as constituent regions of TEs ($n = 9{,}981$) or SEs ($n = 646$)[2,3], and of SE extended regions ($n = 231$) as defined previously[2,3], where enhancers were defined by coenrichment for OCT4, SOX2, NANOG and MED1. Details about peak calls, CDK8/19 abundance at called peaks and loci annotations are provided in Supplementary Table 7. Naive-specific or primed-specific enhancer regions were defined by filtering the PREStige database of enhancers[77], which identifies enhancers by enrichment of H3K4me1 monomethylation in multiple tissues and lineages. Using the PREStige dataset, we identified enhancer regions with H3K4me1 enrichment of >20 units, and that were specific to only preimplantation naive PSCs or postimplantation EpiSCs versus all other tissue-specific enhancer regions listed in the database (~120,000), by subtracting overlapping enhancers (1 bp overhang threshold) as outlined in Extended Data Fig. 8e. Source data are available online, including lists of naive ESC-specific enhancers ($n = 1{,}424$) and EpiSC-specific enhancers ($n = 1{,}005$). To identify the single nearest target gene to each PSC SE and analyse their biological functions, we performed an analysis using GREAT v.3.0.0 (ref. [90]) with the standard settings, using the list of CDK8/19 peaks identified above (Supplementary Table 7). We used GREAT v.3.0.0 for GO analysis of target-gene functions, reporting the $-\log_{10}$-transformed binomial P value with conservative Bonferroni correction for multiple-hypothesis testing[90]. Correlation matrix of ChIP–seq data in Extended Data Fig. 7d was produced using Morpheus, which is available from the Broad Institute (https://software.broadinstitute.org/morpheus/).

For Fig. 7b,c, GSEA was run with a gene set of the single nearest genes to SEs (as identified in the GREAT analysis described above using the standard settings (GREAT v.3.0.0)[90], using the SEs that were previously described in mouse PSCs[2,3] versus the ranked list of differential gene expression determined by RNA-seq for serum/LIF control compared with CDK8/19i-adapted mouse ES PSCs. Source data are available online, including lists of SE-target and expression-matched control genes.

For Extended Data Fig. 8h, GREAT analysis using the standard settings (GREAT v.3.0.0)[90] was used to identify the set of single nearest genes ($n = 3{,}553$ genes) to enhancer regions that were previously identified in mouse PSCs ($n = 10{,}627$)[2,3]. The $\log_2$-transformed fold change in RNA expression of these genes from this study was then ranked from high to low (serum/LIF versus 2i; serum/LIF versus CDK8/19i), and the extent of the overlap, calculated using a hypergeometric test of significance of these two ranked lists, is shown as a heat map in Extended Data Fig. 8h, performed using RRHO[85] with the standard settings (http://systems.crump.ucla.edu/rankrank/rankranksimple.php). The colour intensity of the RRHO heat map indicates the $-\log_{10}$-transformed P value after Benjamini–Yekutieli correction of the hypergeometric overlap.

Statistics and reproducibility. Unless otherwise specified, quantitative data are presented as mean ± s.d. and significance was assessed using two-tailed Student's t-tests; $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. RRHO was performed as described previously[85] using the standard settings and after Benjamini–Yekutieli correction of the hypergeometric overlap. Except when annotated otherwise, each experiment shown was performed three times with similar outcomes. Statistical analyses are described in detail for each panel. No statistical methods were used to predetermine the sample size. In brief, for differential gene expression using RNA-seq analysis, a threshold of $q < 0.05$ or $q < 0.01$ was applied, as indicated in each case. For GSEA, the standard threshold for significance was applied ($P < 0.05$ and $q < 0.25$). Genes that were differentially expressed in the RNA-seq analysis were called using DESeq2 or Cufflinks v.1.3.0 (as described above). Immunofluorescence image analysis is described in detail above in the section about embryo analysis. Statistics were performed using MACS for peak calling of the ChIP–seq experiments. Statistical analyses of ChIP–qPCR, quantitative RT–qPCR and cell culture experiments was performed using Prism (v.7.03; GraphPad) or Microsoft Excel.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

RNA-seq and ChIP–seq data are available from the GEO database under accession numbers GSE112208 and GSE127186. The MS proteomics data are available from the ProteomeXchange Consortium/PRIDE repository under the dataset identifier PXD009200. Details on the published datasets used in Fig. 4e,k are provided in Supplementary Table 3. All other data supporting the findings of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

References

81. Diéguez-Hurtado, R. et al. A Cre-reporter transgenic mouse expressing the far-red fluorescent protein Katushka. *Genesis* **49**, 36–45 (2011).
82. Zalzman, M. et al. *Zscan4* regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).
83. Porter, D. C. et al. Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities. *Proc. Natl Acad. Sci. USA* **109**, 13799–13804 (2012).
84. Lynch, C. J. et al. The RNA polymerase II factor RPAP1 is critical for mediator-driven transcription and cell identity. *Cell Rep.* **22**, 396–410 (2018).
85. Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, e169 (2010).
86. Castro-Diaz, N. et al. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
87. Papadopoulou, T., Kaymak, A., Sayols, S. & Richly, H. Dual role of Med12 in PRC1-dependent gene repression and ncRNA-mediated transcriptional activation. *Cell Cycle* **15**, 1479–1493 (2016).
88. Xiang, L. et al. A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**, 537–542 (2020).
89. Broude, E. V. et al. Expression of CDK8 and CDK8-interacting genes as potential biomarkers in breast cancer. *Curr. Cancer Drug Targets* **15**, 739–749 (2015).
90. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

Acknowledgements

We thank A. Smith, T. MacFarlan, Z. Izsvák, M. Ko, J. Hanna, D. Grégoire and U. Hibner for gifts of reagents; N. Prats and L. César Fernández for assistance, as well as staff at the CNIO and IRB core facilities. M.N.S. was funded by a Leverhulme Trust early career fellowship. Work in the laboratory of M.Z.-G. was funded by the Wellcome Trust (098287/Z/12/Z) and the European Research Council (ERC) (669198). I.C. was funded by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and European Social Fund. I.A. and P.S. were supported by the Fondation pour la Recherche Medicale (DEQ20170336757), Infrastructure Nationale en Biologie et Santé INGESTEM (ANR-11-INBS-0009), IHU-B CESAME (ANR-10-IBHU-003), LabEx REVIVE (ANR-10-LABX-73), LabEx DEVweCAN (ANR-10-LABX-0061) and LabEx CORTEX (ANR-11-LABX-0042) of University of Lyon within the programme 'Investissements d'Avenir' (ANR-11-IDEX-0007). Research by J.P., S.M. and C.B.-A. was supported in part by a grant from the Spanish Ministry of Economy and Competitiveness (SAF2013-44267-R) and by the CNIO. Work in the laboratory of D.F. was funded by the Institut National du Cancer (PLBIO10-068 and PLBIO15-005) and the Ligue National Contre le Cancer (EL2018.LNCC/DF). Work in the laboratory of N.M. was funded by the ERC, under the European Union Horizon 2020 research and innovation programme (StG-2014–640525_REGMAMKID), the Spanish Association Against Cancer (AECC/LABAE16006), Carlos III Health Institute (Red TerCel, CardioCel, RD16/0011/0027), Ministry of Economy and Competitiveness (MINECO) projects SAF2017–89782-R, SAF2015–72617-EXP and RYC-2014–16242, and the CERCA/Government of Catalonia (2017 SGR 1306). Work in the laboratory of S.O. was funded by SAF2013–44866-R from MINECO Spain. Work in the laboratory of M.F.F. was funded by Plan Nacional de I+D+I 2013–2016/FEDER (PI15/00892, to M.F.F. and A.F.F.); the ISCIII-Subdireccion General de Evaluación y Fomento de la Investigación and Plan Nacional de I+D+I 2008–2011/FEDER (CP11/00131, to A.F.F.); IUOPA (to M.I.S.); and the Asturias Regional Government (GRUPIN14–052, to M.F.F.). The IUOPA is supported by the Obra Social Liberbank-Cajastur, Spain. Work in the laboratory of M.S. was funded by the CNIO, the IRB and by grants from Spanish Ministry of Economy co-funded by the European Regional Development Fund (SAF2017-82613-R), ERC (ERC-2014-AdG/669622), Botin Foundation, Banco Santander (Santander Universities Global Division), laCaixa Foundation and Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de Catalunya (Grup de Recerca consolidat 2017 SGR 282). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

C.J.L. designed and performed most of the experiments with mouse cells and embryos, contributed to bioinformatics data analysis and cowrote the manuscript. R.B. designed and performed most of the experiments with human cells, and provided general experimental support. A.M.-V. performed proteomic and bioinformatics analysis. M.N.S. performed embryo experiments, immunofluorescence and data analysis. S.N.-P., I.C., L.R.-G., N.A. and M.M.-M. contributed to experimental work and data analysis. C.T. and E.G. contributed to research with human PSCs and performed differentiation, immunofluorescence and confocal analysis of these experiments, supervised by N.M.; O.G.-C., G.G.-L. and C.S.-O.A., contributed to bioinformatics analyses. C.B.-A., S.M. and J.P. selected, synthesized and characterized small-molecule inhibitors. S.O. provided reagents, contributed to experimental design and supervised mouse embryo research. I.A. and P.S. performed human–rabbit interspecies chimaera and STAT3 assays. S.P., E.S., A.C. and D.F. generated the CDK8-KO mouse, provided reagents and performed additional inhibitor analyses. A.F.F., M.I.S. and M.F.F. performed DNA methylation analysis. P.S., D.F., J.M. and M.Z.-G. provided reagents, discussion and revisions. M.S. designed and supervised the study, secured funding, analysed the data and cowrote the manuscript. All of the authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

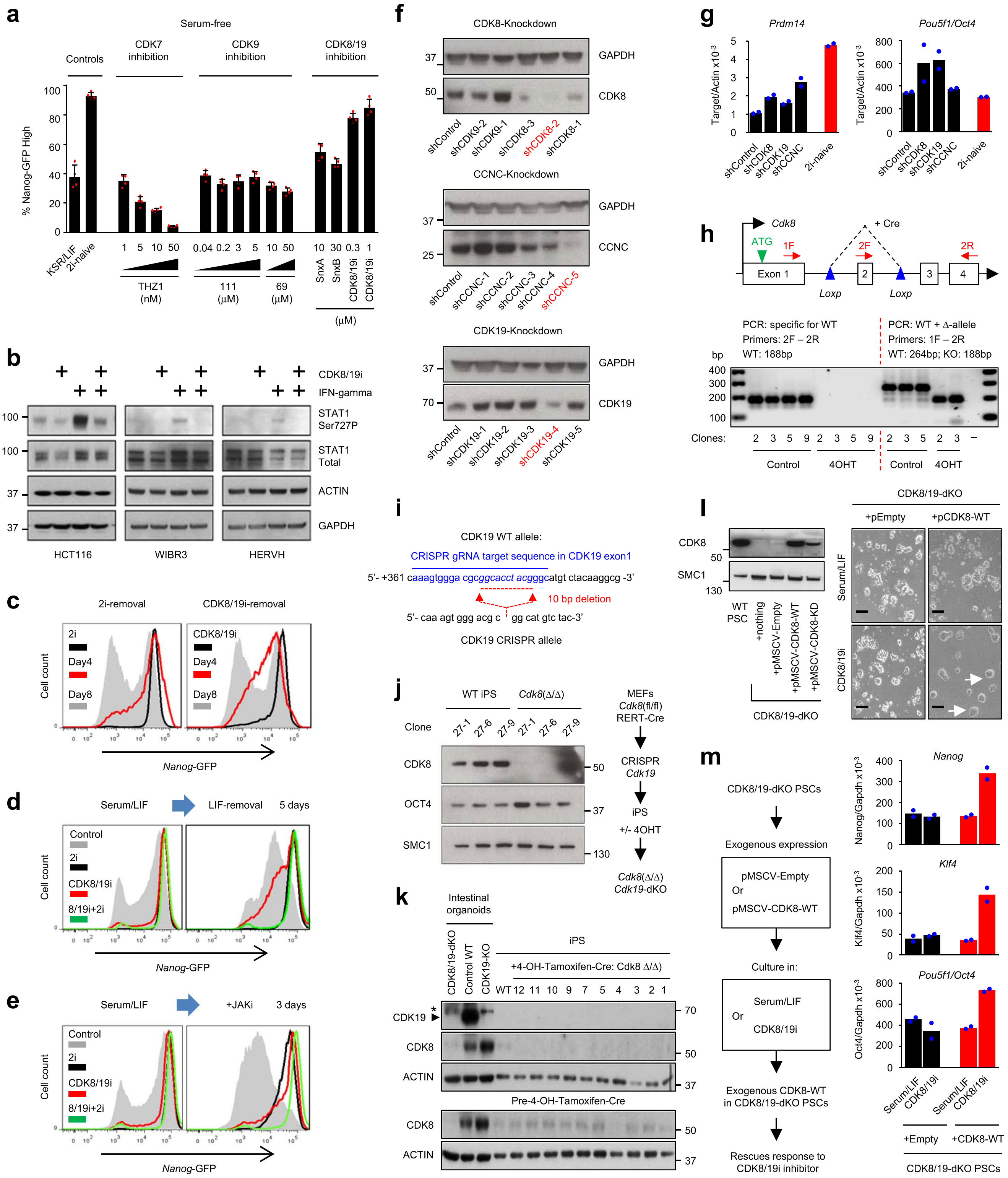
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-020-0573-1.

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-020-0573-1.

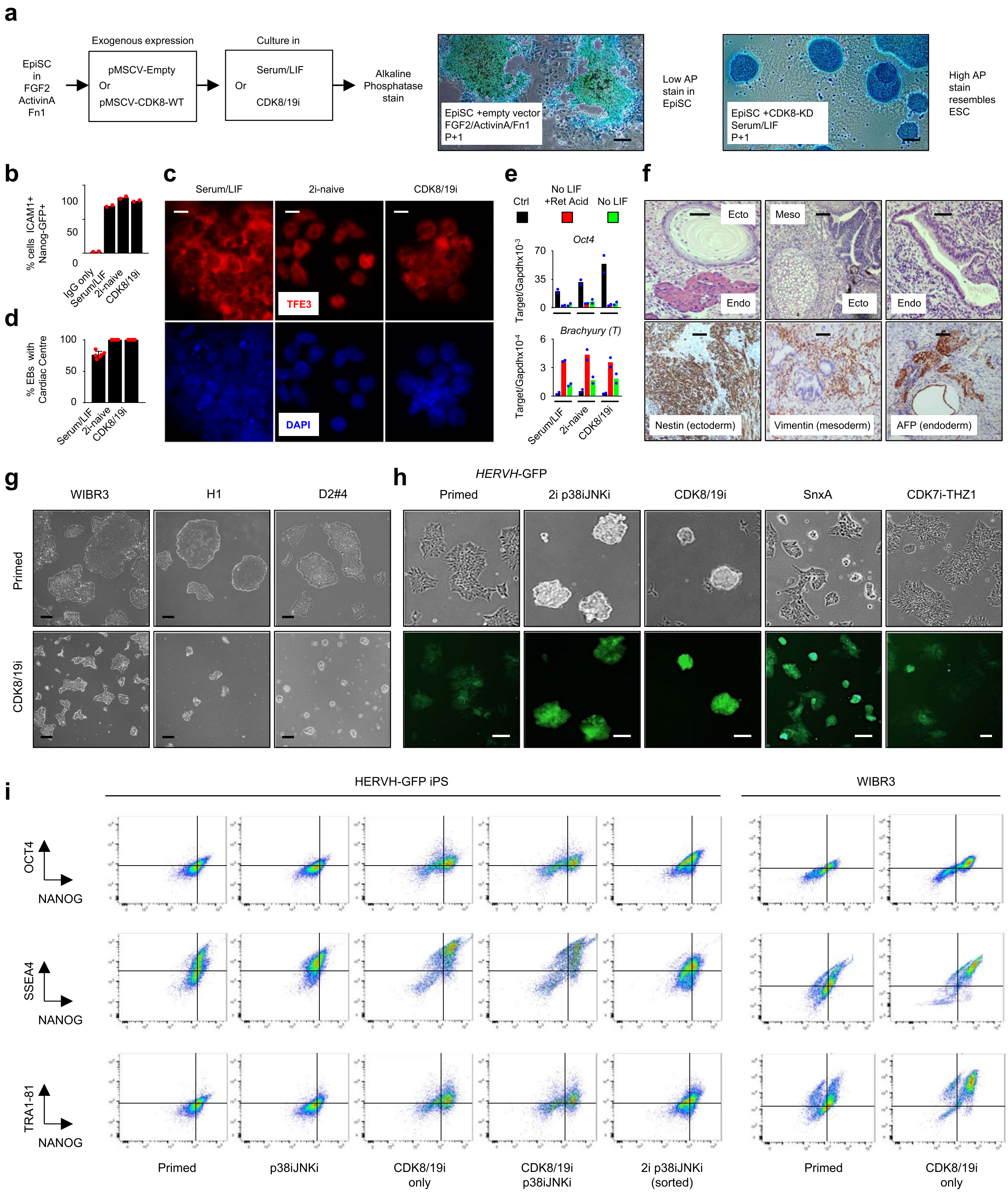
Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.



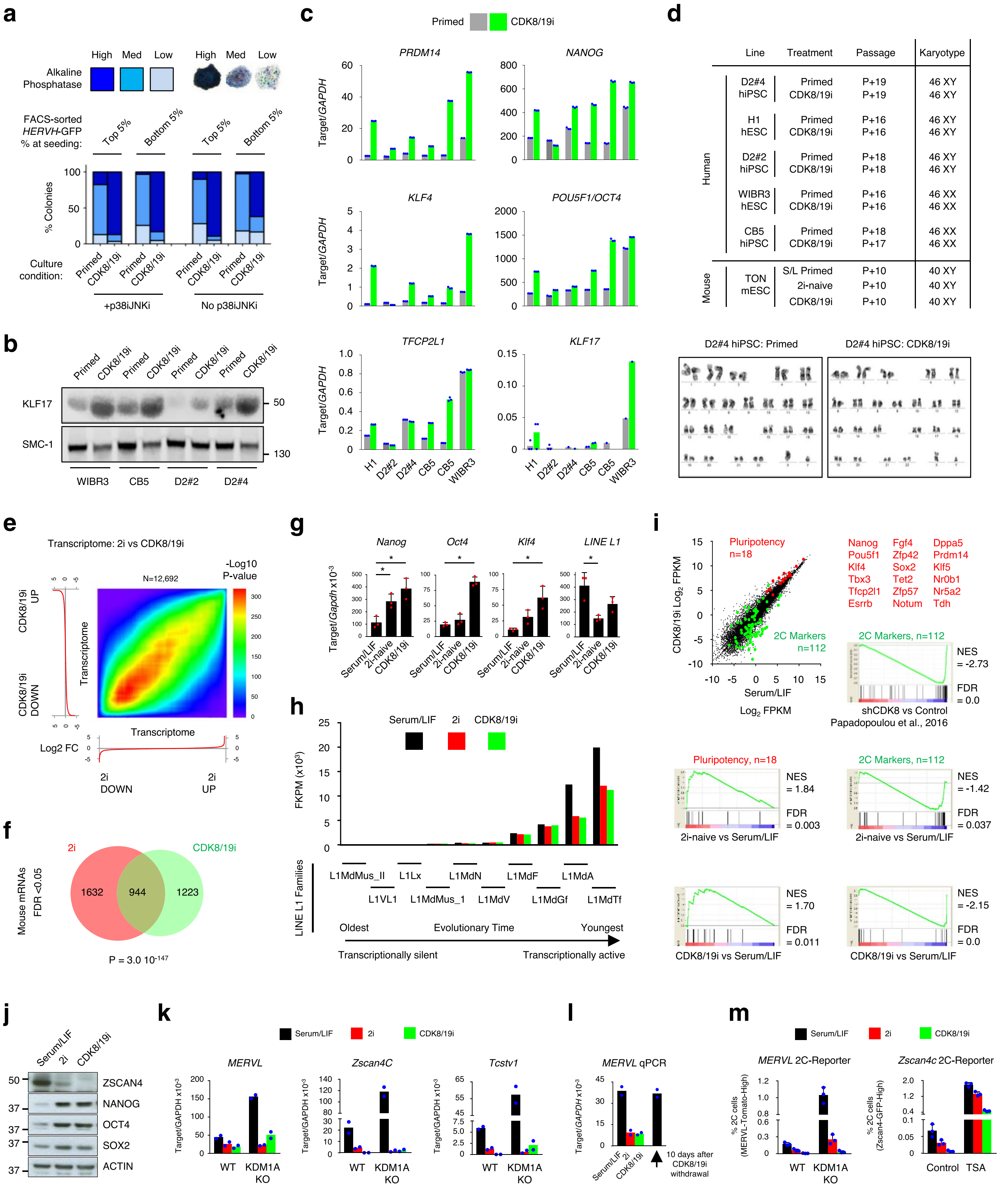
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | An Inhibitor Screen for Promotion of ES Naïve State identifies a distinct role for Mediator kinase activity. a, FACS: Percent cells Nanog-GFP^{high} in indicated treatments with serum-free KSR/LIF culture. Mean±SD, n=4 independent experiments. **b**, Western blot: CDK8-target STAT1-Ser727-Phosphorylation; n=3 experiments with indicated cell lines. STAT1-Ser727P induction by 3h Interferon-γ, ± simultaneous 1 μM CDK8/19i. **c**–**e**, FACS histograms: mouse *Nanog*-GFP knockin reporter PSC previously adapted to 2i or CDK8/19i, tested at intervals. Decreased proportion of *Nanog*-GFP^{High} cells indicates loss of naïve state. **c**, Changes similar by 2i-removal or CDK8/19i-removal. **d**, 2i protects Nanog-GFP^{High} cells longer than CDK8/19i following LIF removal. **e**, Only [CDK8/19i+2i] protects *Nanog*-GFP^{High} naïve-cells completely following JAK-STAT inhibition. Representative of n=2 experiments. **f**, Western blots: lentiviral shRNA-mediated knockdown of CDK8, CDK19 or Cyclin C (CCNC) in mouse PSC. Efficient shRNAs (red). Representative of n=2 experiments. **g**, Pluripotency marker mRNA expression (qRT-PCR) in mouse PSCs following 7d lentiviral shRNA-mediated knockdown of CDK8, CDK19, or CCNC. Data are the mean of two experiments. **h**, Upper schematic: inducible-CDK8-knockout. 4-hydroxy-tamoxifen (4OHT)- inducible Cre drives excision of Exon2. Lower: PCR confirmation of CDK8 Exon2 deletion using indicated primers. Mouse *Cdk8*(fl/fl)-RERT-Cre iPS (n=3–4 independent clones) treated 6d with 0.5 μM 4OHT. **i**, Indel mutation in one mouse CDK19-KO iPS clone, induced by indicated CRISPR guide-RNA against CDK19 Exon1, using lentiviral CRISPR-Cas9. Indel is 10 bp deletion at predicted CRIPSR target site, generating a frameshift immediately downstream of ATG start codon. **j**, Western blots: 4-OHT-inducible 8*Cdk8*-knockout mouse iPS clones as in (**h**). Schematic summarizes generation of these cells. Data representative of 4 experiments with n=3 iPS clones. **k**, Western blots: 4-OHT-inducible CDK8/19-double-knockout mouse iPS clones, generated as in (**J**). CDK8-knockout confirmed at protein level after 4-OHT-inducible Cre treatment. CDK19 protein was undetectable in PSCs, but readily detectable in intestinal organoid controls. Arrow indicates CDK19, confirmed by CRISPR-knockout as shown; *non-specific band. Data: CDK8/19-knockout with n=10 independent iPS clones. **l**, Left, Western blot: CDK8/19-double knockout (dKO) iPSCs ± empty-vector, catalytically-active CDK8 wild-type (WT), or CDK8 Kinase-Dead (KD). Right, bright field images: mouse iPS lines as indicated. Arrows: naïve-like colony morphology in cells expressing CDK8-WT plus treatment with CDK8/19i. Importantly, CDK8/19-dKO iPS with empty vector do not respond to CDK8/19-inhibitor. Images representative n=3 iPS clones. Scale bars 100 μm. **m**, Re-expression of CDK8-WT in null background rescues response to CDK8/19i. Left, overview. Right, mRNA expression of naïve pluripotency markers (qRT-PCR; Mean, 2 experiments) in CDK8/19-dKO mouse iPS with empty-vector, or catalytically-active CDK8-WT, in serum/LIF or CDK8/19i.



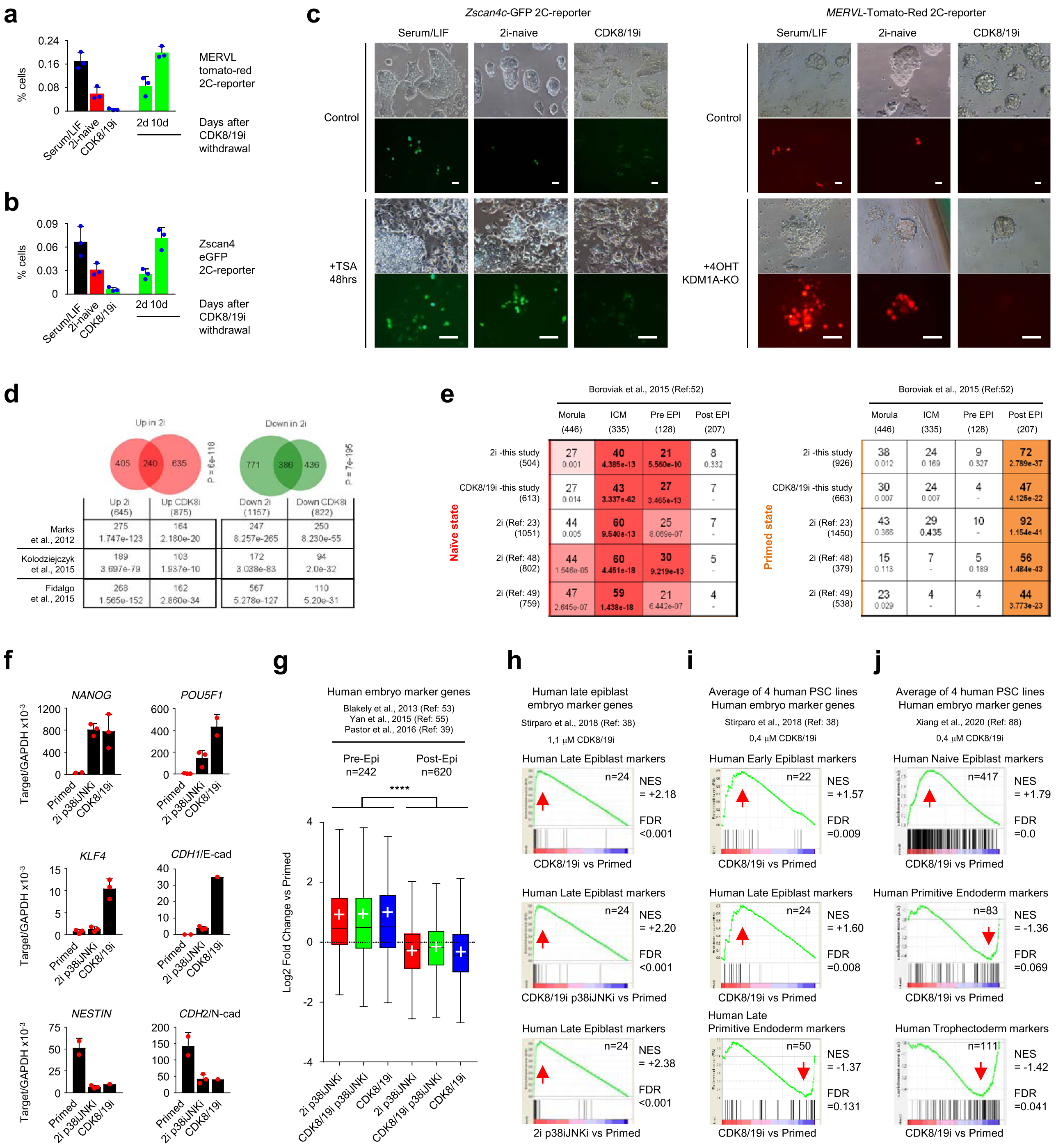
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Positive effect of long- term CDK8/19i on mammalian ES self- renewal and pluripotency. a, Cell morphology and alkaline phosphatase staining of WT mouse EpiSC infected with pMSCV-Empty or pMSCV-CDK8-KD, plus 7d/1 passage in either EpiSC media (Fgf2/ActivinA/fibronectin; Methods) or serum/LIF ES media (see schematic). Images representative of day 7 after pMSCV infection and selection, n=2 cell experiments. Scale bars 100 µm. **b**, FACS: maintenance of pluripotency markers in mouse PSCs. Percentage of double- positive Nanog-GFP+/ICAM1+ PSCs following 3 weeks culture in control serum/LIF, 2i- naïve, or CDK8/19i. Data: Mean, n=2 cell experiments. **c**, Immunofluorescence: TFE3 expression and localization in mouse PSCs adapted to control serum/LIF, 2i-naïve, or CDK8/19i, as in (**b**). Scale bar 10 µm. **d**, Embryoid body (EB) differentiation with beating cardiac centre demonstrates developmental capacity of mouse PSCs previously adapted to indicated conditions. Data: Mean±SD, n=3 cell experiments, two technical replicates each. **e**, Differentiation in vitro demonstrates developmental capacity of mouse PSCs previously adapted to control, 2i or CDK8/19i conditions. PSC differentiation was by LIF-removal or LIF-removal plus retinoic acid, and assessed by qRT-PCR to show loss of pluripotency (Oct4) and induction of differentiation (Brachyury, T). Data: Mean, 2 experiments. **f**, Teratoma assay in vivo demonstrates developmental capacity of mouse PSCs previously adapted to CDK8/19i conditions. Three embryonic germ layers confirmed in teratomas using histology (H+E stain; upper panels), and staining for germ layer markers: NESTIN (ectoderm), VIMENTIN (mesoderm), and Alpha-feto-protein (AFP, endoderm). Data representative of n=6 teratomas. Scale bar 200 µm. **g**, Brightfield images showing colony morphology in 3 human PSC lines in primed state (upper panels), or 14d treatment with CDK8/19i. Images representative n=5 independent experiments. Scale bar 100µm. **h**, Brightfield and live-cell GFP-fluorescence images of human iPS cells (HERVH-GFP reporter) in primed conditions, or following 14d treatment with indicated media cocktails. To derive and maintain the 2i p38iJNKi condition. Images representative of n=5 independent experiments. Scale bar 100 µm. **i**, FACS analysis of pluripotency markers in human PSCs (HERVH iPS or WIBR3 ES), following 3 weeks adaption to indicated culture conditions, as in (**g**, **h**). Data represent one experiment with n=2 independent PSC lines. Primed or CDK8/19i PSCs were routinely passaged in bulk using collagenase. In contrast, for the 2i p38iJNKi condition, cytometric sorting was required to select the top 10% HERVH-GFP cells at each passage, for 3 passages, before fixing the cells 4d after third passage/selection- round.



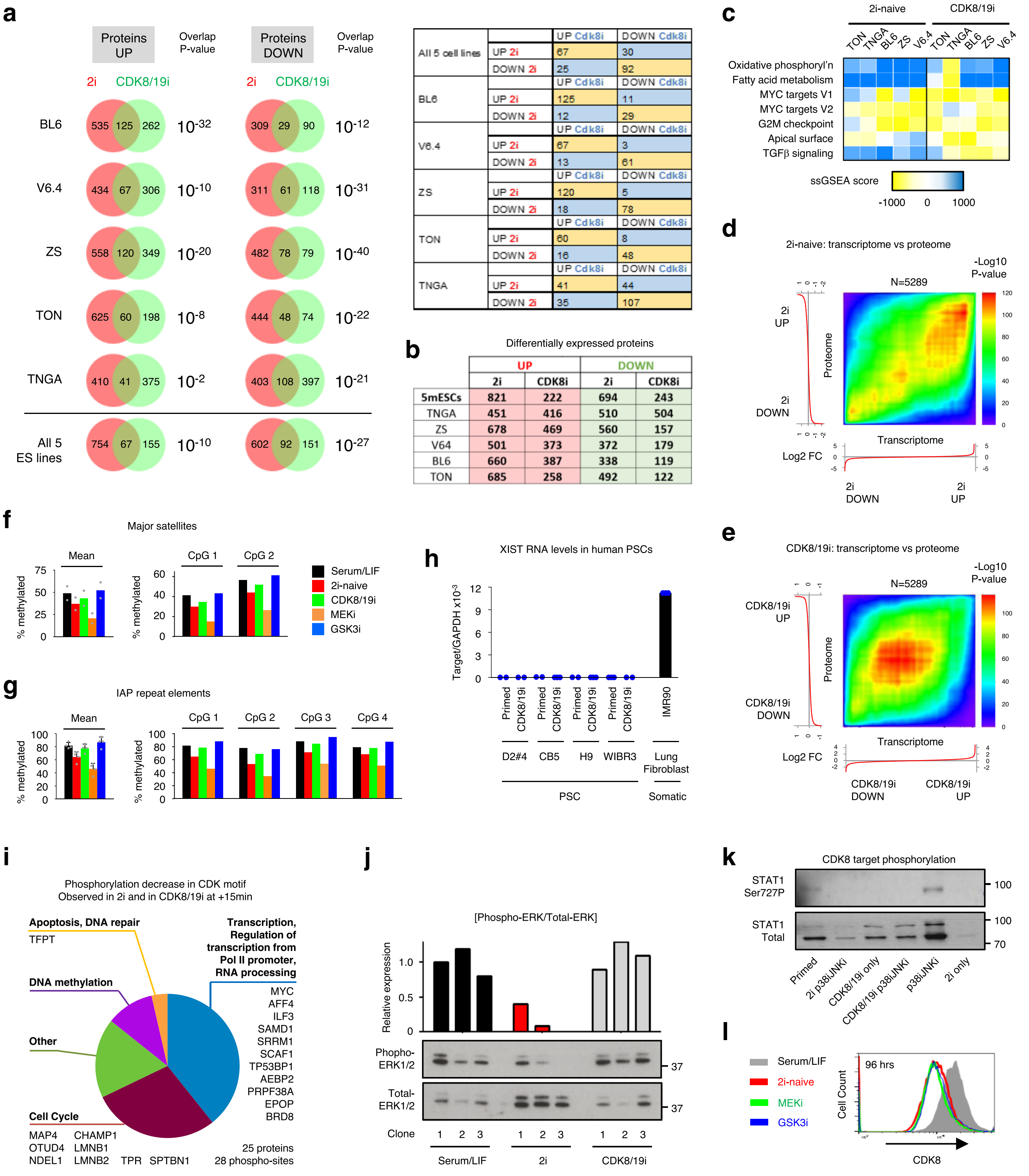
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Self-renewal, genomic stability, and gene expression analysis in mouse and human PSCs in CDK8/19i. a, Single cell clonogenicity: human PSCs in primed or CDK8/19i conditions. Cells were FACS-selected for top or bottom 5% HERVH-GFP intensity, seeded at clonal density in primed or CDK8/19i culture (±p38iJNKi) for 7d (individual colonies arise separately), then Alkaline phosphatase stained (inset, example colonies) to visually score maintenance of pluripotent status. Data: n=10 fields of view, multiple colonies per view, one experiment. **b**, Western blots: naïve pluripotency marker KLF17 in primed or CDK8/19i conditions. SMC1: nuclear internal loading control; n=4 human PSC lines. **c**, Mean pluripotency marker mRNA expression levels (qRT-PCR) in each of n=5 human PSC lines, in primed or CDK8/19i conditions >14d. **d**, Karyotyping indicates genomic stability: Human PSC lines (n=5), 16–19 passages in primed or CDK8/19i conditions. Inset, representative example: karyotype maintenance. **e**, Rank-Rank Hypergeometric Overlap (RRHO):[85] RNA-seq differential expression in mouse PSCs adapted to 2i-naïve or CDK8/19i conditions, versus serum/LIF (n=3 biological replicates; n=12,629 genes). Genes arranged by magnitude-change, then assessed for overlap by RRHO sliding window of 100 genes. Colour intensity: -log10 p- value after Benjamini-Yekutieli correction of hypergeometric overlap. **f**, Overlap and hypergeometric significance of differentially-expressed mRNAs in mouse PSCs in 2i or CDK8/19i, versus control KSR/LIF (serum-free conditions) RNA- seq; n=3 biological replicates; FDR<0.05. **g**, Pluripotency marker and LINE L1 repeat RNA expression (qRT-PCR) in mouse PSCs cultured as in (**e**). Data: n=3 experiments, Mean±SD, t-test, unpaired, two-tailed, *P<0.05. **h**, Effect of 2i or CDK8/19i on LINE L1 super-family expression (RNA-seq) in mouse PSCs cultured as in (**e**). Data: Mean, n=3 biological replicates, for each LINE L1 family, arranged by evolutionary age, which also reflects transcriptional activity and regulatory mechanisms[86]. 2i and CDK8/19i similarly regulate the youngest and most transcriptionally-active families (calculations, notes: Source Data). **i**, Dot plot: RNA-seq expression, in mouse PSCs, cultured as in (**e**). Pluripotency markers (red, n=18); 2C-fluctuation markers (green, n=112);[46,47,82] lists: Source Data). Below: effect of 2i or CDK8/19i (current study), or CDK8-knockdown[87], on these genesets. Significance: GSEA FDR q-values<0.25, indicated. **j**, Western blots, mouse PSCs. Markers of pluripotency, or 2C-fluctuation (ZSCAN4). Representative: n=2 experiments. **k**, **l**, RNA expression: 2C-fluctuation markers (qRT-PCR), mouse PSCs cultured as in (**e**), or after 10d CDK8/19i-withdrawal (**l**). Data: Mean, n=2 experiments. **m**, FACS quantification, percent fluorescence^{high} cells in 2C-fluctuation in two independent mouse PSC 2C-reporter lines, cultured as in (**e**). Induction of 2C- fluctuation: inducible-Kdm1a-knockout[46], or 48h TSA[46]. Data: Mean±SD, n=3 experiments.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Comparison of CDK8/19i and 2i in the current work versus published studies. a, **b**, FACS quantification, percent fluorescence^{high} cells in 2C-fluctuation: n=2 independent mouse PSC 2C-reporter lines[46,82], cultured as indicated. 2i and CDK8/19i repress the 2C-fluctuation, reversible by 2–10d inhibitor-removal. Data: Mean, n=3 experiments. **c**, Bright field or fluorescence^{high} cells in 2C-fluctuation: n=2 independent mouse PSC 2C-reporter lines[46,82], cultured as indicated. Induction of 2C-fluctuation: 7d inducible- Kdm1a-knockout[46], or 48h TSA[46]. **d**, **e**, Overlaps between RNAseq published datasets versus current study in mouse PSCs adapted to 2i or CDK8/19i versus control serum/LIF. Differentially expressed mRNAs in current study (FDR<0.01 and 2-fold change), versus three published studies[23,48,49] (**d**), or, versus developmental stage-specific marker genesets from preimplantation mouse embryos[52] (**e**). (D,E) Overlap hypergeometric significance, and number of genes changing in same direction, are reported below each 2-way comparison. **f**, Pluripotency (NANOG, POU5F1/OCT4, KLF4, CDH1/E-cadherin) or differentiation (CDH2, NESTIN) marker expression (qRT-PCR) in human PSCs adapted to indicated conditions. Mean±SD, n=1–3 biological experiments. **g**, Human embryo stage-specific developmental genesets (scRNA-seq), defining pre- implantation naïve epiblast (n=242 genes) and post-implantation primed epiblast (n=620 genes), can distinguish human PSCs between naïve and primed pluripotent states in vitro by up/down-regulation[39,53,55] (listed in: Source Data). RNA-seq expression of these genesets is shown in the current study in human PSCs adapted to indicated conditions. Tukey box plots; box reflects 25th -75th percentile; horizontal line is median; white-cross indicates mean. Data: n=3 biological replicates, Mean± SD, t-test, unpaired, two-tailed, ****$P < 0.0001$. **h**, **i**, Human embryo lineage-specific genesets[38] defining early/late pre-implantation naïve epiblast (n=22/24 genes), and late primitive endoderm (n=50 genes), were assessed by GSEA in our human PSC cultured in 2i-naïve or CDK8/19i. **h**, 1.1 μM CDK8/19i, one PSC line. **i**, 0.4 μM CDK8/19i, 4 PSC lines. Significance indicated by GSEA FDR q-values<0.25, and up/down-regulation (red arrows), in each panel. **j**, Human embryo lineage-specific scRNA-seq genesets[88], defining pre-implantation naïve epiblast (n=417 genes), primitive endoderm (n=83 genes), or trophectoderm (n=111 genes), were assessed by GSEA in our human PSC cultured in 0.4 μM CDK8/19i versus primed culture (n=4 PSC lines). Significance indicated by GSEA FDR q- values<0.25, and up/down-regulation (red arrows), in each panel.



		UP Cdk8i	DOWN Cdk8i
All 5 cell lines	UP 2i	67	30
	DOWN 2i	25	92
BL6		UP Cdk8i	DOWN Cdk8i
	UP 2i	125	11
	DOWN 2i	12	29
V6.4		UP Cdk8i	DOWN Cdk8i
	UP 2i	67	3
	DOWN 2i	13	61
ZS		UP Cdk8i	DOWN Cdk8i
	UP 2i	120	5
	DOWN 2i	18	78
TON		UP Cdk8i	DOWN Cdk8i
	UP 2i	60	8
	DOWN 2i	16	48
TNGA		UP Cdk8i	DOWN Cdk8i
	UP 2i	41	44
	DOWN 2i	35	107

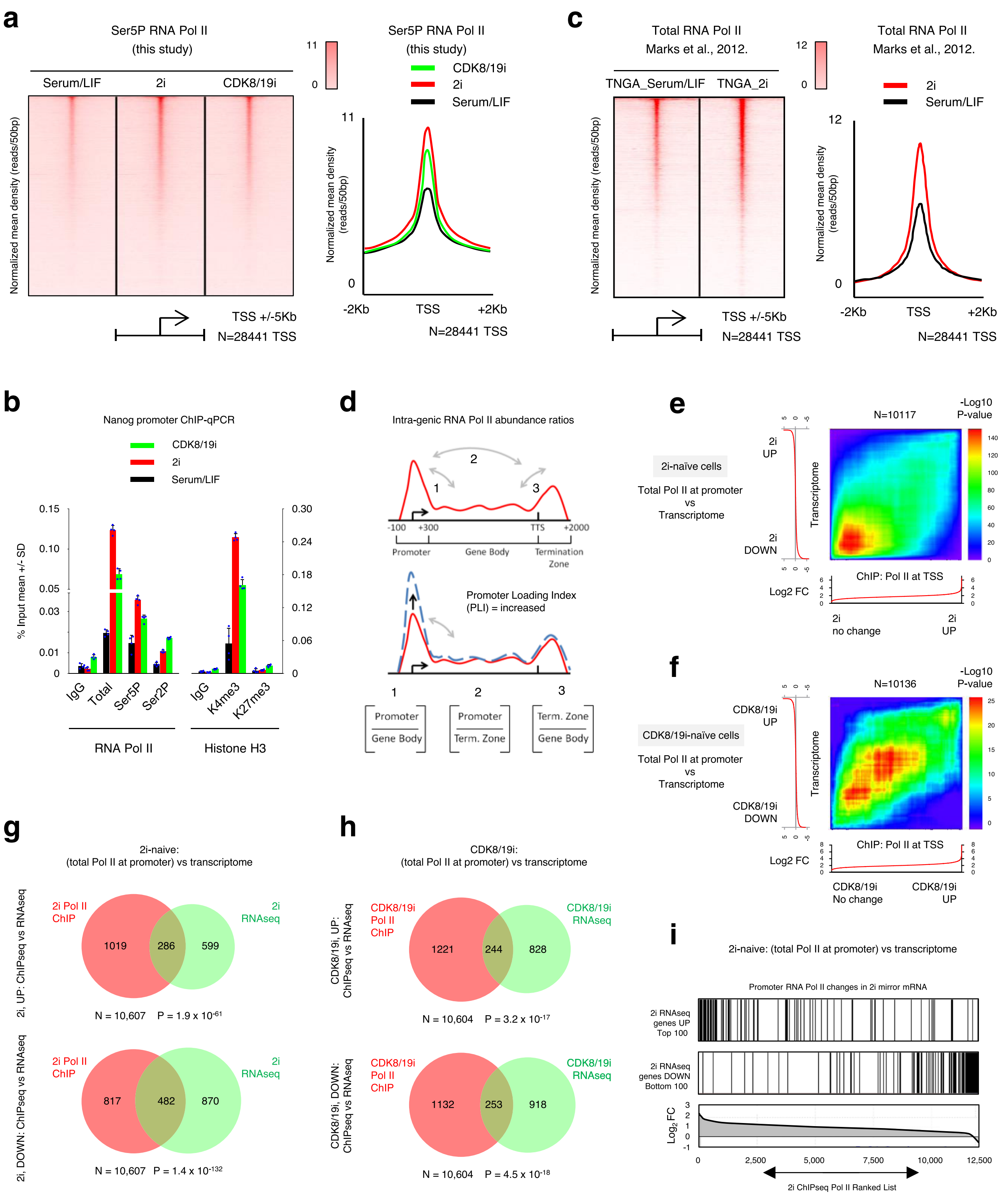
b

Differentially expressed proteins

	UP		DOWN	
	2i	CDK8i	2i	CDK8i
5mESCs	821	222	694	243
TNGA	451	416	510	504
ZS	678	469	560	157
V64	501	373	372	179
BL6	660	387	338	119
TON	685	258	492	122

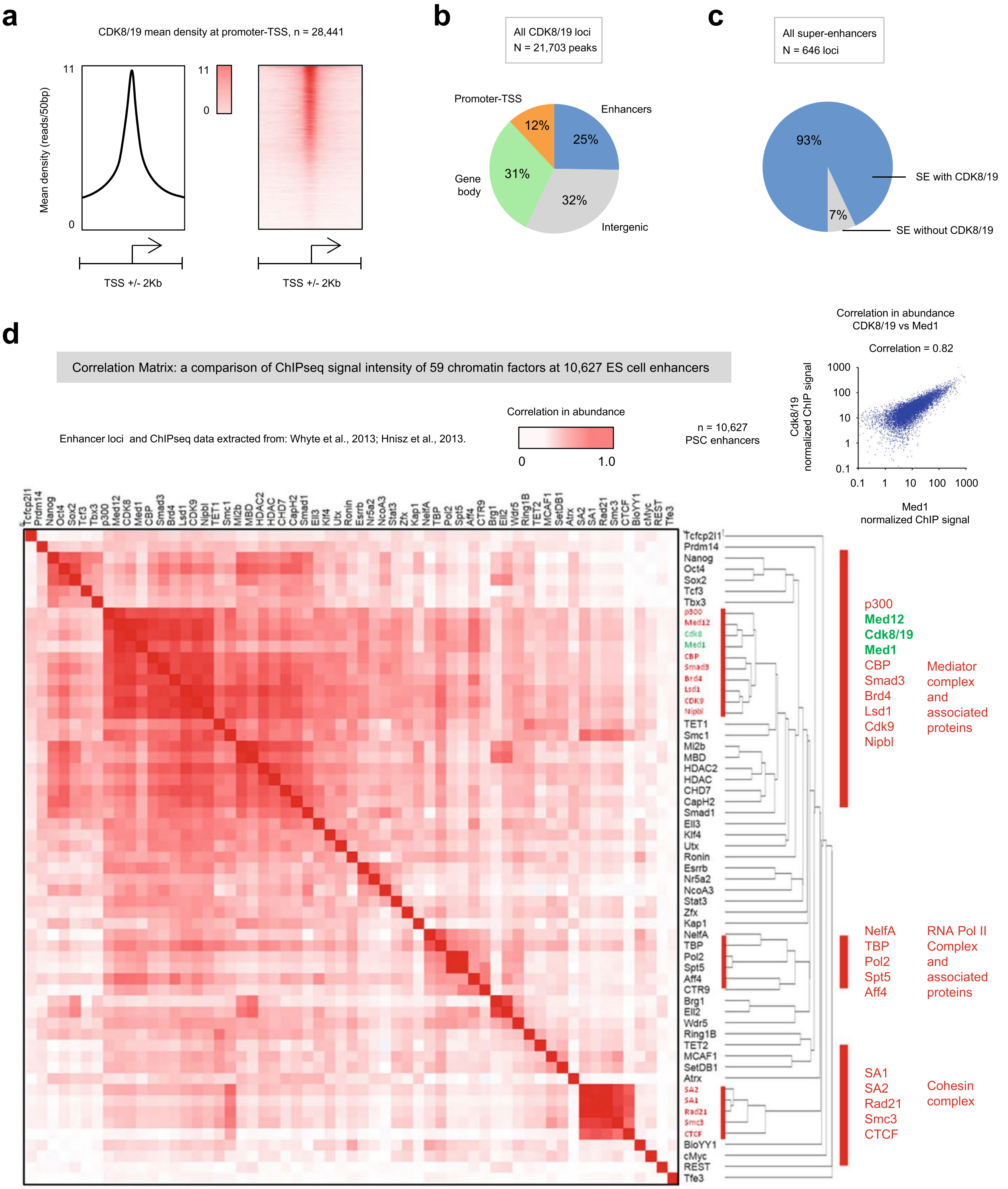
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | CDK8/19i regulates the phospho-proteome and proteome similar to 2i- naïve pluripotency, but not DNA methylation. a, Left: Overlap and hypergeometric significance (P-value) of differentially expressed proteins (FDR<0.05), in mouse PSC lines adapted to 2i-naïve or CDK8/19i, versus standard serum/LIF (n=5 lines, displayed individually). Right: table compares overlap in proteins up/down-regulated, per cell line and per condition, to highlight that the positive correlation (proteins changing in same direction) is greater than the negative correlation in all PSC lines. Supplementary Table 4: list of differentially expressed proteins. **b**, Summary, all proteomic changes, shown per mouse PSC line, adapted to 2i-naïve or CDK8/19i conditions, versus control serum/LIF, as in (**a**). **c**, Heatmap: normalized enrichment of biological pathways identified as significantly up/down-regulated (blue/yellow), by GSEA of proteomic changes, shown per cell line, in mouse PSCs cultured as in (**a**). Significance was confirmed in all pathways shown (GSEA FDR q-values<0.25). **d**, **e**, Rank-Rank Hypergeometric Overlap (RRHO):[85] differential mRNA expression (X- axis) in mouse PSCs adapted to 2i-naïve conditions (**d**) or CDK8/19i (**e**), versus, differential protein expression (Y-axis), for the same genes (n = 5289). Genes arranged by magnitude-change, then assessed for overlap by RRHO sliding window of 100 genes. Colour intensity: -log10 p-value after Benjamini-Yekutieli correction of hypergeometric overlap. **f**, **g**, DNA methylation changes (5-methyl-cytosine; pyrosequencing) in n=4 mouse PSC lines adapted to 2i or CDK8/19i. **f**, Methylation levels at two CpG sites in Major Satellite repeats, shown independently (right), or Mean±SD across the CpG loci (left). **g**, Methylation levels at four CpG sites in IAP repeats, shown independently (right), or Mean±SD across the CpG loci (left). **h**, XIST RNA levels in human PSC lines in this study (qRT-PCR). Female (n=3: WIBR3, CB5 and H9) and Male (n=1: D2#2) PSCs display low/undetectable XIST expression compared to control adult female human somatic cells (lung fibroblasts), suggesting X-silencing erosion may have already occurred in parental cells, as previously observed[67]. Data: n=3 technical replicates. **i**, Functional analysis of proteins containing a CDK phospho-target motif that displays phosphorylation decrease (FDR<0.05) within 15 min treatment of mouse PSCs with 2i or CDK8/19i. Data: n=2 PSC lines. **j**, Western blots: ERK1/2 phosphorylation after long-term adaption (3 weeks) of mouse PSCs to serum/LIF, 2i, or CDK8/19i. Above: relative ERK1/2 phospho-levels, normalized by total ERK1/2 levels. **k**, Western blots: CDK8 kinase-target STAT1-phospho-serine727, in human PSCs, with indicated culture media. **l**, CDK8 protein levels per cell measured by cytometry in mouse PSCs treated with indicated inhibitors. **j**–**l** Representative, n=2 independent experiments.



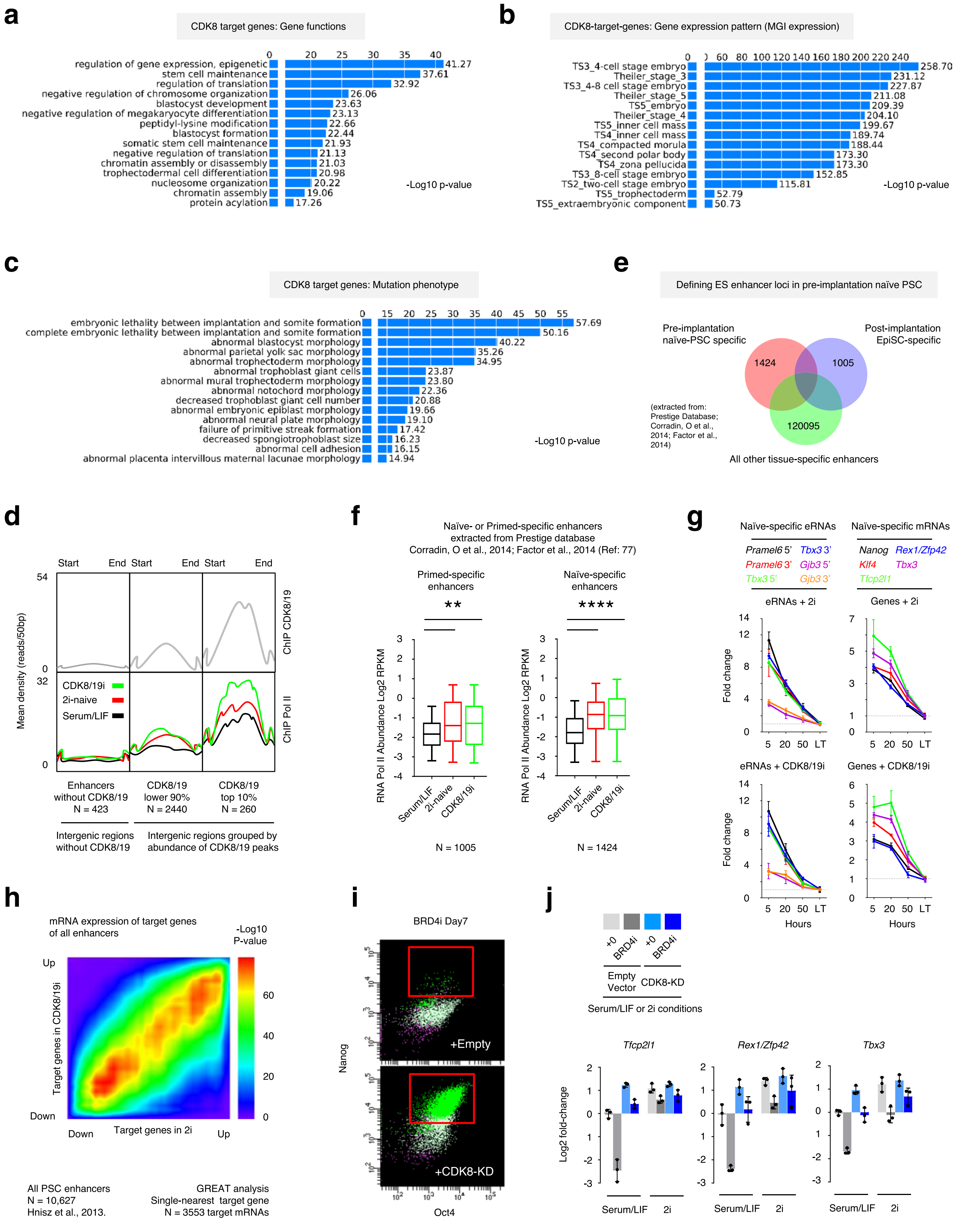
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Analysis of RNA Pol II genomic distribution and corellation with RNA gene expression. a, ChIP-seq: RNA Pol II Serine 5 phosphorylation (Ser5P) abundance at all Refseq Transcription Start Sites (TSS; n=28,441), in mouse PSCs, treated as indicated. Left: Heatmap, TSS +/−5 Kb. Right: Metagene average +/−2 Kb. ChIP-seq n=3 pooled replicates. **b**, ChIP-qPCR: RNA Pol II and histone marks at the Nanog TSS. RNA Pol II and H3K4me^3 (active euchromatin) are increased. Data: Mean±SD, n=4 ChIP replicates. **c**, Re-analysis of published[23,69] ChIP-seq: RNA Pol II abundance at all Refseq Transcription Start Sites (TSS; n=28,441), in mouse PSCs, treated as indicated, (similar conditions to current study, compare with Fig. 5h, i). Left: Heatmap, TSS +/−5Kb. Right: Metagene average +/−2Kb. ChIP-seq n=3 pooled replicates. **d**, Schematic: defining gene regions and Pol II loading ratios used in this study, similar to previous reports[68]. Lower panel: schematic summarizing results in Fig. 5h, i, where Promoter Loading Index is increased (Promoter/Body). **e**, **f**, Rank-Rank Hypergeometric Overlap (RRHO):[85] differential mRNA expression (RNA-seq data; Y-axis) in mouse PSCs adapted to 2i-naïve conditions (**e**; n=10,117) or CDK8/19i (**f**; n=10,136), versus, differential RNA Pol II abundance at promoter-TSS (ChIP-seq data; X-axis), for the same genes. Genes arranged by magnitude-change, then assessed for overlap by RRHO sliding window of 100 genes. Colour intensity: - log10 p-value after Benjamini-Yekutieli correction of hypergeometric overlap. **g**, **h**, Venn diagrams: genes with differential mRNA expression up/down-regulated (green circles; FDR<0.01) in 2i (**g**) or CDK8/19i (**h**) overlap significantly with genes where the promoter has the greatest/least change in RNA Pol II abundance (red circles; promoters with fold change > one standard deviation from mean). Overlap significance: hypergeometric test; P-values, and number of genes "n", indicated in each panel. Genes up (top Venn diagrams), and genes down (lower Venn diagrams), refer to inhibitor-treated cells versus control serum/LIF conditions. **i**, Genes with the greatest change in RNA Pol II abundance (lower panel; ranked list of promoters by magnitude of RNA Pol II abundance fold-change; n=12,693; ChIP-seq) correlate with the top 100 most differentially expressed mRNAs up/down-regulated in 2i-naïve conditions (upper two panels; RNA-seq). All changes refer to 2i-treated cells versus control serum/LIF conditions.



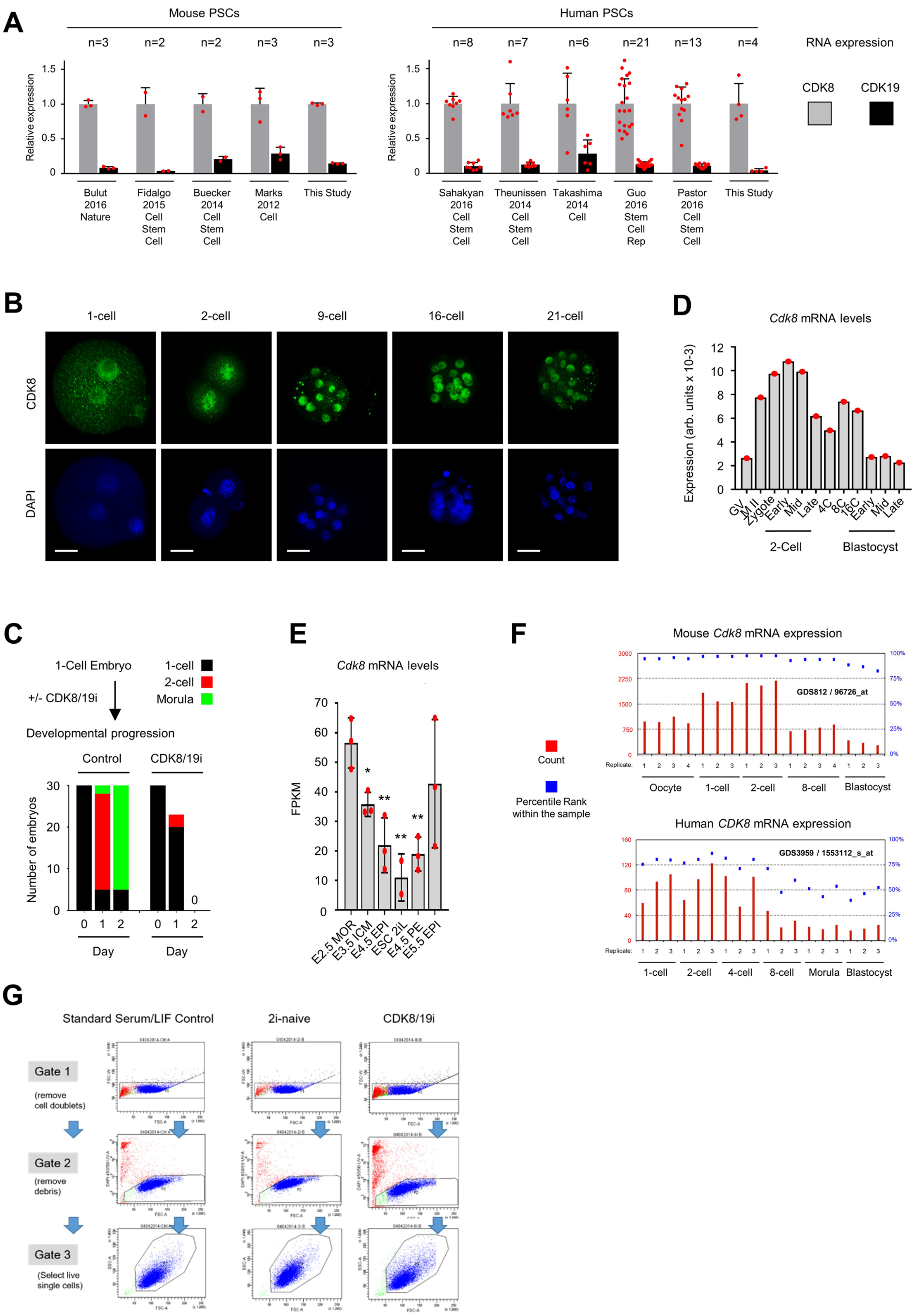
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | ChIP-seq for CDK8 and analysis of its genomic distribution. a, CDK8/19 average ChIP-seq enrichment[2,3] density in mouse PSCs at Promoter-TSS regions +/− 2 Kb, n=28,441 TSS (Refseq). **b**, CDK8/19 binding loci defined in mouse PSCs by ChIP-seq[2,3], MACS peak calling, and categorized by functional annotation of the region by HOMER (n=21,703, see Supplementary Table 7). Note: ChIP antibody binds both CDK8 and CDK19, see Methods. Promoter-TSS: TSS+/− 1Kb. Gene Body: Exons, Introns, and transcription termination site TTS +/− 1Kb. Enhancer constituent regions, as defined[2,3]. **c**, Percentage of SE-constituent regions[2,3] enriched for CDK8/19 binding (see also Supplementary Table 7). **d**, CDK8 is an integral part of Enhancer-Mediator in mouse PSCs. Pearson correlation Matrix summarizes correlation/co-occupancy between 59 factors in 10627 Enhancers in mouse PSCs, based on comparison of ChIP-seq signal intensity in published datasets. Enhancer loci and ChIPseq data extracted from[2,3]. The 59 factors indicated are a range of chromatin modifiers and transcription factors. Each square of the matrix represents a comparison between the corresponding pair of factors for their similarity in ChIP signal ranking across the 10,627 enhancer regions, to calculate a r^2 correlation of their similarity, where 1.0 = exactly similar. An example of a single correlation between two factors is shown for the Mediator subunit Med1 and CDK8/19 abundance within mouse PSC enhancers, in the upper-right of the panel. Hierarchical clustering groups those factors by similarity in ChIP signal pattern across all 10,627 enhancers. Thus, high correlation between two factors (red), indicates co- enrichment to similar levels and at the same set of enhancers, which is suggestive of functional co-operation. Co-enrichment patterns for subunits and co-factors of the Mediator, RNA Pol II and Cohesin complexes can be observed (indicated), consistent with their reported combinatorial roles at enhancers. CDK8/19 clusters most closely with the Mediator complex and other critical regulators of enhancer function. See Methods, Supplementary Table 7, and Source Data for analysis of the published ChIP datasets and enhancer loci defined by[2,3].



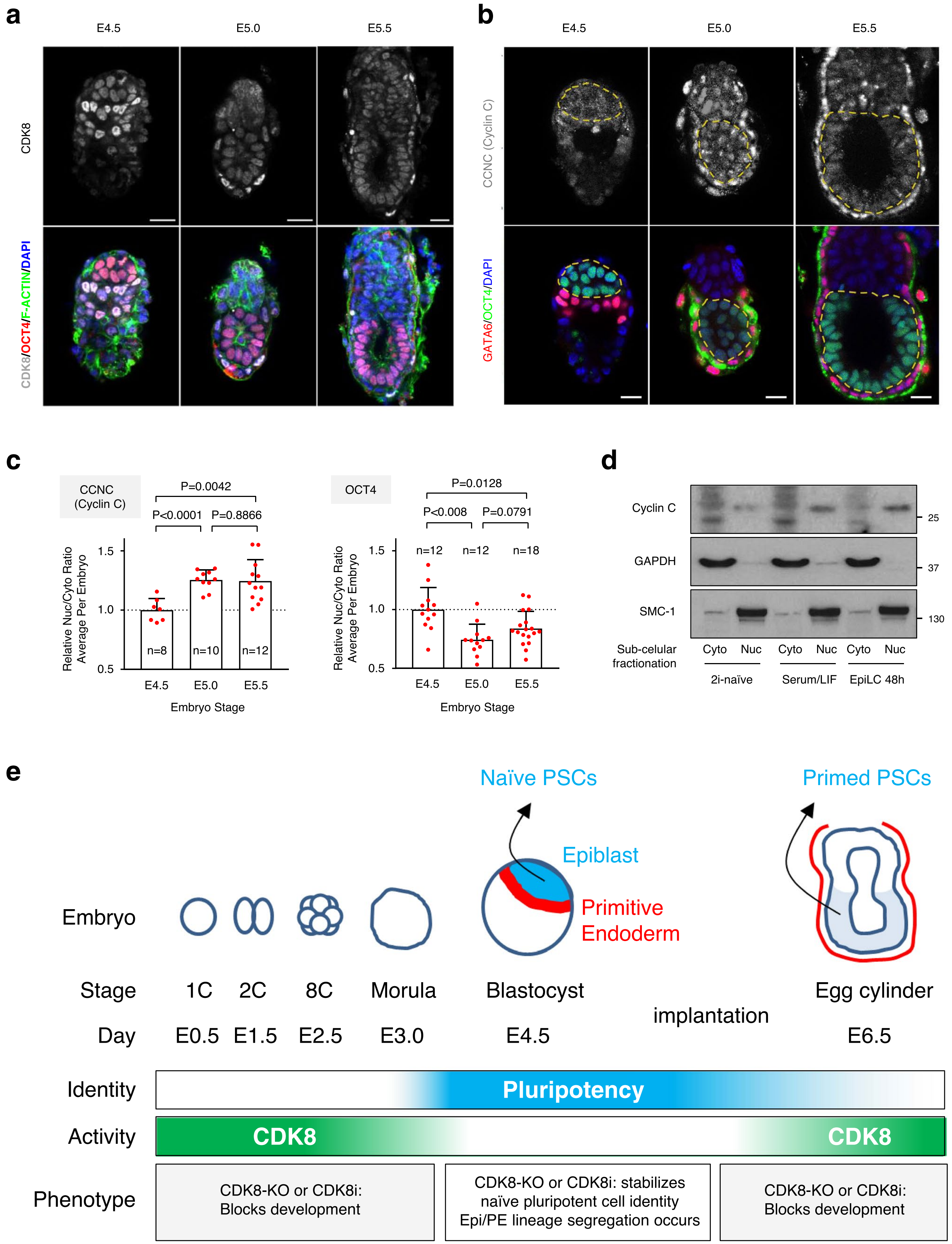
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | RNA Pol II and CDK8 genomic distribution. 2i and CDK8/19i hyper- actívate naïve-state enhancer activity. a–**c**, Gene ontology enrichment and functional annotation of CDK8/19-target genes identified by the single-nearest gene to each CDK8/19 binding site in mouse PSCs (ChIP-seq, n=21,703 peaks: Supplementary Table 7) using GREAT analysis[90]. Data: -log 10 binomial P-value with Bonferroni correction for multiple hypothesis testing indicating the significance of each gene ontology category. **d**, Metagene enrichment in the indicated genomic regions for CDK8/19 or RNA Pol II abundance, as determined by ChIP-seq in mouse PSCs above. Genomic regions were defined in groups by the CDK8/19 peak intensity as defined in (**a**). **e**, Identification of mouse PSC super-enhancer (SE) loci specific to pre-implantation naïve epiblast, or post-implantation primed epiblast. Enhancer loci were extracted from the Prestige Database[77]. The SEs in naïve or primed epiblast were first identified, and then any SEs common to a panel of 16 somatic tissues (threshold: 1 bp overlap) were subtracted (Methods). Enhancer loci lists: Source Data. **f**, RNA Pol II abundance in mouse PSC primed-specific super-enhancers (on left, n=1005), or naïve-specific super-enhancers (on right, n=1424), as defined in (**e**). RNA Pol II levels are significantly higher in 2i or CDK8/19i conditions versus serum/LIF control: t-test, unpaired, two-tailed, Welchs correction, $**P<0.01$; $****P<0.0001$). Tukey box plot centre lines show median values, box limits represent upper and lower quartiles, and whiskers show 1.5× interquartile range. **g**, qRT-PCR: pluripotency marker genes and naïve-specific eRNA[71] abundance in mouse PSC at short time intervals after withdrawal of 2i or CDK8/19i from culture. Data: Mean±SEM, n=3 experiments. **h**, Rank-Rank Hypergeometric Overlap (RRHO)[85]: heatmap shows differential mRNA expression (RNA-seq) of the single-nearest target genes (n=3,553; GREAT analysis) identified for all PSC enhancers[2,3] (n=10,627) in 2i-naïve conditions (X-axis) or CDK8/19i (Y-axis), compared to control serum/LIF conditions. The enhancer-target mRNA expression changes are arranged by magnitude, then assessed for overlap by RRHO sliding window of 100 genes. Colour intensity: $-\log_{10}$ *P* value after Benjamini- Yekutieli correction of hypergeometric overlap. Highly significant overlap along the diagonal indicates similar regulation of enhancer-target gene mRNA expression in 2i and CDK8/19i. **i**, FACS: NANOG and OCT4 protein expression following 7d treatment with 500 nM BRD4i(JQ1) in CDK8/19-dKO iPS clones expressing pMSCV-Empty or pMSCV-CDK8- Kinase-Dead (CDK8-KD). Representative of n=3 cell experiments. **j**, qRT-PCR: expression of naïve marker genes following 48h treatment with 500 nM BRD4i(JQ1). CDK8/19-dKO iPS ± CDK8-KD were cultured in 2i or standard serum/LIF, as indicated. Mean±SD, n=3 independent clones.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | CDK8 expression in vivo and the role of Mediator during mouse preimplantation development. a, CDK8 and CDK19 mRNA relative expression levels in PSCs, as detected by RNAseq in 5 mouse datasets[23,49–51] and 6 human datasets[32,34,39,58,66] (including the current study). Mean±SD of independent RNAseq replicates in each published study (see: n replicates indicated in panel; data calculation, see: Source Data). **b**, Immunofluorescence for CDK8 protein levels during mouse preimplantation development from 1-Cell to early blastocyst stage (E3.5), representing one set of embryos. Scale bars 25µm. **c**, CDK8/19-inhibition blocks embryo development at 1–2 Cell stage. Day E0.5 zygotes were harvested from females and immediately cultured in vitro in KSOM ± CDK8/19i for 2 days, with assessment of their developmental progression by visual inspection of cell number and morphology at intervals. Data represents n=30 embryos per condition, across two independent experiments. **d**, **e**, CDK8 mRNA expression levels in specific embryo stages and lineages during mouse preimplantation development. CDK8 mRNA expression declines until blastocyst stage, both in mouse and human pre-implantation embryos. In (**d**), mean data values from published microarray studies (Methods). In (**e**), CDK8 mRNA expression levels detected by RNA-seq in specific embryo stages and lineages during mouse preimplantation development; data from[52], Mean±SD, n=2–3 replicates per time point, significance assessed by one-way ANOVA unpaired T-test. *P<0.05; **P<0.01. (see Source Data). **f**, CDK8 mRNA expression levels during mouse or human embryo pre-implantation development, as detected by microarray in published datasets. Mouse: https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS812:96726_at Human: https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS3959:1553112_s_at (**g**) Schematic showing example of FACS gating strategy in this study. DAPI was added to live cell suspension 2mins before analysis, as a live/dead discriminator. Gates 1, 2, and 3, sequentially act to exclude cell doublets, and debris, thus selecting live single cells for analysis of Nanog-GFP profile in mouse PSCs.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Cyclin C expression localization during mouse preimplantation development. a, Immunofluorescence: CDK8, OCT4, and F-ACTIN in mouse early embryos from E4.5 to E5.5. Scale bars 20 µm. Images representative of n=3 experiments. **b**, **c**, Immunofluorescence (**b**): cyclin C protein levels during mouse development from preimplantation blastocyst stage (E4.5) to post-implantation cylinder stage (E5.5). Co-staining with OCT4 to mark epiblast, and GATA6, to mark primitive endoderm at E4.5 and its maturation into post-implantation visceral endoderm. **c**,, Cyclin C nuclear-cytoplasmic ratio was quantified and plotted, where each data point represents the mean Nuc-Cyto ratio for epiblast cells of one embryo (Source Data). As internal control, the Nuc-Cyto ratio for OCT4 was also quantified. Nuclear abundance of cyclin C increases in epiblast cells during development from E4.5 to E5.5. In contrast, OCT4 nuclear-cytoplasmic ratio does not follow this pattern. This implies that cyclin C pattern is not related to staining or imaging artefacts. **b**,, representative examples shown; n=2 experiments. **c**,, data: Mean±SD, T-test, unpaired, two-tailed, P-values and number of embryos "n" is indicated. **d**, Western blot: cyclin C localization by sub-cellular localization. Nuclear and cytoplasmic fractions were prepared from mouse cells across developmental spectrum: naïve (adapted to 2i), primed (adapted to serum/LIF), or epiblast-like stem cells (abbreviated EpiLC; derived by treating PSCs for 48 h with EpiSC media[50]) (Methods). Relative abundance of nuclear cyclin C is greater in primed state EpiLC and in serum/LIF conditions, compared to 2i-naïve. Data represent n=2 experiments. **e**, Summary: The developmental requirement for CDK8 activity mirrors its embryonic expression pattern. Maxima in CDK8 expression coincide with a requirement for development around the zygote-morula and post-implantation stages. Between these two periods, a transient minima in CDK8 expression occurs during emergence of naïve epiblast, where CDK8 function appears dispensable. We suggest that CDK8/19 chemical inhibition in vitro mimics CDK8 downregulation during pre-implantation development in vivo, providing mechanistic insight on how naïve pluripotency may arise in the embryo: (i) CDK8/19 is required during zygote-to-morula development, where its expression is high. (ii) During morula-to-blastocyst pre-implantation development, CDK8 expression declines, and nuclear cyclin C decreases. This coincides with the emergence of E4.5 pre-implantation naïve epiblast and, accordingly, CDK8/19 inhibition does not interfere with naïve epiblast specification. In contrast to MEK inhibition, CDK8/19 inhibition does not affect the epiblast/PE lineage segregation. (iii) During the subsequent developmental transition of pre-implantation naïve epiblast to post-implantation primed state, CDK8 expression becomes increased and CDK8/19 activity is required for morphogenic events.

nature research

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.