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

Cian Lynch, Raquel Bernad, Ana Martínez-Val, Marta Shahbazi, Sandrina Nóbrega-Pereira, Isabel Calvo, Carmen Blanco-Aparicio, Carolina Tarantino, Elena Garreta, Laia Richart-Ginés, et al.

Global hyperactivation of enhancers stabilizes human and mouse naïve 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], Sandrina Nóbrega-Pereira[5], Isabel Calvo[1,2], Carmen Blanco-Aparicio[6], Carolina Tarantino[7], Elena Garreta[7], Laia Richart-Ginés[8], Noelia Alcazar[1,2], Osvaldo Graña-Castro[9], Gonzalo Gómez-Lopez[9], Irene Aksoy[10], Maribel Muñoz-Martín[1,2], Sonia Martinez[6], Sagrario Ortega[11], Susana Prieto[12], Elisabeth Simboeck[12], Alain Camasses[12], Camille Stephan-Otto Attolini[13], Agustin F. Fernandez[14], Marta I. Sierra[14], Mario F. Fraga[14], Joaquin Pastor[6], Daniel Fisher[12], Nuria Montserrat[7,15], Pierre Savatier[10], Javier Muñoz[3],

Magdalena Zernicka-Goetz[4] and Manuel Serrano[1,2,16,*]

[1] Tumour Suppression Group
Spanish National Cancer Research Centre (CNIO)
Madrid 28029, Spain

[2] Cellular Plasticity and Disease Group
Institute for Research in Biomedicine (IRB Barcelona)
Barcelona Institute of Science and Technology (BIST)
Barcelona 08028, Spain

[3] ProteoRed-ISCIII Proteomics Unit,
Spanish National Cancer Research Centre (CNIO)
Madrid 28029, Spain

[4] Zernicka-Goetz Lab
Mammalian Embryo and Stem Cell Group,
University of Cambridge,
Department of Physiology, Development and Neuroscience,
Downing Street,
Cambridge CB2 3EG, UK

[5] Instituto de Medicina Molecular João Lobo Antunes
Faculdade de Medicina
Universidade de Lisboa
Lisboa 1649-028, Portugal

[6] Experimental Therapeutics Program,
Spanish National Cancer Research Centre (CNIO)
Madrid 28029, Spain

[7] Pluripotency for organ regeneration. Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute for Science and Technology (BIST), Barcelona, Spain.

[8] Epithelial Carcinogenesis Laboratory,
Spanish National Cancer Research Centre (CNIO)
Madrid 28029, Spain

[9] Bioinformatics Unit
Spanish National Cancer Research Centre (CNIO)
Madrid 28029, Spain

[10] Univ Lyon, Université Lyon 1, INSERM U1208,
Stem Cell and Brain Research Institute,
69500 Bron, France

[11] Transgenic Mice Unit
Spanish National Cancer Research Centre (CNIO)
Madrid 28029, Spain

[12] Laboratory of Nuclear Control of Cell Proliferation
Institute of Molecular Genetics of Montpellier
1919 Route de Mende
34293 Montpellier, Cedex 5, France

[13] Bioinformatics-Biostatistics Unit,
Institute for Research in Biomedicine (IRB Barcelona)
Barcelona Institute of Science and Technology (BIST)
Barcelona 08028, Spain

[14] 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

[15] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Barcelona, Spain.

[16] Catalan Institution for Research and Advanced Studies (ICREA)
Barcelona 08010, Spain

* *Correspondence*: manuel.serrano@irbbarcelona.org

Abstract

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

(149 words, = limit 150)

Each cell-type contains a unique repertoire of active enhancer complexes at specific DNA regions, which arise by high concentration of lineage-specific transcription factors and cell signaling pathways [1–3]. The Mediator complex is highly enriched at enhancers and it is thought to integrate multiple signals, eventually leading to the recruitment of RNA polymerase II (RNA Pol II) to nearby and distant promoters thus having a major contribution to the transcriptional program characteristic of each cell type [4–8]. A large fraction of cellular Mediator and transcriptional machinery clusters within a small number of unusually long enhancers, known as super-enhancers (SEs)[3]. SEs maintain high expression of the master transcription factors driving cell identity, yet SEs also possess inherent sensitivity to perturbation [2,3,9]. Given Mediator´s central role in enhancer-driven transcription, modulation of its activity may influence cell identity transitions [9–11]. Indeed, global enhancer activation was recently identified as a common feature across multiple human cancers [12]. Pluripotent stem cells (PSC) provide a prototypical model of cellular plasticity, whose transcriptional program can be stabilized, extinguished or re-captured [13]. Here, we manipulate Mediator function in pluripotent stem cells, in order to elucidate the transcriptional basis of their plasticity.

There have been great advances in recent years regarding the structure and mechanistic functioning of Mediator[5,8,10,14–16]. The 30 subunits of Mediator are organized in four general domains: the “head”, “middle”, and “tail” domains that constitute “core-Mediator”, plus a fourth accessory domain known as the “CDK8-module”. This module contains the only enzymatic activity of Mediator, namely the kinase CDK8 or its highly similar, but poorly studied, paralog CDK19, together with their binding partner, cyclin C (CCNC), essential for kinase activity. The CDK8/19-kinase module associates with core-Mediator and it is a negative regulator of RNA Pol II recruitment [17–21]. Negative regulation of Mediator by CDK8/19 may involve its kinase activity or sterically hindering the association between Mediator and RNA Pol II [5,15,20–22]. Supporting the negative role of the kinase activity, it has been shown that chemical inhibition of CDK8/19 results in global hyperactivation of enhancer function in the context of cancer cells [17]. In relation to this, CDK8/19 phosphorylates multiple Mediator subunits [20,22,23]. Additional layers of complexity

derive from the fact that CDK8 can phosphorylate the RNA Pol II C-terminal regulatory domain [24], chromatin regulators, and specific transcription factors [8,19]. Of note, CDK8-dependent phosphorylation of transcription factors often leads to their degradation [25–30].

While the basic principles of PSC identity were initially established in mice, our understanding of pluripotency across mammalian species is in progress [13,31]. In particular, capturing and stabilizing therapeutically-useful human PSCs has been challenging [32,33]. Therefore, a major current focus is to identify common principles, and accommodate apparent differences, between mouse and human pluripotency. Optimal culture conditions to shield mouse ES cells from extra-cellular differentiation stimuli involve chemical inhibition of MEK and GSK3 kinases with a two inhibitor cocktail known as "2i" [34,35]. Mouse ES cells cultured in 2i (referred to as "2i-naïve" cells) phenocopy the stable and homogenous state of undifferentiated naïve pluripotency which exists transiently *in vivo* in the E4.5 pre-implantation embryo epiblast. In contrast, culture of ES cells in the absence of 2i triggers a shift in cell identity towards post-implantation epiblast, also known as the primed state [36–38]. Enhancer destabilization is well known to trigger loss of Mediator-driven gene expression in many cell types, and in primed PSCs, this induces differentiation [39–42]. Remarkably, 2i-naïve PSCs were recently found to be highly resistant to enhancer destabilization by BRD4 inhibition [43], indicating that a defining feature of the naïve state may involve distinct enhancer regulatory mechanisms. Within the 2i cocktail, MEK-inhibition has been implicated as the key catalyst of these effects on PSC identity [34,43–45]. MEK inhibition in PSCs is associated with potent and rapid reconfiguration of the transcriptome, proteome, and DNA methylome [36–38,46]. However, it remains unclear which are the critical steps for the stabilization of the naïve state. Moreover, chemical cocktails based on MEK inhibition appear less successful in stabilizing naïve human PSC identity compared to mouse [32,33]. Thus, better mechanistic understanding of the naïve state is desirable to stabilize human stem cell pluripotency.

Here, we investigate a specific role for Mediator-kinase activity in PSC identity. To this end, we assess the impact of CDK8/19-kinase inhibition on

molecular parameters that characterize PSC identity and compare them to 2i-induced naïve pluripotency. We document the effects of CDK8/19i *versus* 2i in terms of immediate phospho-proteomic changes, enhancer activity, RNA Pol II genomic distribution, mRNA transcriptome, proteome, cellular morphology, self-renewal and developmental capacity. The role of CDK8/19 *in vivo*, during mouse pre-implantation, is also explored. In summary, targeting Mediator through its kinase module selectively stabilizes an early pluripotent cell identity, repressing differentiation, favoring self-renewal, and up-regulating pre-implantation naïve epiblast gene expression patterns in mouse and in human.

Inhibition of Mediator kinase stabilizes mouse naïve pluripotency

GFP knock-in reporters at key stem cell marker genes such as *Nanog* represent well-established and precise indicators of the naïve and primed states [47,48]. For example, in 2i-naïve state, *Nanog* promoter activity is enhanced, yielding a homogenous high *Nanog*-GFP expression pattern and uniform dome-shaped colonies across the cell population. In contrast, the *Nanog* promoter is metastable in primed state ES cells, reversibly oscillating between high and low activity. This is associated with a heterogeneous *Nanog*-GFP expression pattern and flattened diffuse colonies in the primed cell population, indicative of a general underlying switch in transcriptional program [38,44,47,48].

Here, we employed the *Nanog*-GFP knock-in system to distinguish the primed (GFP^{low}) or naïve (GFP^{high}) states [47]. As a positive control, treatment with 2i induced a characteristic shift of the culture into a homogeneous GFP^{high} naïve state with uniform colonies (**Figure 1A-C, and S1A**). Conversely, as a negative control, we used the BRD4 inhibitor JQ1 to evict Mediator from enhancers [41,43], and as reported, this triggered GFP^{low} status, colony dispersion, and rapid differentiation [39,40,42,43]. 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. Of note, while CDK8/19 act in a Mediator-dependent manner, the other transcriptional CDKs play general roles in RNA Pol II promoter escape and elongation, in particular, CDK7 as part of TFIIH, and CDK9 as part of the pTEFb complex. Specific inhibitors of CDK7 or CDK9 did

not increase the ratio of *Nanog*-GFP^{high} mouse PSCs (**Figure 1A and S1A**). In contrast, however, several potent and structurally-unrelated CDK8/19 inhibitors had a clear positive effect, inducing the formation of homogenous dome-shaped colonies, and up-regulating both the *Nanog*-GFP reporter and endogenous *Nanog* expression, similar to PSC in the 2i-induced naïve state (**Figures 1A-E and S1A**; see also: **Table S1**). Potency and selectivity of CDK8/19-inhibitors, commercially available or developed in-house, were assessed by multiple methods: (i) selectivity was suggested by KinomeScan, which includes a panel of 456 kinases (see **Table S1**); (ii) Lanthascreen assays demonstrated inhibitory activity at nanomolar concentrations against pure recombinant CDK8/CCNC and CDK19/CCNC (see **Table S1**); (iii) luciferase reporter cell assays (TOP-FLASH) (see **Table S1**); and (iv) potent inhibition of STAT1 Ser727 phosphorylation in human PSCs, a well-documented CDK8 target site [17,22,26,49] (**Figures 1F and S1B**). Based on these data, we focused on the CNIO molecule ETP-47799, which was the most effective at improving mouse PSCs, and we will refer to it simply as CDK8/19i (**Figures 1A, 1B, and S1A**; also, for the structure and characterization of this inhibitor, and comparison with other inhibitors used in this study, see: **Table S1**). In addition to the improvements in *Nanog*-GFP profile and colony morphology mentioned above, the effect of CDK8/19i on mouse PSCs resembled 2i in three additional ways: (i) it occurred in serum- and serum-free based media (**Figures 1A and S1A**); (ii) it was reversible upon withdrawal of CDK8/19i with a kinetics similar to that of 2i-removal (**Figure S1C**); and, (iii) upon removal of LIF or inhibition of LIF signaling with a JAK inhibitor, the presence of CDK8/19i delayed the down-regulation of *Nanog*-GFP expression (**Figures S1D and S1E**). Taken together, we conclude that inhibition of Mediator kinase CDK8/19 shifts mouse PSC morphology and *Nanog* expression towards their characteristic status in the naïve state.

Genetic validation was next explored. Depletion via shRNA-knockdown of CDK8, CDK19, but most successfully, their regulatory partner cyclin C (CCNC; which is essential for full kinase activity), led to up-regulation of *Nanog* expression and naïve-like colony morphology (**Figures 1G, S1F and S1G**). In a second genetic approach, CDK8 and CDK19 double-knockout (dKO) mouse PSCs were generated (**Figure S1H to S1K**). CDK8/19-dKO PSCs could self-renew

151 indefinitely, but they did not acquire naïve morphological features or *Nanog*
152 upregulation, and no longer responded to CDK8/19-inhibitors (**Figures S1L and**
153 **S1M**). This suggested that the beneficial effects observed above may require the
154 physical presence of the inactive-kinase. Consistent with this hypothesis, we
155 found that CDK8/19-dKO PSCs reconstituted with a CDK8-kinase dead mutant
156 (CDK8-KD; D173A) displayed homogenous naïve colony morphology, promoted
157 high expression of naïve-state markers (**Figures 1H-M**), and down-regulated *Fgf5*,
158 a key marker of the primed state (**Figure 1J**). It is important to emphasize that
159 reconstituted-CDK8-KD PSCs acquire these naïve features without the need of
160 any chemical inhibitor and despite maintaining active MEK-ERK signaling (**Figure**
161 **1K**). Thus, reconstituted-CDK8-KD PSCs closely recapitulate the effects observed
162 above by small molecule inhibition of CDK8/19.
163
164 Post-implantation epiblast stem cells (EpiSC; cultured with FGF2, Activin A
165 and fibronectin) are considered to exist in a deeper, or developmentally more
166 advanced, primed state than mouse PSCs in serum/LIF, and they are marked by
167 high *Fgf5* expression and low expression of naïve markers [36]. Conversion of
168 mouse EpiSC into naïve PSCs is highly inefficient [50–52]. To address whether
169 CDK8/19 inhibition can confer the naïve state in EpiSCs, we derived EpiSCs from
170 ES cells (see **Methods**) and exogenously over-expressed CDK8-KD.
171 Interestingly, EpiSC/CDK8-KD cells lose *Fgf5*, express *Nanog, Rex1* and *Klf4*, and
172 form dome-shaped colonies with high alkaline phosphatase staining, characteristic
173 of the naïve state (**Figures 1N, 1O, and S1N**). Together, these data indicate that
174 the presence of a kinase-inactive CDK8 is sufficient to down-regulate primed
175 features and promote several key characteristics of naïve pluripotency, despite the
176 continued presence of MEK-ERK signaling.
177
178 **CDK8/19i-adapted mouse PSCs maintain developmental potential**
179 To extend and explore the effect of CDK8/19 inhibition, we cultured mouse PSCs
180 for over 10 passages in CDK8/19i. This maintained ICAM1 cell surface expression
181 and enhanced naïve features, including colony morphology, high alkaline
182 phosphatase (AP), *Nanog*-GFP^{high}, and high endogenous *Nanog* (**Figures 2A to**
183 **2C, and S2A**). Another characteristic of the naïve state is TFE3 nuclear

localization [53] and this was also observed in long-term passaged CDK8/19i PSCs (**Figure S2B**). Long-term CDK8/19i-adapted PSCs displayed typical developmental capacity following inhibitor withdrawal, specifically, retinoic-acid-induced differentiation, embryoid body cardiac centre formation, spheroid polarization and lumenogenesis, generation of teratomas containing three germ layers, and robust chimera contribution after morula aggregation and blastocyst micro-injection assays (traced by constitutive GFP or RFP) evaluated at E4.5, E7.5, E14.5 and fully-developed adults which subsequently completed germline transmission (**Figures 2C to 2H, S2C to S2E**). Of note, maintaining the continued presence of CDK8/19i impaired the early developmental events of polarization and lumenogenesis *in vitro* (**Figure 2D**), an observation discussed further below. Thus, pluripotent stem cells long-term adapted to CDK8/19i maintain both self-renewal and developmental capacity.

CDK8/19i induces and stabilizes the naïve state in human PSCs

We next tested the effect of CDK8/19i on human stem cell identity. STAT3 overexpression in combination with 2i/LIF induces the human naïve state [54], and we observed that CDK8/19i can replace 2i in this system (**Figure 2I**). Interestingly, STAT3 overexpression was dispensable, and 14 days of CDK8/19i treatment, in the absence of other chemicals or transgenes, progressively converted human iPS colonies from flat and primed-like to a dome-shaped bi-refringent morphology. This was observed for a total of 7 human PSC lines treated with 0.4 μM or 1.1 μM CDK8/19i/LIF (**Figures 2J and S2F**). This included human iPS cells carrying a specific *HERVH*-GFP reporter insertion that marks human naïve cell identity[55] (**Figure S2G**). A 2i-based chemical cocktail combined with selection by cell-sorting (abbreviated as 2i p38iJNKi) induced naïve colony morphology, as expected [55–57], with homogeneous high *HERVH*-GFP (**Figure S2G**). Interestingly, treatment with CDK8/19 inhibitors (CNIO-47799 or SnxA) also produced morphology conversion and increased GFP, similar to 2i p38iJNKi (**Figures 2K and S2G**). The changes induced by CDK8/19 inhibition were gradual, required no selection upon passage (sorting or manual picking), required no additional supplements except rhLIF, and were stable in the continuous presence of the inhibitor. In contrast, inhibition of CDK7 failed to change colony morphology or

GFP fluorescence, and was instead associated with cell death (**Figure S2G**). Culture of human PSCs in CDK8/19i increased clonogenicity and alkaline phosphatase intensity, as well as, endogenous pluripotency markers NANOG, OCT4, SSEA4, TRA1-81, TFCP2L1, and KLF17 (**Figures 2L, 2M, and S2H to S2K**). In contrast, MYC, known to be reduced in naïve cells [34,58], was also reduced in cells maintained in CDK8/19i (**Figure 2M**). Therefore, similar to observations in mouse PSCs above, CDK8/19i-adaption of human PSCs can reset several characteristics indicative of naïve state stabilization.

Developmental potential of CDK8/19i-adapted human PSCs

Recent reports suggest that chemical induction of the human naïve state can trigger genomic instability, severely impairing subsequent developmental potential [33,59]. We found that naïve CDK8/19i-adapted human PSCs had normal karyotype after 8-10 passages (>40 days), suggesting genomic stability comparable to control primed cells (**Figure S2L**). Next, the developmental potential of CDK8/19i-adapted human PSCs was examined. We observed that upon inhibitor withdrawal, CDK8/19i-adapted human PSCs maintained the capacity to contribute towards all three embryonic germ layers by embryoid body differentiation *in vitro* (**Figure 2N and Table S1**), and by teratoma assay *in vivo* (**Figures 2O, 2P, and Table S1**). Thus, CDK8/19i-adapted naïve human PSCs matched the capacity of the parental primed PSCs.

A naïve-specific developmental assay was next pursued. Preimplantation interspecies chimerism is emerging as a test for naïve-specific characteristics, namely, capacity for clonal survival in a host embryo [60,61]. We tested CDK8/19i-adapted human iPS carrying a constitutive Tomato-red marker for human-rabbit interspecies chimerism. Specifically, 5 or 10 CDK8/19i-adapted human PSCs were micro-injected into pre-implantation E2.5 rabbit morulae. Interestingly, the presence of human cells (Tomato-positive) was detected 72 h later in up to 50% of rabbit blastocysts (**Figure 2Q**). In contrast, human PSC in the primed state were unable to integrate or survive in rabbit embryos (0/24 rabbit embryos), similar to several previous reports for primed state human PSCs within the embryos of mice, pigs, and cattle [62–68].

Overall, we conclude that long-term adaptation of mouse and human PSCs to CDK8/19i stabilizes naïve pluripotency while preserving their developmental potential. This suggests that the role of CDK8/19 in pluripotency may be conserved across mammals.

CDK8/19i resets the transcriptome similar to 2i

We next compared by RNA sequencing global gene expression in mouse ES cells long-term adapted to CDK8/19i *versus* 2i. Overall, in mouse ES cells, CDK8/19i dramatically altered gene expression with a magnitude similar to 2i conditions, and with a highly significant overlap in the identity and biological functions of genes up- or down-regulated (FDR<0.05) (**Figure 3A, S3A, and Table S2**). These overlaps were also observed in serum-free media conditions (**Figure S3B; Table S2**). Of note, naïve and core pluripotency markers were maintained or enhanced in CDK8/19i or 2i, compared to control serum/LIF conditions (**Figure 3B**), with changes validated by qRT-PCRs (**Figure S3C**). Moreover, differentiation markers were globally down-regulated in CDK8/19i and 2i states (**Table S2**).

Endogenous retrovirus (ERV) expression is highly stage-specific during mouse pre-implantation, and ERV-mediated transcriptional control is integral to ES cell identity [66,69–74]. The transcriptomic overlap between CDK8/19i and 2i states extended to ERVs, with similar viral families significantly up- or down-regulated in both CDK8/19i and 2i (**Figure 3C; Table S2**). In particular, LINE L1 families, each with thousands of copies across the genome, were regulated in close parallel, displaying a highly similar level of expression in CDK8/19i and 2i-naïve states (**Figures S3C and S3D**). Another aspect of the plasticity of mouse PSCs is their ability to transition to a 2-cell-like (2C) state, specifically marked by hyper-activation of the MERVL family of ERVs and by *Zscan4c* expression [74–76]. Stabilization of the naïve state with 2i limits the 2C-like fluctuation [74,75]. This was also the case in CDK8/19i-treated PSCs, as confirmed by multiple 2C-markers, including *MERVL* and *Zscan4c*, and by both *MERVL*-Tomato and *Zscan4c*-eGFP 2C-reporter models (**Figures 3C, and S3E to S3L**). Finally, we observed significant correlation between our CDK8/19i and 2i transcriptomic data and

published transcriptomes from several independent studies of 2i-naïve mouse cells (**Figures 3E and S3M**), and also with the transcriptome of E4.5 epiblast single-cells (**Figures 3E and S3N**).

RNA-seq analyses of human PSCs adapted to CDK8/19i or a 2i-based naïve cocktail (2i p38iJNKi) overlapped significantly (**Figure 3F; Table S3**). In addition, recently identified markers of human and primate pre-implantation epiblasts and *in vitro* naïve human PSCs [56,59,66,77–84] were strongly up-regulated by CDK8/19i, including *NANOG*, *TFCP2L1*, *KLF5, KLF17*, *CDH1*, *NODAL*, *TDGF1*, *FGF4*, *GDF3*, and *SOX15*, while differentiation markers were repressed (**Figures 3G, S4A to S4D; Table S3**). Moreover, the global human ERV transcriptomes of CDK8/19i-adapted and 2i-adapted cells overlapped heavily, including up-regulation of the SVA-family, LTR7-family, and HERV-family viral elements (**Figures 3H to 3J; Table S3**). These changes are highly consistent with previous characterizations of ERV expression for human and primate naïve PSCs and pre-implantation epiblast [66,72,73]. Lastly, we observed a high correlation between our transcriptome data in human PSCs in CDK8/19i and RNA expression datasets from seven independent studies in human and primate PSCs, in the *in vitro* naïve state, and in embryo naïve epiblast single-cell analyses (**Figure 3K**).

In summary, in both mouse and human PSCs, CDK8/19i up-regulates naïve and core pluripotency markers, re-shapes the endogenous retroviral transcriptome, and represses differentiation markers, in a manner remarkably similar to the transcriptomic resetting observed in multiple previous studies of naïve pluripotency, *in vitro* and *in vivo*, in mouse and in human.

CDK8/19i resets the proteome similar to 2i

While PSC plasticity has been heavily explored in terms of RNA expression, its proteome remains relatively ill-defined. Hence, we next analyzed the proteome of mouse ES cells in standard serum/LIF *versus* 2i-naïve or CDK8/19i-adapted conditions. Across five mouse PSC lines, CDK8/19 inhibition altered the expression levels of 465 proteins, 159 (35%) of which changed in the same direction in 2i conditions (**Figures 3L, S4E, S4F, and Table S4**). Importantly,

among the overlapping proteins in both 2i-naïve and CDK8/19i conditions, key pluripotency regulators such as KLF4, and metabolic pathways such as oxidative phosphorylation, featured amongst the most upregulated, while LIN28A, MYC-target genes, and differentiation markers were down-regulated (**Figures 3M, S4G and Table S4**). In addition, proteomic changes in 2i and CDK8/19i significantly correlated with the transcriptomic changes observed above (**Figures S4H and S4I**).

CDK8/19i does not reset global DNA methylation levels

Many 2i-based chemical cocktails induce global DNA hypomethylation, both in human and mouse PSCs, similar to the pre-implantation naïve epiblast [44,58,66,85–87]. More recently, 2i-variant cocktails have been reported to induce naïve features while largely preserving global DNA methylation, both in mouse and human PSCs [56,88,89]. Importantly, neither mouse nor human CDK8/19i-adapted PSCs showed evidence of global DNA hypomethylation (**Figures 3N and 3O**). In addition to global DNA hypomethylation, and as previously reported [86], 2i or MEK-inhibition alone, induced demethylation of LINE L1 repeat regions (**Figure 3P**), and major satellite regions (**Figure S4J**), but had no effect on the methylation of IAP repeats (**Figure S4K**). In contrast to this, CDK8/19i did not reduce methylation at any of these mouse repeat elements (**Figures 3P, S4J and S4K**). Therefore, CDK8/19i, like some 2i-variant cocktails, is able to induce naïve features in the absence of global DNA hypomethylation or hypomethylation of repeated elements.

X-chromosome reactivation status is another molecular signature reported in human naïve pluripotency during MEK-inhibition [90–93], which may be inferred by assessing XIST RNA expression in female cells. However, analysis by qPCR revealed very low XIST expression in primed human PSCs (**Figure S4L**), suggesting that erosion of X-silencing may have already occurred in the parental cells under primed conditions, as commonly observed previously [90–94]. Notably, CDK8/19i treatment did not reactivate XIST expression (**Figure S4L**), a phenomenon which was recently reported using specific media cocktails that also induce several aspects of the naïve human pluripotent state [90–92]. In summary therefore, CDK8/19i treatment does not recapitulate this reported reactivation of

XIST RNA expression after X-silencing erosion, indicating another distinction with media cocktails based on MEK-inhibition [90–92].

CDK8/19i induces phospho-changes similar to 2i

To further explore the similarity between CDK8/19i and 2i, we next assessed the phospho-proteome of mouse PSCs just 15 min after exposure to CDK8/19i or 2i. Intriguingly, out of the 622 phospho-sites altered, 495 (79.6%) were similarly regulated by CDK8/19i and 2i (**Figures 4A and 4B**). The phospho-sites co-regulated (both up and down) by CDK8/19i and 2i occurred on proteins heavily enriched for functions in transcriptional regulation and key stem cell signaling pathways (**Figure 4B, S5A; Table S5**). As shown before, CDK8/19i does not inhibit the kinase activity of purified recombinant GSK3 or MEK (**Table S1**). To further reinforce this, we confirmed in cellular assays that CDK8/19i does not reduce the relative levels of phospho-ERK (**Figures 4C, 4D, and S5B**). Instead, 2i treatment reduced CDK8/19 kinase activity on its target STAT1 (**Figures 4C, and S5C**) with a moderate downregulation of CDK8 protein levels (**Figure S5D**). Taken together, these data suggest that CDK8/19 inhibition occurs downstream of 2i, such that both treatments result in highly overlapping phospho-site changes.

CDK8/19i resets global RNA Pol II loading similar to 2i

To understand how CDK8/19-inhibition phenocopies the transcriptome of 2i-induced naïve pluripotency, we next investigated global regulation of RNA Pol II abundance on chromatin by ChIP-seq in mouse ES cells cultured in three conditions: serum/LIF, 2i, or CDK8/19i. Similar to published resources of Pol II ChIP-seq in mouse ES cells [38,95,96] (ENCODE: https://www.encodeproject.org/), we found RNA Pol II, both total and Ser5-phosphorylated, at transcriptional start sites (TSS) in all three conditions (**Figures 4E, 4F, and S5E, S5F**). In close agreement with data extracted from a previous report [38], we observed that 2i increases RNA Pol II binding to promoters (**Figures 4E, 4F, and S5E to S5G**). Remarkably, this global effect of 2i was phenocopied by CDK8/19i, regarding both total and Ser5-phosphorylated RNA Pol II (**Figures 4E, 4F, and S5E to S5G**). Changes in RNA Pol II abundance at promoters and in other DNA elements may hold mechanistic insights. Following similar previous analyses [95,97], we measured

RNA Pol II abundance in the promoter, gene body and transcription termination site (TTS) for each gene (**Figures 4G and S5H; Table S6**). 90% of genes possessed a promoter to gene body loading ratio > 2.0 (**Figure 4G; Table S6**), consistent with previous reports in mouse ES cells [38,95]. Comparison of the ratios of RNA Pol II between the promoter, gene body, or termination sub-regions of each gene indicated that 2i induces a selective increase in RNA Pol II binding to the promoter region of genes (**Figures 4G and S5H**), as reported [38]. Importantly, this was recapitulated by CDK8/19i, increasing RNA Pol II binding to promoters at a similar magnitude to that observed in 2i-induced naïve pluripotency, including at the level of individual genes (**Figures 4G to 4M; Table S6**). Therefore, 2i- and CDK8/19i-induced naïve pluripotency are accompanied by widespread accumulation of RNA Pol II abundance at promoters.

We next investigated how the resetting of RNA Pol II abundance reflected on the behavior of individual genes. For this, we ranked genes by their differential RNA Pol II loading at the promoter in 2i or CDK8/19i conditions, and compared this to their relative mRNA or protein levels. We observed a gene-specific relationship between RNA Pol II promoter abundance in 2i or in CDK8/19i conditions, and relative mRNA changes (**Figures S5I and S5J, S6A to S6C**). In summary, gene-specific changes in RNA Pol II promoter loading may explain a significant proportion of the gene expression profile characteristic of 2i- or CDK8/19i-induced naïve pluripotency.

CDK8/19i and 2i trigger activation of super-enhancers

The primary role of Mediator is at enhancers, regulating RNA Pol II recruitment to promoters [4,5,7]. Therefore, we hypothesized that CDK8/19 inhibition may trigger changes in enhancer activity, which could explain the observed increase of RNA Pol II loading at promoters. To this end, we first identified the genomic localization of CDK8/19 in mouse PSCs using published ChIP-seq datasets [3] (**Table S7**). CDK8/19 was particularly enriched at PSC promoters and enhancers, including previously defined[2] super-enhancer (SE) and typical enhancer (TE) regions (**Figures 5A, S6D to S6F)**. Specifically, the majority of super-enhancers in mouse PSCs contained CDK8/19 (**Figure 5A, S6F**), and there was a strong quantitative

correlation between the abundance of CDK8/19, Mediator subunits, and other factors critical for enhancer activity (such as, p300, CBP, Pol II, or BRD4) at enhancers (**Figure S6G**). Moreover, we noted that the highest levels of CDK8/19 occurred within PSC super-enhancers (**Figure 5A and Table S7**). In addition, the putative target genes proximal to genomic CDK8/19 binding loci were highly enriched in preimplantation functions (**Figure S7A to S7C**).

Since CDK8/19 protein is particularly enriched at super-enhancers (SE), we next examined the impact of CDK8/19i on SE function. Enhancers contain RNA Pol II which transcribes enhancer-RNAs (eRNAs), a process that faithfully reflects enhancer activity [4,98,99]. Accordingly, we measured the effect of CDK8/19i or 2i on the levels of RNA Pol II and eRNAs at SEs in mouse PSCs, to assess their activity. 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 to typical enhancers (**Figures 5B, 5C, and S7D**). Consistent with this, mouse PSCs treated with 2i or CDK8/19i displayed elevated enhancer-derived eRNA levels within enhancers previously described as specifically activated in the naïve-state [100] (**Figures 5D, and S7E, S7F**). The induction of naïve-specific eRNAs and naïve marker genes was an early event, occurring within 48 hours of adding either 2i or CDK8/19i, and moreover, it was also rapidly reversible (**Figures 5D and S7G**). Lastly, consistent with naïve-specific enhancer activation, the expression levels of SE target-genes were selectively up-regulated in both 2i and CDK8/19i (**Figures 5E, 5F and S7H**). We conclude that in PSCs, CDK8/19i and 2i hyper-activate existing SE, and up-regulate SE target-genes, in a manner which reinforces naïve pluripotency.

CDK8/19 inhibition or 2i compensates decreased Mediator function

Across multiple cell types, loss of Mediator function triggers a selective decrease in expression of enhancer target genes [4,5,8]. For example, BRD4-inhibition in primed state PSCs decreases the ability of Mediator to recruit RNA Pol II, and this rapidly results in loss of Mediator-driven transcription, collapse of pluripotency gene expression, and differentiation (see **Figure 1A**) [39,40,42]. Compared to primed PSCs, naïve PSCs were recently found to be highly resistant to decreased

Mediator activity and enhancer destabilization induced by BRD4-inhibition [43]. Interestingly, mouse PSCs reconstituted with kinase-dead CDK8 were resistant to enhancer destabilization by BRD4-inhibition for 10 passages (>3 weeks), maintaining naïve morphology, high expression of alkaline phosphatase, and high expression of naïve-specific pluripotency markers genes and eRNAs, similar to 2i-naïve wild-type cells (**Figures 5G, 5H, and S7I, S7J**). Thus, PSCs expressing kinase-dead CDK8 phenocopy the robust resistance to enhancer destabilization which is characteristic of 2i-naïve PSCs.

Roles of CDK8/19 during early embryonic development

Given the above observations that CDK8/19 inhibition stabilizes naïve pluripotency, we lastly investigated CDK8/19 function during early embryonic development. We focused on CDK8, because it is highly expressed compared to CDK19, both in mouse and human PSCs (**Figure S1K and S8A**) and, for this, we used a CDK8-specific antibody (see **Figure S1J**). We detected CDK8 protein at the fertilized oocyte and morula states (**Figure S8B**). To test the role of CDK8/19 during these stages, we exposed fertilized oocytes to CDK8/19i and followed their development *in vitro*. We observed that inhibition of CDK8/19 severely impaired the progression of oocytes to the 2-cell stage (**Figure 6A**). In agreement with this, it was reported that CDK8-knockout is embryonic lethal before the mouse 4-8 cell stage [101].

We next investigated the role of CDK8 post-morula. CDK8 mRNA expression declined until blastocyst stage, both in mouse and human pre-implantation embryos, according to several published datasets [78,102,103] (**Figures 6B, S8C and S8D**). We observed that CDK8 protein expression per cell was homogenous in the inner cell mass (ICM) at E3.5 (**Figures 6C and 6D**). Interestingly, at E4.5, when the ICM segregates into the naïve epiblast (EPI) and the primitive endoderm (PE), CDK8 protein levels diverged, with lower levels in EPI compared to PE (**Figures 6C, 6D and S8E**). This pattern was transient, and it became reversed in post-implantation EPI at E5.5 (**Figures 6C, 6D and S8E**). To further document that CDK8 levels are reduced in the naïve epiblast, embryos were cultured from E3.5 to E4.5 in the presence of MEK inhibitor (MEKi). It is well

established that MEK inhibition blocks PE formation in E4.5 embryos, permitting only the emergence of the naïve epiblast [37,45,104,105]. As expected, the presence of MEKi prevented PE formation and CDK8 expression was homogenously reduced in the epiblast (**Figures 6E and 6F**). We also assessed the CDK8 binding partner and essential activating subunit cyclin C. From E4.5 to E5.5, cyclin C altered its nuclear-cytoplasmic ratio, specifically, E4.5 epiblast contained significantly less nuclear cyclin C than E5.5 *in vivo* (**Figures S9A and S9B**), while a similar pattern was also observed comparing 2i-naïve versus primed state PSCs *in vitro* (**Figure S9C**). In summary therefore, the emergence of naïve pluripotency during embryo development, at E4.5, coincides with decreased CDK8 expression and decreased availability of its essential subunit cyclin C. Notably, this parallels the effect of MEK-inhibition on CDK8 expression and stabilization of naïve epiblast identity in ES cells *in vitro* (see above, **Figures 4 and S5**).

We next asked whether inhibition of CDK8/19 affects the emergence of naïve pluripotency. Similar to MEKi, exposure of E3.0-3.5 embryos to CDK8/19i did not interfere with epiblast development (**Figures 6G and 6H**) and allowed the derivation of ES cell lines. However, in contrast to MEKi, CDK8/19i permitted PE formation (**Figure 6G and 6H**). This is consistent with the observation above that MEK activity is unaffected by CDK8/19i (**Figures 4C and 4D**), and it suggests that the critical roles of MEK for PE segregation are independent of CDK8/19.

Having established that CDK8/19 activity is dispensable for the emergence of the naïve epiblast, similar to MEK, we examined its requirement in subsequent developmental transitions. Particularly, considering the elevation of CDK8 levels observed during the pre-implantation to post-implantation epiblast developmental transition (**Figures 6B to 6D**). As a readout of epiblast progression, we focused on lumen formation within the post-implantation epiblast, which marks the initiation of morphogenesis downstream of naïve pluripotency exit [106,107]. We found that CDK8/19i treatment during E4.5 to E5.5 impaired post-implantation epiblast lumenogenesis of embryos (**Figure 6I;** see also Figure 2D, above, for spheroids). This indicates a requirement for CDK8/19 activity to support epiblast development,

512 from the naïve pre-implantation to primed post-implantation embryonic stage,
513 consistent with the observed elevation in CDK8 expression at this time.
514
515 Altogether these data suggest that CDK8/19 function in early embryonic
516 development mirrors its expression pattern (**Figure 6J**), and may be summarized
517 in three periods: (i) CDK8/19 is required during 1C to morula development, where
518 its expression is high. (ii) During morula to blastocyst pre-implantation
519 development, CDK8 and cyclin C expression declines. This coincides with the
520 emergence of the E4.5 pre-implantation naïve epiblast and, accordingly, CDK8/19
521 inhibition does not interfere with naïve epiblast specification. Notably, in contrast
522 to MEK inhibition, CDK8/19 inhibition does not affect the EPI/PE lineage
523 segregation. (iii) During the subsequent developmental transition of pre-
524 implantation naïve epiblast to the post-implantation primed state, CDK8 expression
525 becomes increased and CDK8/19 activity is required for the morphogenic events
526 during this transition.
527
528 We conclude that CDK8 inhibition coincides with the emergence of naïve
529 pluripotent epiblast identity *in vivo*, a feature which can be exploited to stabilize
530 naïve PSC culture by CDK8/19i *in vitro*.
531
532 **DISCUSSION**
533 Here, we uncover a role for the Mediator-kinases CDK8/19 in defining the
534 equilibrium between naïve and primed pluripotent states. Moreover, the findings
535 provide a chemical method to transition from naïve to primed identity both in mouse
536 and human pluripotent cells. Molecular analyses reveal how the RNA Pol II
537 transcriptional machinery is reorganized to coordinate this cell identity conversion
538 (**see Extended Discussion in Supplementary Material**). Collectively, our data
539 point towards the following model: 2i and CDK8/19i rapidly induce a highly
540 overlapping set of phospho-changes focused on the transcriptional machinery,
541 triggering enhancer hyperactivation, global increase in RNA Pol II recruitment to
542 promoters, and resetting of gene expression. This also includes the upregulation
543 of enhancer-derived RNAs (eRNAs), and the resetting of endogenous retroviral
544 and repeat element expression. Thus, the ability of 2i and CDK8/19i to induce

naïve features appears to originate from their common effect on Mediator and RNA Pol II transcriptional activity. This model is consistent 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 [98,100,108]. Thus, the transcriptional landscape of naïve pluripotency can be stabilized by Mediator stimulation, and this can be achieved by chemical inhibition of CDK8/19, a process which appears to mimic CDK8 downregulation during pre-implantation development *in vivo*.

Further studies are required to reveal if and how MEK-ERK signaling may regulate CDK8/19 activity in PSCs (see **Extended Discussion**). However, based on the current evidence, we suggest this model to explain how CDK8/19-inhibition can recapitulate many molecular events typically observed during the induction of the naïve state, for example, downstream of 2i-treatment of PSCs in vitro, or in vivo. An important difference between 2i-naïve cells and CDK8/19i-naïve cells is the fact that CDK8/19i-naïve cells do not present global DNA hypomethylation. This is relevant because prolonged DNA hypomethylation associated with MEK-inhibition is known to have detrimental side effects, specifically chromosomal instability and imprinting erasure, which are appear particularly acute in human PSCs [33,59,88,89]. Indeed, stabilization of the naïve state in human PSCs remains to be optimized [32,33]. In this regard, CDK8/19i-naïve human cells retain normal karyotype after prolonged culture. Therefore, chemical inhibition of CDK8/19 offers a new approach that may help to solve remaining challenges in human naïve PSC culture.

Taken together, we report that the transcriptional landscape of naïve pluripotency can be stabilized by Mediator stimulation, and this can be achieved by chemical inhibition of CDK8/19. The extent to which CDK8/19i mimics 2i suggests a central role of Mediator during the induction of naïve pluripotency, and it provides a mechanism by which naïve pluripotency may arise *in vivo*. Lastly, chemical inhibition of CDK8/19 may help to stabilize other intrinsically unstable cell states.

EXTENDED DISCUSSION

Identifying a conserved role for Mediator/CDK8/19 in naïve pluripotency

To explore the role of the transcriptional machinery in the maintenance of the naïve state, we began by focusing on the transcriptional CDKs (CDK7, CDK8/19, and CDK9). We observed that inhibition of CDK7 or CDK9 was deleterious to mouse PSCs, consistent with their general roles in RNA Pol II transcription. In contrast, selectively targeting the activity of the Mediator kinases CDK8/19 had a positive effect in inducing and stabilizing naïve identity. To confirm these positive effects, we utilized a number of structurally-unrelated and commercially available small molecules, in addition to developing and validating a novel potent inhibitor of CDK8/19. We recapitulated the small molecule approach using three genetic models: (i) shRNA knockdown of cyclin C, the rate-limiting partner of CDK8/19, resulted in upregulation of naïve features; (ii) CDK8/19 double-knockout PSCs no longer responded to CDK8/19 small molecule inhibitors, indicating the specificity of the inhibitors, and suggesting that CDK8/19 may promote naïve pluripotency by a kinase-independent process; (iii) CDK8/19 double-knockout PSCs reconstituted with a CDK8-Kinase Dead protein recaptured the upregulation of naïve features, confirming that CDK8 promotes naïve identity by a critical, kinase-independent, process. Importantly, kinase-independent functions for CDK8 have been reported, for example, the CDK8/19 sub-module can act as a negative regulator of core-Mediator, by steric hindrance against recruitment of RNA Pol II [5,15,21].

An intriguing observation was that CDK8/19-dKO PSCs required re-constitution with a kinase-dead CDK8 protein in order to recapitulate the effects of the CDK8/19-small molecule inhibitor. This requirement for the physical presence of CDK8 could be explained in two ways. In a simple model, CDK8 may play a kinase-independent structural role together with CCNC and the other subunits of the kinase Mediator module, and indeed, this effect has been widely reported [21,109–113]. In an alternative possibility, which we cannot exclude, the absence of both CDK8 and CDK19 may allow the atypical incorporation of other CDKs which are now able to associate with the unpartnered CCNC, for example as shown for CDK2

610 [114], re-constituting a functional Mediator complex that is no longer responsive to
611 CDK8/19 chemical inhibition.
612
613 Importantly, we also find that CDK8/19i induces naïve features in human
614 PSCs. This suggests that the role of CDK8/19 in pluripotency is conserved across
615 mammalian species. Stabilization of the naïve state in human PSCs remains to
616 be optimized [32,33]. In this regard, chemical inhibition of CDK8/19 offers the
617 advantage of a new approach that may help to solve remaining challenges in
618 human naïve PSC culture.
619
620
621 **CDK8/19-inhibition phenocopies molecular features of naïve pluripotency**
622 Based on the cellular behaviour described above, we assessed molecular events
623 upon induction of naïve pluripotency in PSCs, and then compared them to the
624 effect of CDK8/19-inhibition.
625
626 While the 2i kinase-inhibitor cocktail promotes the transition of PSC identity
627 from primed to a stable naïve state, little is known regarding the early phospho-
628 proteomic changes which mediate the conversion mechanism. We observed that
629 just 15 minutes after exposure of PSCs to 2i or to CDK8/19i there is a highly
630 overlapping change in the phospho-proteome, largely focused on components of
631 the transcriptional machinery. Of note, we also observe that 2i down-regulates
632 CDK8/19 activity. This suggests that the primary effects of MEK and GSK3
633 inhibition are rapidly exerted on the transcriptional apparatus, and that CDK8/19
634 may lie downstream of these pathways. Nevertheless, our data do not exclude the
635 possibility that 2i and CDK8/19i may also exert non-transcriptional and/or
636 secondary effects on the transcriptome. The data implies a simple model whereby
637 MEK signaling may regulate CDK8 activity and Mediator function downstream, yet
638 many questions remain regarding the connecting steps. Our observations provide
639 a conceptual framework for further mechanistic dissection of how MEK signaling
640 may regulate Mediator-driven gene transcription.
641

642 In line with the similarity in the early phospho-proteome, we also found
643 highly significant overlap in several molecular profiles of CDK8/19i-adapted and 2i-
644 adapted PSCs. These analyses include enhancer activity (measured by RNA Pol
645 II abundance and enhancer RNA transcription), recruitment of RNA Pol II to
646 promoters, transcriptome, and proteome. Of note, the parallels between 2i and
647 CDK8/19i treatments extend to their effect on the pattern of expression of eRNAs
648 and repetitive elements such as endogenous retroviruses, where a role for
649 Mediator had not been previously demonstrated to our knowledge. In addition, 2i
650 and CDK8/19i display a similar gene-specific correlation between the magnitude
651 of change in RNA Pol II loading at promoters and the ultimate changes in mRNA
652 and protein expression levels. All together, these data point towards the following
653 model: 2i and CDK8/19i rapidly induce a highly overlapping set of phospho-
654 changes focused on the transcriptional machinery, triggering enhancer
655 hyperactivation and highly similar resetting of global RNA Pol II loading and gene
656 expression. Thus, the ability of 2i and CDK8/19i to induce similar naïve features
657 originates in their common pattern of RNA Pol II transcriptional control. This model
658 is consistent with the concept that transitions in cell identity are driven by early
659 reconfiguration of the active enhancer network, which resets the transcriptional
660 machinery to the new program [98,100,108].
661
662 **Global enhancer hyperactivation underlies naïve pluripotency**
663 The Mediator complex is considered an integrative hub of upstream signals and
664 plays a central role in cell identity [1,4–8,115]. By analyzing previous ChIP-seq
665 datasets [2,3], we found that in PSCs, CDK8/19 co-localizes with the Mediator
666 complex essentially at all active enhancers and promoters, and is particularly
667 enriched at super-enhancers. As mentioned before, the induction of naïve identity
668 using 2i or CDK8/19i treatments can stimulate Mediator function, which we detect
669 by a global increase in RNA Pol II recruitment, global hyper-activation of existing
670 PSC enhancer loci, and upregulation of enhancer-driven transcription. We
671 propose that this reinforces the pluripotency network underlying naïve PSC
672 identity. In agreement with a recent report [43], we observe that in 2i, naïve-specific
673 enhancer activity is resistant to enhancer/Mediator destabilization by BRD4-
674 inhibition. Importantly, this property can also be conferred by expression of CDK8-

675 Kinase Dead mutant protein. This suggests a simple mechanism where removal
676 of the inhibitory influence of CDK8/19, hyperactivates Mediator function at
677 enhancers, and that this occurs similarly in 2i or via CDK8/19 inhibition. In support
678 of global activation of super-enhancers in the naïve state, a recent study of
679 chromatin looping has revealed that super-enhancers interact with more target
680 promoters, and engage in more long-range interactions, during naïve pluripotency
681 compared to primed pluripotent cells [116], while furthermore, a state of global
682 hypertranscription has been suggested in PSCs [117,118].
683
684 We note that a similar mechanism of Mediator hyperactivation via CDK8/19
685 inhibition has been reported in cancer cells [17]. However intriguingly, this resulted
686 in cell death in acute myeloid leukemia (AML) cells [17], while we find that a similar
687 approach in PSC reinforces naive cell identity. Cancer cells commonly develop
688 novel oncogenic SEs that can result in addiction to a defined range of enhancer-
689 driven transcription [119,120]. Thus cancer cell oncogenic SEs may be sensitive to
690 perturbation, either when hyperactivated, as in the case of CDK8 inhibition [17], or
691 when inhibited, as in the case of CDK7 inhibition [119,120]. This provides an
692 interesting parallel with MEK inhibition, which is also detrimental to many cancer
693 cells, but is beneficial to the naïve state.
694
695 **Role of Mediator/CDK8/19 during early development**
696 Embryos null for *Cdk8* are lethal before the 4-8-cell stage [101]. Consistent with this,
697 we found a peak of CDK8 expression around the 2-cell stage, and that CDK8/19i
698 treatment of zygotes blocked development before the 4-cell stage. Related to this,
699 PSCs can transiently re-activate part of the transcriptional program of 2-cell (2C)
700 stage embryos, in a fluctuation that remains poorly understood [74–76]. Interestingly,
701 the 2C-like transcriptional-fluctuation was strongly repressed by either 2i- or
702 CDK8/19i-treatment of mouse PSCs. This further reinforces the concept that
703 CDK8/19-inhibition stabilizes cells in an homogenous naïve state, similar to 2i.
704 Taken together, our data suggest a role for CDK8/19 around the 2C stage, both
705 during normal development, and also when acquired in vitro through the intrinsic
706 plasticity of PSCs.
707

Lineage specification towards epiblast (EPI) or primitive endoderm (PE) initiates within the blastocyst inner cell mass (ICM) between E3.5 and E4.5 [37,121,122]. Moreover, EPI/PE lineage divergence is known to be heavily dependent on FGF-MEK-ERK signaling, and lineage segregation is complete by E4.5 [45,104,105]. We detected differential CDK8 expression in the E4.5 blastocyst ICM, between preimplantation EPI ($CDK8^{low}$) and PE ($CDK8^{high}$). In parallel, cyclin C nuclear availability is comparatively low in epiblast at E4.5, also suggesting restricted CDK8 activity. Therefore, CDK8/19i treatment of PSCs in vitro mimics the down-regulation of CDK8 protein levels and activity during the pre-implantation epiblast at E4.5. Although CDK8 protein is upregulated in PE cells, we note that blocking its kinase activity using CDK8/19i does not seem to impair exit from ICM identity and entry towards the formation of PE cells. A simple explanation for this may be that CCNC levels are low at E4.5 in PE, and remain low until after implantation. This suggests that several modes of controlling CDK8 function seem to be operating in early development.

The data also infer a role for CDK8/19 activity later in development. We observed that both CDK8 protein levels and cyclin C nuclear localization increase in the epiblast after implantation. To explore the possible role of CDK8 at this stage, we treated E4.5 blastocysts with CDK8/19i and followed their development in vitro until lumen formation, a process that is characteristic of post-implantation differentiation [106,107,123]. Interestingly, CDK8/19i blocked lumen formation in the epiblast, which is consistent with the idea that low CDK8/19 activity stabilizes the naïve state characteristic of the pre-implantation epiblast. Considering the current data collectively, it appears that CDK8 function follows its expression pattern, with three phases: (i) a peak of CDK8 expression coincides with its requirement at the 2-cell stage; (ii) CDK8 expression reaches a minima in the pre-implantation epiblast, at E4.5, a state that coincides with naïve pluripotency and which we mimic in vitro with CDK8/19i; and (iii) CDK8 re-expression in the post-implantation epiblast, at E5.5, is required for further developmental progression. Consistent with this, an important role was recently suggested for CDK8 during induction of early developmental gene expression [124].

Convergence of inhibitor cocktails in CDK8/19 function

All current PSC media cocktails reported to stabilize the naïve state contain small molecule inhibitors targeting one or more factors in the MEK signalling pathway (FGFRi, RAFi, SRCi, PKCi, p38i, JNKi, MEKi), reviewed in: [32]. It is notable that many components of this pathway have also been shown to regulate CDK8 activity, including KRAS, RAF, SRC, PKC, p38, JNK, MEK, and ERK [125–128]. Thus, CDK8/19-inhibition may be a common feature of naïve-inducing media cocktails. Further studies are required to reveal the mechanism by which MEK-ERK signaling regulates CDK8/19 activity in PSCs. At present, based on the current evidence, we suggest this model to explain how CDK8/19-inhibition can recapitulate molecular events typically observed during the induction of the naïve state, for example, downstream of 2i-treatment of PSCs in vitro, or in vivo.

CDK8/19-independent effects of 2i

As outlined above, CDK8/19i recapitulates a significant proportion of 2i-associated effects on cell identity, however, global DNA hypomethylation was an exception. The ability of 2i to trigger global DNA hypomethylation is thought to be heavily dependent on the inhibition of MEK signaling [32,37,85,86,129]. Transcriptional mechanisms have been proposed to connect MEK-inhibition with DNA demethylation, specifically through *Prdm14* and the *Dnmt3* gene family [44,130,131]. We observed that CDK8/19i recapitulates these same transcriptional changes, however, CDK8/19i did not trigger global DNA hypomethylation. We speculate that MEK inhibition may also contribute to DNA demethylation through non-transcriptional mechanisms, such as direct phosphorylation of DNMTs [132]. The ability of CDK8/19i to implement the transcriptional landscape of naïve pluripotency, without global DNA hypomethylation, may avoid the detrimental side effects of imprint erasure recently reported during conditions of MEK-inhibition [88,89].

Finally, inhibition of CDK8/19 during blastocyst formation did not interfere with the specification of the primitive endoderm (PE), a process highly sensitive to MEK inhibition [37,45,104,105]. The phosphorylation of the transcription factor GATA6 by MEK has been recently shown to be a key event in the determination of the PE [133]. Since CDK8/19i does not affect the kinase activity of MEK, it is possible that

774 the presence of an active MEK/GATA6 circuit is sufficient to determine PE
775 formation in the face of CDK8/19 inhibition. Lastly, differential regulation of TGFβ
776 signaling was suggested at the proteomic level, and this may contribute towards
777 differences we observe between 2i and CDK8/19i treatments.
778
779 **Placing CDK8/19i-treated PSCs along the pluripotency spectrum**
780 Our current understanding of stem cell identity indicates a continuum of molecular
781 changes along a spectrum from naïve to primed states, which also reflects the
782 developmental path in early embryos[13,32,36,121]. Here, we find that CDK8/19
783 inhibition is sufficient to recapitulate the majority of molecular characteristics
784 associated with a transition from the primed to the naïve state, in particular, cellular
785 morphology, the global changes in phospho-signaling, gene expression, and RNA
786 Pol II regulation. In contrast, CDK8/19i does not recapitulate some other reported
787 features of the naïve state [32,134], specifically, global DNA hypomethylation, X-
788 chromosome reactivation[91,135], and, in the particular case of human PSCs, SSEA4
789 down-regulation [33,59,136]. The stabilization of the human naïve pluripotent state has
790 been proven to be challenging and it has become clear that alternate methods
791 affect specific aspects differently. It is notable that many 2i-based media cocktails
792 used for human PSCs result in SSEA4 down-regulation, global DNA
793 hypomethylation, genomic instability and impaired developmental potency [33,59,135–
794 138]. However, we note that the 2i-based cocktail reported by Hanna (2ip38iJNKi)
795 [56], does not downregulate SSEA4, does not produce large DNA demethylation,
796 and is not associated with genomic instability [33,56]. Therefore, similar to the Hanna
797 cocktail [56], CDK8/19i installs many naïve features in human cells while maintaining
798 SSEA4, DNA global methylation and genomic stability, but CDK8/19i does not
799 affect MEK-ERK signaling.
800

ACKNOWLEDGMENTS

We are grateful to Austin Smith, Todd MacFarlan, Zsuzsanna Izsvák, Minoru Ko, and Jacob Hanna for generous gifts of reagents; Damien Grégoire and Ula Hibner for CDK19 sgRNA and CDK8-WT constructs; and, Neus Prats and the IRB Histopathology Unit for assistance in the analyses of human teratomas. We are also thankful to Maribel Muñoz-Martín (CNIO; IRB) and Luis César Fernández (CNIO) for assistance in post-implantation E6.5 embryo isolation, and to the CNIO core facilities, in particular the Proteomic, Bioinformatic, Cytometry, Confocal and Transgenic Mice Units. M.N.S is funded by a Leverhulme Trust early career fellowship. Work in the lab of M.Z-G is funded by the Wellcome Trust (098287/Z/12/Z) and ERC (669198). I.A. and P.S. were supported by the Infrastructure Nationale en Biologie et Santé INGESTEM (ANR-11-INBS-0009), the IHU-B CESAME (ANR-10-IBHU-003), and the LabEx “DEVweCAN” (ANR-10-LABX-0061) and the LabEx "CORTEX" (ANR-11-LABX-0042) of the University of Lyon within the program ‘Investissements d’Avenir’ (ANR-11-IDEX-0007). Work in the laboratory of N.M. was funded by the European Research Council (ERC), under the European Union Horizon 2020 research and innovation program (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), through the projects SAF2017-89782-R, SAF2015-72617-EXP and RYC-2014-16242, and the CERCA/Government of Catalonia (2017 SGR 1306). Work in the Unit of S.O. was funded by grant 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 (M.I.S.); 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 the Spanish Ministry of Economy co-funded by the European Regional Development Fund (ERDF) (SAF2013-48256-R), the European Research Council (ERC-2014-AdG/669622), the Botin Foundation and

Banco Santander (Santander Universities Global Division), and the “laCaixa” Foundation.

AUTHOR CONTRIBUTIONS

C.J.L. performed molecular analyses in cells and embryos, contributed to experimental design, bioinformatic data analysis, and co-wrote the manuscript. R.B. performed human cell culture, teratoma assays, and molecular analyses. A.M.-V performed proteomic and bioinformatic 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. adapted human PSCs to the different experimental conditions, and performed three germ layer 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 bioinformatic 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 work; I.A. and P.S. performed human-rabbit interspecies chimera assays; S.P., E.S., A.C., and D.F. generated the CDK8-knockout 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, analyzed the data, and co-wrote the manuscript. All authors discussed the results and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
7. D'Alessio, J. A., Wright, K. J. & Tjian, R. Shifting Players and Paradigms in Cell-Specific Transcription. *Mol. Cell* **36**, 924–931 (2009).
8. Soutourina, J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **4**, 262–274 (2017).
9. 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).
10. Fant, C. B. & Taatjes, D. J. Regulatory functions of the Mediator kinases CDK8 and CDK19. *Transcription* **00**, 1–15 (2018).
11. Dannappel, M. V., Sooraj, D., Loh, J. J., Firestein, R. & Chan, G. Molecular and in vivo Functions of the CDK8 and CDK19 Kinase Modules. **6**, 1–9 (2019).
12. Chen, H. *et al.* A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* **173**, 386–399.e12 (2018).
13. Wu, J. & Izpisua Belmonte, J. C. Dynamic Pluripotent Stem Cell States and Their Applications. *Cell Stem Cell* **17**, 509–525 (2015).

14. Flanagan, P. M., Kelleher, R. J., Sayre, M. H., Tschochner, H. & Kornberg, R. D. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **350**, 436–438 (1991).

15. Clark, A. D., Oldenbroek, M. & Boyer, T. G. Mediator kinase module and human tumorigenesis. *Crit. Rev. Biochem. Mol. Biol.* **50**, 393–426 (2015).

16. Malik, S. & Roeder, R. G. Mediator: A Drawbridge across the Enhancer-Promoter Divide. *Mol. Cell* **64**, 433–434 (2016).

17. Pelish, H. E. *et al.* Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **526**, 273–276 (2015).

18. Hengartner, C. J. *et al.* Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* **2**, 43–53 (1998).

19. Galbraith, M. D., Donner, A. J. & Espinosa, J. M. CDK8: a positive regulator of transcription. *Transcription* **1**, 4–12 (2010).

20. 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).

21. 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).

22. 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).

23. Gonzalez, D. *et al.* Suppression of Mediator is regulated by Cdk8-dependent Grr1 turnover of the Med3 coactivator. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2500–5 (2014).

24. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012).

25. Alarcón, C. *et al.* Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-β Pathways. *Cell* **139**, 757–769 (2009).

26. Bancerek, J. *et al.* CDK8 Kinase Phosphorylates Transcription Factor

STAT1 to Selectively Regulate the Interferon Response. *Immunity* **38**, 250–262 (2013).

27. Fryer, C. J., White, J. B. & Jones, K. A. Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Mol. Cell* **16**, 509–520 (2004).

28. Nelson, C., Goto, S., Lund, K., Hung, W. & Sadowski, I. Srb10/Cdk8 regulates yeast filamentous growth by phosphorylating the transcription factor Ste12. *Nature* **421**, 187–90 (2003).

29. Raithatha, S., Su, T.-C., Lourenco, P., Goto, S. & Sadowski, I. Cdk8 regulates stability of the transcription factor Phd1 to control pseudohyphal differentiation of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **32**, 664–74 (2012).

30. Chi, Y. *et al.* Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev.* **15**, 1078–92 (2001).

31. Savatier, P., Osteil, P. & Tam, P. P. L. Pluripotency of embryo-derived stem cells from rodents, lagomorphs, and primates: Slippery slope, terrace and cliff. *Stem Cell Res.* **19**, 104–112 (2017).

32. 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).

33. Liu, X. *et al.* Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods* **11**, 1055–1062 (2017).

34. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).

35. Ying, Q.-L. & Smith, A. The Art of Capturing Pluripotency: Creating the Right Culture. *Stem Cell Reports* **8**, 1457–1464 (2017).

36. Hackett, J. A. & Surani, M. A. Regulatory principles of pluripotency: From the ground state up. *Cell Stem Cell* **15**, 416–430 (2014).

37. Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell*

4, 487–492 (2009).

38. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).

39. Horne, G. A. *et al.* Nanog Requires BRD4 to Maintain Murine Embryonic Stem Cell Pluripotency and Is Suppressed by Bromodomain Inhibitor JQ1 Together with Lefty1. *Stem Cells Dev.* **24**, 879–891 (2015).

40. Liu, W. *et al.* BRD4 regulates Nanog expression in mouse embryonic stem cells and preimplantation embryos. *Cell Death Differ.* **21**, 1950–60 (2014).

41. Bhagwat, A. S. *et al.* BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. *Cell Rep.* **15**, 519–530 (2016).

42. 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).

43. Finley, L. W. S. *et al.* Pluripotency transcription factors and Tet1/2 maintain Brd4-independent stem cell identity. *Nat. Cell Biol.* (2018). doi:10.1038/s41556-018-0086-3

44. 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).

45. 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).

46. 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).

47. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).

48. Torres-Padilla, M.-E. & Chambers, I. Transcription factor heterogeneity in

pluripotent stem cells: a stochastic advantage. *Development* **141**, 2173–2181 (2014).

49. 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).

50. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. **1069**, 1063–1069 (2009).

51. Greber, B. *et al.* Conserved and Divergent Roles of FGF Signaling in Mouse Epiblast Stem Cells and Human Embryonic Stem Cells. *Cell Stem Cell* **6**, 215–226 (2010).

52. Bao, S. *et al.* Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461**, 1292–1295 (2009).

53. Betschinger, J. *et al.* Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* **153**, 335–347 (2013).

54. Chen, H. *et al.* Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *Nat. Commun.* **6**, 7095 (2015).

55. Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409 (2014).

56. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).

57. Wang, J. *et al.* Isolation and cultivation of naive-like human pluripotent stem cells based on HERVH expression. *Nat. Protoc.* **11**, 327–346 (2016).

58. Takashima, Y. *et al.* Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* **158**, 1254–1269 (2014).

59. 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).

60. Wu, J. *et al.* Stem cells and interspecies chimaeras. *Nature* **540**, 51–59 (2016).

61. Huang, K. *et al.* BMI1 enables interspecies chimerism with human pluripotent stem cells. *Nat. Commun.* **9**, 4649 (2018).

62. Wu, J. *et al.* Interspecies Chimerism with Mammalian Pluripotent Stem Cells. *Cell* **168**, 473–486.e15 (2017).

63. Osteil, P. *et al.* Induced pluripotent stem cells derived from rabbits exhibit some characteristics of naive pluripotency. *Biol. Open* **2**, 613–628 (2013).

64. Tachibana, M. *et al.* Generation of Chimeric Rhesus Monkeys. *Cell* **148**, 285–295 (2012).

65. Masaki, H. *et al.* Interspecific in vitro assay for the chimera-forming ability of human pluripotent stem cells. *Development* **142**, 3222–30 (2015).

66. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19**, 502–515 (2016).

67. Mascetti, V. L. & Pedersen, R. A. Contributions of Mammalian Chimeras to Pluripotent Stem Cell Research. *Cell Stem Cell* **19**, 163–175 (2016).

68. Mascetti, V. L. & Pedersen, R. A. Human-Mouse Chimerism Validates Human Stem Cell Pluripotency. *Cell Stem Cell* **18**, 67–72 (2016).

69. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273–287 (2011).

70. 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).

71. Fadloun, A. *et al.* Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **20**, 332–338 (2013).

72. 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).

73. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).

74. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with

endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).

75. 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).

76. Zalzman, M. *et al.* Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).

77. Guo, G. *et al.* Epigenetic resetting of human pluripotency. *Development* **144**, 2748–2763 (2017).

78. Boroviak, T. *et al.* Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev. Cell* **35**, 366–382 (2015).

79. Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3613–3613 (2015).

80. Nakamura, T. *et al.* A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016).

81. 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).

82. Xue, Z. *et al.* Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).

83. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012–1026 (2016).

84. 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).

85. Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).

86. Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA

hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–6 (2013).

87. 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).

88. Choi, J. *et al.* Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219–223 (2017).

89. Yagi, M. *et al.* Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548**, 224–227 (2017).

90. Sahakyan, A. *et al.* Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation Human Naive Pluripotent Stem Cells Model. *Stem Cell* 1–15 (2017). doi:10.1016/j.stem.2016.10.006

91. Vallot, C. *et al.* XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell* **20**, 102–111 (2017).

92. Vallot, C. *et al.* Erosion of X chromosome inactivation in human pluripotent cells initiates with XACT coating and depends on a specific heterochromatin landscape. *Cell Stem Cell* **16**, 533–546 (2015).

93. 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).

94. Mekhoubad, S., Bock, C., Boer, A. S. De, Kiskinis, E. & Meissner, A. Erosion of Dosage Compensation Impacts Human iPSC Disease Modeling. *Cell Stem Cell* **10**, 595–609 (2012).

95. Rahl, P. B. *et al.* C-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).

96. 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).

97. Chen, F. X. *et al.* PAF1, a molecular regulator of promoter-proximal pausing by RNA Polymerase II. *Cell* **162**, 1003–1015 (2015).

98. Arner, E. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science (80-.).* **347**, 1010–1014 (2015).

99. Zhu, Y. *et al.* Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res.* **41**, 10032–10043 (2013).

100. Respuela, P. *et al.* Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification. *Cell Stem Cell* **18**, 118–133 (2016).

101. Westerling, T., Kuuluvainen, E. & Makela, T. P. Cdk8 Is Essential for Preimplantation Mouse Development. *Mol. Cell. Biol.* **27**, 6177–6182 (2007).

102. Wang, Q. T. *et al.* A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev. Cell* **6**, 133–144 (2004).

103. Xie, D. *et al.* Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* **20**, 804–815 (2010).

104. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Dev. Cell* **10**, 615–624 (2006).

105. Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724 (2010).

106. Bedzhov, I. & Zernicka-Goetz, M. Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* **156**, 1032–1044 (2014).

107. Shahbazi, M. N. *et al.* Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* (2017). doi:10.1038/nature24675

108. 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).

109. Audetat, K. A. *et al.* A Kinase-Independent Role for Cyclin-Dependent Kinase 19 in p53 Response. *Mol. Cell. Biol.* **37**, e00626-16 (2017).

110. Zhu, Y. *et al.* CYCLIN-DEPENDENT KINASE8 Differentially Regulates Plant Immunity to Fungal Pathogens through Kinase-Dependent and -Independent Functions in Arabidopsis. *Plant Cell* **26**, 4149–4170 (2014).

111. Tsutsui, T. *et al.* Mediator complex recruits epigenetic regulators via its two cyclin-dependent kinase subunits to repress transcription of immune response genes. *J. Biol. Chem.* **288**, 20955–20965 (2013).

112. Kuuluvainen, E. *et al.* Cyclin-dependent kinase 8 module expression profiling reveals requirement of Mediator subunits 12 and 13 for transcription of Serpent-dependent innate immunity genes in Drosophila. *J. Biol. Chem.* **289**, 16252–16261 (2014).

113. Kuuluvainen, E., Domènech-Moreno, E., Niemelä, E. H. & Mäkelä, T. P. Depletion of Mediator kinase module subunits represses super-enhancer associated genes in colon cancer cells. *Mol. Cell. Biol.* MCB.00573-17 (2018). doi:10.1128/MCB.00573-17

114. Li, N. *et al.* Cyclin C is a haploinsufficient tumour suppressor. *Nat. Cell Biol.* **16**, 1080–1091 (2014).

115. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772 (2010).

116. 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).

117. Percharde, M., Bulut-Karslioglu, A. & Ramalho-Santos, M. Hypertranscription in Development, Stem Cells, and Regeneration. *Dev. Cell* **40**, 9–21 (2017).

118. Efroni, S. *et al.* Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2**, 437–447 (2008).

119. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**, 629–643 (2017).

120. Lovén, J. *et al.* Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell* **153**, 320–334 (2013).

121. Morgani, S., Nichols, J. & Hadjantonakis, A.-K. The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. *BMC Dev. Biol.* **17**, 7 (2017).

122. Boroviak, T. & Nichols, J. The birth of embryonic pluripotency. *Philos. Trans. R. Soc.* **369**, 20130541 (2014).

123. Harrison, S. E., Sozen, B., Christodoulou, N., Kyprianou, C. & Zernicka-Goetz, M. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science (80-.).* **356**, eaal1810 (2017).

124. Dimitrova, E. *et al.* FBXL19 recruits CDK-Mediator to CpG islands of developmental genes priming them for activation during lineage commitment. *Elife* **7**, 1–27 (2018).

125. Staab, J., Herrmann-Lingen, C. & Meyer, T. CDK8 as the STAT1 serine 727 kinase? *JAK-STAT* **2**, e24275 (2013).

126. McDermott, M. S. J. *et al.* Inhibition of CDK8 mediator kinase suppresses estrogen dependent transcription and the growth of estrogen receptor positive breast cancer. *Oncotarget* **8**, 12558–12575 (2017).

127. Xu, W. *et al.* Mutated K-ras activates CDK8 to stimulate the epithelial-to-mesenchymal transition in pancreatic cancer in part via the Wnt/β-catenin signaling pathway. *Cancer Lett.* **356**, 613–27 (2015).

128. Li, N., Mclaren, J. E., Michael, D. R., Fielding, C. A. & Ramji, D. P. Uptake of Modified Lipoproteins by Human Genes Implicated in Atherosclerosis, and the Activation of STAT1, the Expression of Key ERK Is Integral to the IFN-g-Mediated Macrophages. **185**, 3041–3048 (2010).

129. Kunath, T. *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895–2902 (2007).

130. Hackett, J. A. *et al.* Synergistic Mechanisms of DNA Demethylation during Transition to Ground- State Pluripotency. *Stem Cell Reports* **1**, 518–531 (2013).

131. Sim, Y.-J. *et al.* 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. *Stem cell reports* **8**, 1312–1328 (2017).

132. Jeltsch, A. & Jurkowska, R. Z. Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic Acids Res.* **44**, 8556–8575 (2016).

133. Meng, Y. *et al.* GATA6 phosphorylation by Erk1/2 propels exit from pluripotency and commitment to primitive endoderm. *Dev. Biol.* (2018). doi:10.1016/j.ydbio.2018.02.007

134. De Los Angeles, A. *et al.* Hallmarks of pluripotency. *Nature* **525**, 469–478 (2015).

135. Sahakyan, A. *et al.* Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* **20**, (2017).

136. Collier, A. J. *et al.* Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. *Cell Stem Cell* 1–17 (2016). doi:10.1016/j.stem.2017.02.014

137. Di Stefano, B. *et al.* Reduced MEK inhibition preserves genomic stability in naive human embryonic stem cells. *Nat. Methods* **15**, 732–740 (2018).

138. Lee, J. *et al.* Lineage-Specific Differentiation Is Influenced by State of Human Pluripotency Article Lineage-Specific Differentiation Is Influenced by State of Human Pluripotency. *CellReports* **19**, 20–35 (2017).

139. Bedzhov, I., Leung, C. Y., Bialecka, M. & Zernicka-Goetz, M. In vitro culture of mouse blastocysts beyond the implantation stages. *Nat. Protoc.* **9**, 2732–9 (2014).

140. 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).

141. Castro-Diaz, N. *et al.* Evolutionally dynamic L1 regulation in embryonic

stem cells. *Genes Dev.* **28**, 1397–1409 (2014).

142. Pezic, D., Manakov, S. A., Sachidanandam, R. & Aravin, A. A. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev.* **28**, 1410–1428 (2014).

143. Sookdeo, A., Hepp, C. M., McClure, M. A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob. DNA* **4**, 1–15 (2013).

144. Boissinot, S. & Sookdeo, A. The Evolution of Line-1 in Vertebrates. *Genome Biol. Evol.* **8**, evw247 (2016).

145. Amano, T. *et al.* Zscan4 restores the developmental potency of embryonic stem cells. *Nat. Commun.* **4**, 1–10 (2013).

146. 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).

147. Fernandez, A. F., Valledor, L., Vallejo, F., Cañal, M. J. & Fraga, M. F. Quantification of Global DNA Methylation Levels by Mass Spectrometry. in *Methods in molecular biology (Clifton, N.J.)* **1708**, 49–58 (2018).

148. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).

149. Buecker, C. *et al.* Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).

Fig. 1

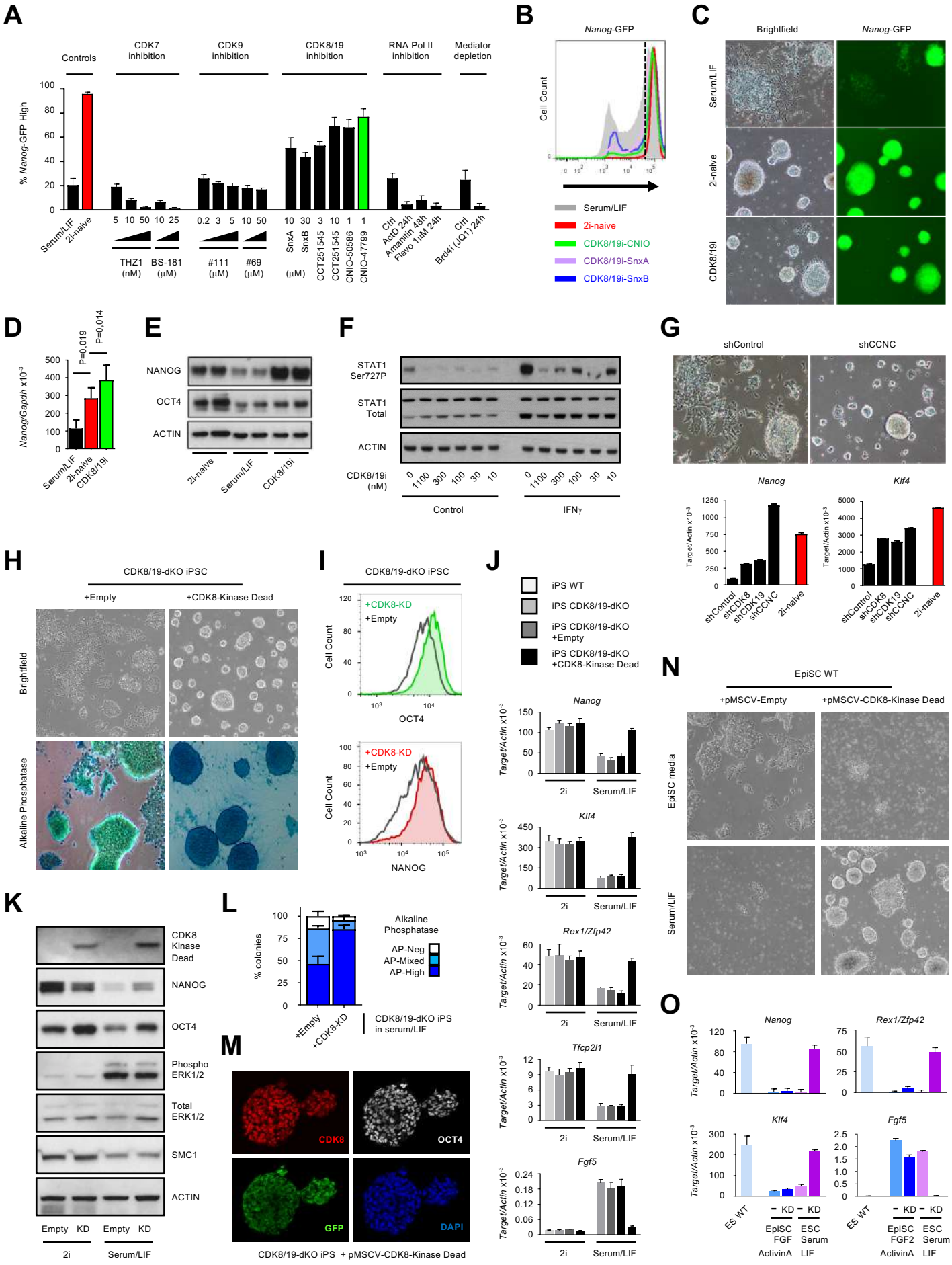

A
Controls
CDK7 inhibition
CDK9 inhibition
CDK8/19 inhibition
RNA Pol II inhibition
Mediator depletion
% Nanog-GFP High
Serum/LIF
2i-naive
THZ1 (nM)
BS-181 (μM)
#111 (μM)
#69 (μM)
SnxA
SnxB
CCT251545
CCT251545
CNIO-50586
CNIO-47799
Ctrl
Acid 24h
Amanitin 48h
Flavo 1μM 24h
Ctrl
Brd4i (JQ1) 24h
B
Nanog-GFP
Cell Count
Serum/LIF
2i-naive
CDK8/19i-CNIO
CDK8/19i-SnxA
CDK8/19i-SnxB
C
Brightfield
Nanog-GFP
Serum/LIF
2i-naive
CDK8/19i
D
P=0,019
P=0,014
Nanog/Gapdh x10-3
Serum/LIF
2i-naive
CDK8/19i
E
NANOG
OCT4
ACTIN
2i-naive
Serum/LIF
CDK8/19i
F
STAT1 Ser727P
STAT1 Total
ACTIN
CDK8/19i (nM)
Control
IFNγ
G
shControl
shCCNC
Nanog
Klf4
Target/Actin x10-3
shControl
shCDK8
shCDK19
shCCNC
2i-naive
H
CDK8/19-dKO iPSC
+Empty
+CDK8-Kinase Dead
Brightfield
Alkaline Phosphatase
I
CDK8/19-dKO iPSC
+CDK8-KD
+Empty
Cell Count
OCT4
NANOG
J
iPS WT
iPS CDK8/19-dKO
iPS CDK8/19-dKO +Empty
iPS CDK8/19-dKO +CDK8-Kinase Dead
Nanog
Klf4
Rex1/Zfp42
Tfcp2l1
Fgf5
Target/Actin x10-3
2i
Serum/LIF
K
CDK8 Kinase Dead
NANOG
OCT4
Phospho ERK1/2
Total ERK1/2
SMC1
ACTIN
Empty
KD
2i
Serum/LIF
L
% colonies
Alkaline Phosphatase
AP-Neg
AP-Mixed
AP-High
+Empty
+CDK8-KD
CDK8/19-dKO iPS in serum/LIF
M
CDK8
OCT4
GFP
DAPI
CDK8/19-dKO iPS + pMSCV-CDK8-Kinase Dead
N
EpiSC WT
+pMSCV-Empty
+pMSCV-CDK8-Kinase Dead
EpiSC media
Serum/LIF
O
Nanog
Rex1/Zfp42
Klf4
Fgf5
Target/Actin x10-3
ES WT
KD
EpiSC FGF ActivinA
ESC Serum LIF
EpiSC FGF2 ActivinA
ESC Serum LIF

FIGURE LEGENDS

Figure 1: An Inhibitor Screen for Promotion of ES Naïve State identifies a distinct role for Mediator kinase activity.

(A) Effect of indicated treatments on Nanog expression per cell, using a mouse *Nanog*-GFP knockin reporter ES cell line in standard serum/LIF base media. Percentage of *Nanog*-GFP-high determined by FACS. Data representative of 3 independent experiments.

(B) Typical FACS histogram of *Nanog*-GFP expression. $Nanog^{Low}$ and $Nanog^{High}$ cell populations are evident in the serum/LIF population (grey). Dotted line indicates the threshold where >95% of the cells are $Nanog^{High}$ in the 2i-naïve culture condition. Data representative of 3 independent experiments.

(C) ES cell colony morphology in the indicated treatments. Brightfield and *Nanog*-GFP shown.

(D,E) Endogenous *Nanog* mRNA (D), or protein (E), expression levels in mouse ES cells adapted to culture in the indicated conditions. Data representative of 3 independent experiments. In (D), Mean +/- Std Dev, * $P < 0,05$.

(F) Phosphorylation levels of the CDK8-target STAT1 Ser727 by Western blot. HERVH human iPS cells were treated with the indicated range of CDK8/19i concentrations for 3 hours, +/- simultaneous induction of STAT1-Ser727P by γ-interferon for 3 hours. Data representative of 2 independent experiments.

(G) Cell morphology and qRT-PCR of mRNA expression in mouse ES cells following 7 days of lentiviral shRNA-mediated knockdown of CDK8, CDK19, or CyclinC (CCNC). Mean +/-SD of 3 replicates.

(H) Typical cell morphology and alkaline phosphatase staining of CDK8/19-double-knockout (CDK8/19-dKO) iPS cells stably expressing pMSCV-Empty or pMSCV-CDK8-Kinase Dead (CDK8-KD). Data representative of 3 independent clones.

(I) FACS analysis of endogenous NANOG and OCT4 protein levels in CDK8/19-dKO iPS cells stably expressing pMSCV-Empty or pMSCV-CDK8-KD. Data representative of 3 independent clones.

(J) qRT-PCR of mRNA expression in WT iPS, or CDK8/19-dKO iPS cells stably expressing pMSCV-Empty or pMSCV-CDK8-KD, adapted to the indicated media conditions. Data are mean and Std Dev from 3 independent clones.

(K) Western blot analysis of protein expression in CDK8/19-dKO iPS cells stably expressing pMSCV-Empty or pMSCV-CDK8-KD, adapted to the indicated media conditions. Data representative of 2 independent experiments.

(L) Alkaline phosphatase (AP) staining becomes homogenously high in cells expressing CDK8-KD. Cells were fixed and stained at day 14 after retroviral infection and selection with pMSCV-Empty or pMSCV-CDK8-KD. Cells were seeded at clonal density and allowed to form colonies for 5-7 days. Staining intensity was scored visually for each colony, using ten fields of view. Data represent three cell experiments.

(M) Immunofluorescence for the indicated markers in CDK8/19-dKO iPS cells expressing pMSCV-CDK8-KD-puro-IRES-GFP, confirming stable expression of CDK8-KD and GFP in all cells.

(N,O) Cell morphology (N) and qRT-PCR of mRNA expression (O) of WT mouse EpiSC infected with pMSCV-Empty or pMSCV-CDK8-KD, which were then maintained for 1 passage in EpiSC media (Fgf2/ActivinA/fibronectin; see **Methods**) or 1 passage in standard serum/LIF ES media (see also Figure S1N). Data representative of 3 independent expts at day 7 after pMSCV infection and selection.

Fig. 2

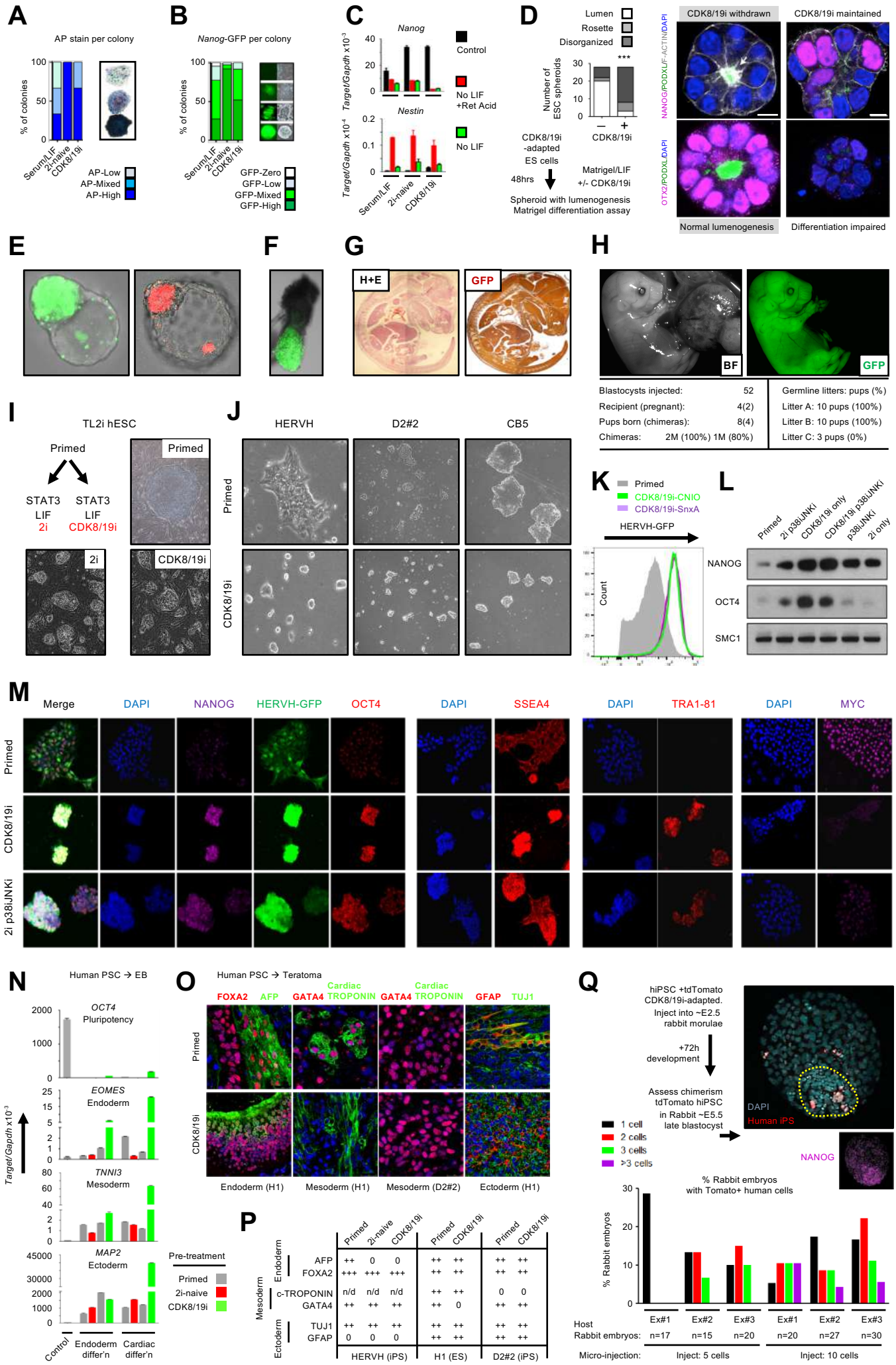

A
AP stain per colony
% of colonies
Serum/LIF
2i-naive
CDK8/19i
AP-Low
AP-Mixed
AP-High
B
Nanog-GFP per colony
GFP-Zero
GFP-Low
GFP-Mixed
GFP-High
C
Nanog
Nestin
Target/Gapdh x10-3
Target/Gapdh x10-4
Control
No LIF +Ret Acid
No LIF
D
Lumen
Rosette
Disorganized
Number of ESC spheroids
CDK8/19i -adapted ES cells
48hrs
Matrigel/LIF +/- CDK8/19i
Spheroid with lumenogenesis
Matrigel differentiation assay
CDK8/19i withdrawn
CDK8/19i maintained
NANOG/PODXL/F-ACTIN/DAPI
OTX2/PODXL/DAPI
Normal lumenogenesis
Differentiation impaired
E
F
G
H+E
GFP
H
BF
GFP
Blastocysts injected: 52
Recipient (pregnant): 4(2)
Pups born (chimeras): 8(4)
Chimeras: 2M (100%) 1M (80%)
Germline litters: pups (%)
Litter A: 10 pups (100%)
Litter B: 10 pups (100%)
Litter C: 3 pups (0%)
I
TL2i hESC
Primed
STAT3 LIF 2i
STAT3 LIF CDK8/19i
2i
CDK8/19i
J
HERVH
D2#2
CB5
K
Primed
CDK8/19i-CNIO
CDK8/19i-SnxA
HERVH-GFP
Count
L
Primed
2i p38iJNKi
CDK8/19i only
CDK8/19i p38iJNKi
p38iJNKi
2i only
NANOG
OCT4
SMC1
M
Merge
DAPI
NANOG
HERVH-GFP
OCT4
SSEA4
TRA1-81
MYC
N
Human PSC → EB
OCT4 Pluripotency
EOMES Endoderm
TNNI3 Mesoderm
MAP2 Ectoderm
Target/Gapdh x10-3
Pre-treatment
Primed
2i-naive
CDK8/19i
Control
Endoderm differ'n
Cardiac differ'n
O
Human PSC → Teratoma
FOXA2
AFP
GATA4
Cardiac TROPONIN
GFAP
TUJ1
Endoderm (H1)
Mesoderm (H1)
Mesoderm (D2#2)
Ectoderm (H1)
P
Endoderm
Mesoderm
Ectoderm
AFP
FOXA2
c-TROPONIN
GATA4
TUJ1
GFAP
HERVH (iPS)
H1 (ES)
D2#2 (iPS)
Q
hiPSC +tdTomato CDK8/19i-adapted. Inject into ~E2.5 rabbit morulae
+72h development
Assess chimerism tdTomato hiPSC in Rabbit ~E5.5 late blastocyst
1 cell
2 cells
3 cells
>3 cells
DAPI
Human iPS
NANOG
% Rabbit embryos with Tomato+ human cells
% Rabbit embryos
Host Rabbit embryos:
Ex#1 n=17
Ex#2 n=15
Ex#3 n=20
Ex#1 n=20
Ex#2 n=27
Ex#3 n=30
Micro-injection:
Inject: 5 cells
Inject: 10 cells

Figure 2: Positive effect of long-term CDK8/19i on mammalian ES self-renewal and pluripotency

(A,B) Assays of clonogenicity of mouse PSCs cultured in control, 2i, or CDK8/19i conditions. *Nanog*-GFP reporter ES cells were FACS sorted to 1-cell per well in 96-well plates, followed by culture in the indicated media for 7 days. At day 7, cell colonies were fixed and stained for Alkaline phosphatase (A), or scored for their *Nanog*-GFP staining intensity (B), to assess their pluripotent status. *Nanog*-GFP or alkaline phosphatase staining intensity was scored visually for each colony in ten fields of view (examples of single-colony staining are shown on right of each graph). Data suggests the overall ability of the seeded cells to retain or enhance pluripotency (medium or homogenously high staining, respectively) in either standard serum/LIF, 2i-naïve, or CDK8/19i culture conditions. Data representative of 3 independent experiments.

(C, D) Differentiation assays in vitro to confirm developmental capacity of mouse PSCs previously adapted to control, 2i or CDK8/19i conditions. In (C), PSC differentiation was by LIF-removal or LIF-removal plus retinoic acid (see **Methods**). Differentiation was assessed by qRT-PCR to show loss of pluripotency (*Nanog*) and induction of differentiation (*Nestin*). Data representative of 2 independent experiments, showing Mean +/- SD of 3 technical replicates. In (D), PSC pluripotency exit was assessed by seeding PSCs in 3D-matrigel/LIF to permit observation of the three earliest stages of epiblast development, that is, rosette formation, polarization, and lumenogenesis. PSCs that exit from the naïve pluripotent state undergo polarization and lumenogenesis whereas cells locked in the naïve state fail to undergo lumenogenesis [107] [106,123]. Assessment of these morphological events was by immunofluorescence in the developing embryoids for NANOG, PODOCALYXIN (PODXL), F-ACTIN, and OTX2. Data representative of 3 independent experiments. Scale bars 10 μm.

(E-H) In vivo assays to confirm developmental capacity of mouse PSCs previously adapted to CDK8/19i conditions. Constitutively-labelled PSCs (ROSA26-GFP or Tg.CAG-Katushka) were used to generate chimeric embryos by aggregation or micro-injection into host E2.5 morulae (see **Methods**). Chimerism was assessed visually at E4.5 blastocyst (E), E6.5 egg cylinder (F), E14.5 (G), perinatal E19.5

(H), and by germline transmission after mating adult chimeras (H). In (H), the parenthesis refer to the percent chimerism determined by coat colour, and for germline transmission, the percentage of pups with coat colour gernline transmission.

(I) Induction of naïve colony morphology in human OSCAR ES cells by tamoxifen-inducible consitutively active STAT3, LIF, and 2i (TL2i) as reported [54], or by substituting CDK8/19i for 2i (TLCDK8/19i). The inclusion of 2i or CDK8/19i was required to induce naïve morphology. Images representative of 3 independent cell experiments.

(J) Brightfield images showing colony morphology in 2 human PSC lines in primed state (upper panels), or following 14 days treatment with CDK8/19i. Images representative of at least 5 independent cell experiments.

(K) Cytometry of *HERVH*-GFP intensity per cell in human PSCs adapted to primed or CDK8/19i conditions, as in (J) above. Data representative of at least 3 independent cell experiments.

(L) Western blots of pluripotency markers NANOG and OCT4 in human PSCs cultured as indicated in (K) above with 2i-based or CDK8/19i-based media (see **Methods**). SMC1 is a nuclear protein as internal loading control.

(M) Immunofluorescence for pluripotency markers in human PSCs cultured as indicated, as in (L) above.

(N) Human PSCs previously adapted to the indicated media conditions (primed, 2i-naïve, or CDK8/19i) were subsequently tested by embryoid body differentiation assay in vitro, using endoderm-directed, or caridac-directed protocols (see **Methods**). Plots display the expression levels of selected markers for pluripotency or the 3 embryonic germ layers, assessed by qRT-PCR (data for H9 cell line shown). Data are representative of two human PSC lines (H1 and H9). See also Table S1, Sheets #8, #9, for full summary of all tested lineage markers, determined by qRT-PCR (17) or immunofluorecence (6). A specific neural-directed differentiation protocol was not performed, since EBs maintained in the same differentiation conditions used for endoderm or mesoderm also spontaneously gave rise to neural cell clusters.

(O) Human PSCs previously adapted to the indicated media conditions (primed, or CDK8/19i) were subsequently tested by teratoma differentiation assay (see **Methods**). Data are representative of three human PSC lines (H1, D2#2 and HERVH). Immunofluorescence of selected markers for the 3 embryonic germ layers is shown for the H1 and D2#2 cell lines (labelled in parentheses below each panel).

(P) Summary of all tested lineage markers (6) determined by immunofluorecence in teratomas generated from the three cell lines in (O), above. “+” = detected; “0“ = not detected; “n/d” = not determined.

(Q) Interspecies chimera developmental assay in vivo to confirm developmental capacity of human PSCs previously adapted to primed or CDK8/19i conditions. Constitutively-labelled human PSCs (td:tomato-red; HERVH iPS cell line) were used to test chimerism by introduction into host rabbit morulae of ~E2.5 (see **Methods**). Chimerism was assessed visually at the ~E5.5 rabbit blastocyst. The number of human cells introduced, and the number of human cells observed in the rabbit embryo +72hrs later are indicated in the graph. Data are from three independent experiments. A representative image shows immunofluorescence of a E.5.5 rabbit blastocyst, with the inner cell mass indicated (determined by Nanog staining, inset panel). Human PSCs previously adapted to CDK8/19i displayed moderate contribution to human-rabbit chimeras.

Fig. 3

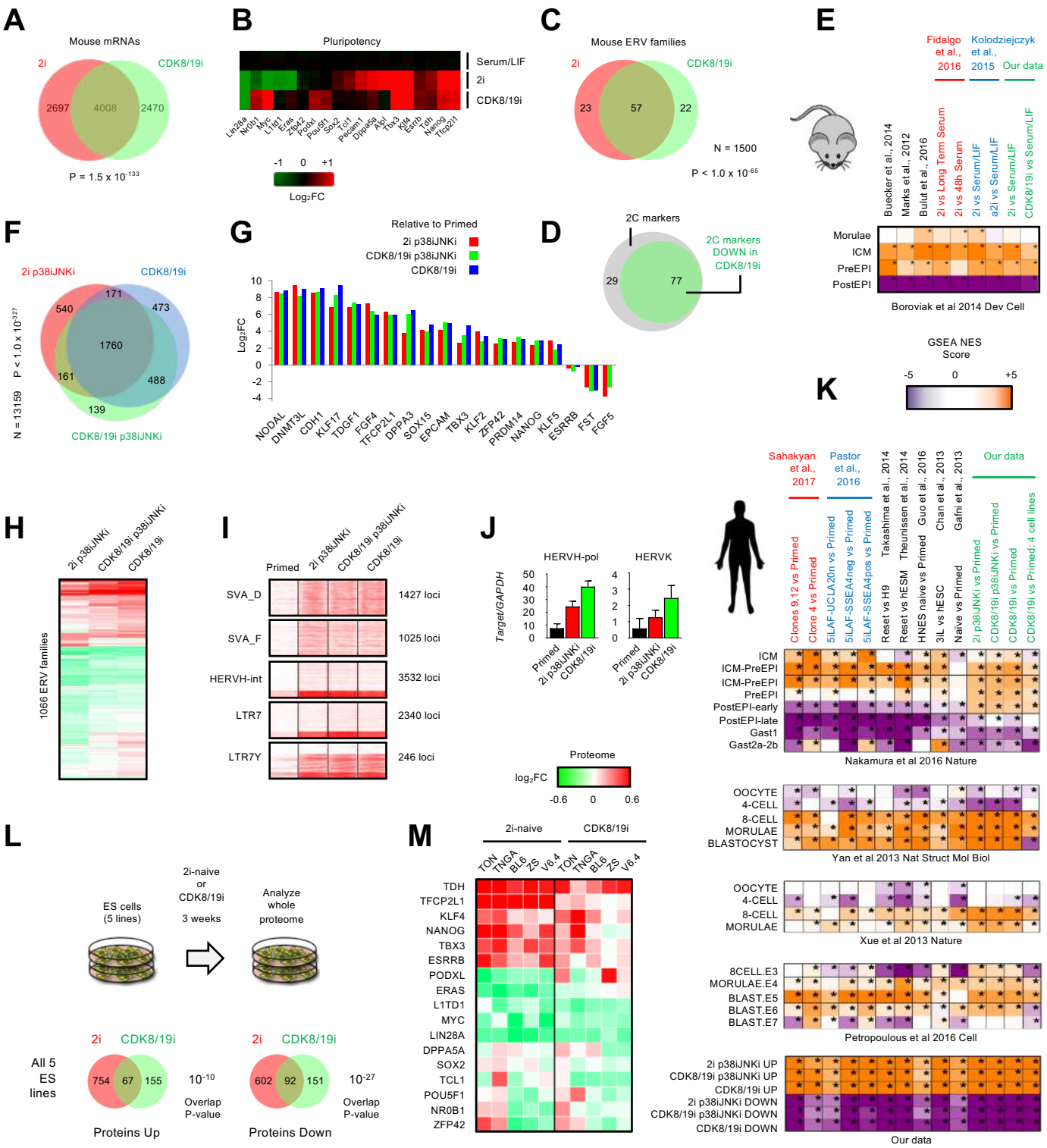

A
Mouse mRNAs
2i
CDK8/19i
2697
4008
2470
P = 1.5 x 10-133
B
Pluripotency
Serum/LIF
2i
CDK8/19i
Lin28a Nr0b1 Myc L1td1 Eras Zfp42 Podxl Pou5f1 Sox2 Tcl1 Pecam1 Dppa5a Ajap1 Tbx3 Klf4 Esrrb Tdh Nanog Tfcp2l1
-1 0 +1
Log2FC
C
Mouse ERV families
2i
CDK8/19i
23
57
22
N = 1500
P < 1.0 x 10-65
E
Fidalgo et al., 2016
Kolodziejczyk et al., 2015
Our data
Buecker et al., 2014
Marks et al., 2012
Bulut et al., 2016
2i vs Long Term Serum
2i vs 48h Serum
2i vs Serum/LIF
a2i vs Serum/LIF
2i vs Serum/LIF
CDK8/19i vs Serum/LIF
Morulae
ICM
PreEPI
PostEPI
Boroviak et al 2014 Dev Cell
F
2i p38iJNKi
CDK8/19i
171
540
473
1760
161
488
139
CDK8/19i p38iJNKi
N = 13159
P < 1.0 x 10-327
G
Relative to Primed
2i p38iJNKi
CDK8/19i p38iJNKi
CDK8/19i
Log2FC
NODAL DNMT3L CDH1 KLF17 TDGF1 FGF4 TFCP2L1 DPPA3 SOX15 EPCAM TBX3 KLF2 ZFP42 PRDM14 NANOG KLF5 ESRRB FST FGF5
D
2C markers
2C markers DOWN in CDK8/19i
29
77
K
GSEA NES Score
-5 0 +5
H
2i p38iJNKi
CDK8/19i p38iJNKi
CDK8/19i
1066 ERV families
I
Primed
2i p38iJNKi
CDK8/19i p38iJNKi
CDK8/19i
SVA_D 1427 loci
SVA_F 1025 loci
HERVH-int 3532 loci
LTR7 2340 loci
LTR7Y 246 loci
J
HERVH-pol
HERVK
Target/GAPDH
Primed
2i p38iJNKi
CDK8/19i
Proteome
log2FC
-0.6 0 0.6
Sahakyan et al., 2017
Pastor et al., 2016
Our data
Clones 9,12 vs Primed
Clone 4 vs Primed
5iLAF-UCLA20n vs Primed
5iLAF-SSEA4neg vs Primed
5iLAF-SSEA4pos vs Primed
Reset vs H9
Reset vs hESM
HNES naive vs Primed
3iL vs hESC
Naïve vs Primed
Takashima et al., 2014
Theunissen et al., 2014
Guo et al., 2016
Chan et al., 2013
Gafni et al., 2013
2i p38iJNKi vs Primed
CDK8/19i p38iJNKi vs Primed
CDK8/19i vs Primed
CDK8/19i vs Primed: 4 cell lines
ICM
ICM-PreEPI
ICM-PreEPI
PreEPI
PostEPI-early
PostEPI-late
Gast1
Gast2a-2b
Nakamura et al 2016 Nature
OOCYTE
4-CELL
8-CELL
MORULAE
BLASTOCYST
Yan et al 2013 Nat Struct Mol Biol
OOCYTE
4-CELL
8-CELL
MORULAE
Xue et al 2013 Nature
8CELL.E3
MORULAE.E4
BLAST.E5
BLAST.E6
BLAST.E7
Petropoulous et al 2016 Cell
2i p38iJNKi UP
CDK8/19i p38iJNKi UP
CDK8/19i UP
2i p38iJNKi DOWN
CDK8/19i p38iJNKi DOWN
CDK8/19i DOWN
Our data
L
ES cells (5 lines)
2i-naive or CDK8/19i
3 weeks
Analyze whole proteome
All 5 ES lines
2i
CDK8/19i
754
67
155
10-10
Overlap P-value
Proteins Up
2i
CDK8/19i
602
92
151
10-27
Overlap P-value
Proteins Down
M
2i-naive
CDK8/19i
TON TNGA BL6 ZS V6.4 TON TNGA BL6 ZS V6.4
TDH
TFCP2L1
KLF4
NANOG
TBX3
ESRRB
PODXL
ERAS
L1TD1
MYC
LIN28A
DPPA5A
SOX2
TCL1
POU5F1
NR0B1
ZFP42
N
Global DNA methylation
TON
ZS
B6
V6.4
Mean (4 lines)
%5mC
Serum/LIF
2i-naive
CDK8/19i
O
Global DNA methylation
HERVH
WIBR3
%5mC
Primed
2i p38iJNKi
CDK8/19i
P
LINE L1
CpG 2
CpG 1
CpG 3
% methylated
Serum/LIF
2i-naive
CDK8/19i
MEKi
GSK3i

Figure 3: Gene expression and DNA methylation analysis in PSCs adapted to 2i or CDK8/19i

(A) Overlap and hypergeometric significance test of differentially expressed mRNAs in mouse PSCs adapted to 2i-naïve or CDK8/19i, versus standard serum/LIF (3 replicates; FDR<0.01). See Table S2 for gene lists.

(B) Heatmap of mRNA expression changes for selected pluripotency regulators by RNAseq as in (A) above.

(C) Overlap and hypergeometric significance test of differentially expressed ERV families in mouse PSCs adapted to 2i-naïve or in CDK8/19i, versus standard serum/LIF mouse PSCs (3 replicates; FDR<0.05). See Table S2 for ERV lists.

(D) Overlap of change in expression levels of markers of the 2-Cell (2C) fluctuation in mouse PSC adapted to CDK8/19i versus serum/LIF as in (A) above. See Table S2 for gene lists.

(E) Heatmap of normalized enrichment scores (NES) in a GSEA comparison of our RNAseq data in (A) above, or five other studies (indicated above heatmap), versus stage-specific marker genesets identified during mouse preimplantation development [78]. Data (FDR q<0.05) are considered significant and marked with an asterisk (*) in the heatmaps of GSEA NES scores.

(F) Overlap and hypergeometric significance test of differentially expressed mRNAs in human PSCs adapted the indicated culture conditions versus control primed cells (RNAseq; >2x fold-change, 3 replicates, FDR<0.05).

(G) RNA expression changes determined by RNAseq are shown for selected human naïve pluripotency markers (up), or post-implantation primed epiblast markers (down) in human PSCs adapted to the indicated conditions as in (F) above.

(H) Heatmap showing correlation of differentially expressed ERV families in human PSCs adapted to the indicated culture conditions versus control primed cells, as in (F) above.

(I) Heatmap showing correlation of RNA expression from individual loci of the indicated ERV families in human PSCs adapted to the indicated culture conditions versus control primed cells, as in (F) above.

(J) RNA expression levels of HERVH or HERVK determined by qRT-PCR, in human PSCs adapted to the indicated culture conditions, as in (F) above.

(K) Heatmap of NES scores in a GSEA comparison of our RNAseq data or seven other studies (indicated above heatmap) verses stage-specific marker genesets identified during mouse preimplantation development. Bottom panel is a comparison of our data versus the other studies. Data (3 reolicates, FDR $q<0.05$) are considered significant and marked with an asterisk (*) in the heatmaps of GSEA NES scores.

(L) Upper panel: overview of proteomic analysis. Pie chart shows the proportion of the proteome which undergoes significant change in mouse PSCs adapted to the indicated treatments. Lower panel: overlap and hypergeometric significance test of differentially expressed proteins averaged across five mouse PSC lines adapted to 2i-naïve or in CDK8/19i, versus standard serum/LIF mouse PSCs (FDR<0.05). See also Figure S4E for data per cell line and Table S4 for full list of differentially expressed proteins.

(M) Heatmap of protein expression changes for key pluripotency regulators and markers in five mouse PSC cell lines in 2i-naïve or CDK8/19i conditions, compared to control standard serum/LIF. PSC cell line indicated above heatmap.

(N) Global DNA methylation (5-methyl-cytosine) changes in four mouse PSC lines adapted to 2i or CDK8/19i. Left panel, four separate cell lines. Right panel, Mean +/- SD of all four cell lines.

(O) Global DNA methylation (5-methyl-cytosine) changes in two human PSC lines adapted to 2i or CDK8/19i. Data is Mean +/- SD of 3 replicates per cell line.

(P) CpG methylation status of three specific loci in the LINE L1 family of repeat elements, in mouse PSC adapted to the indicated conditions, as in (N) above. CpG methylation status was assessed by pyrosequencing.

Fig. 4

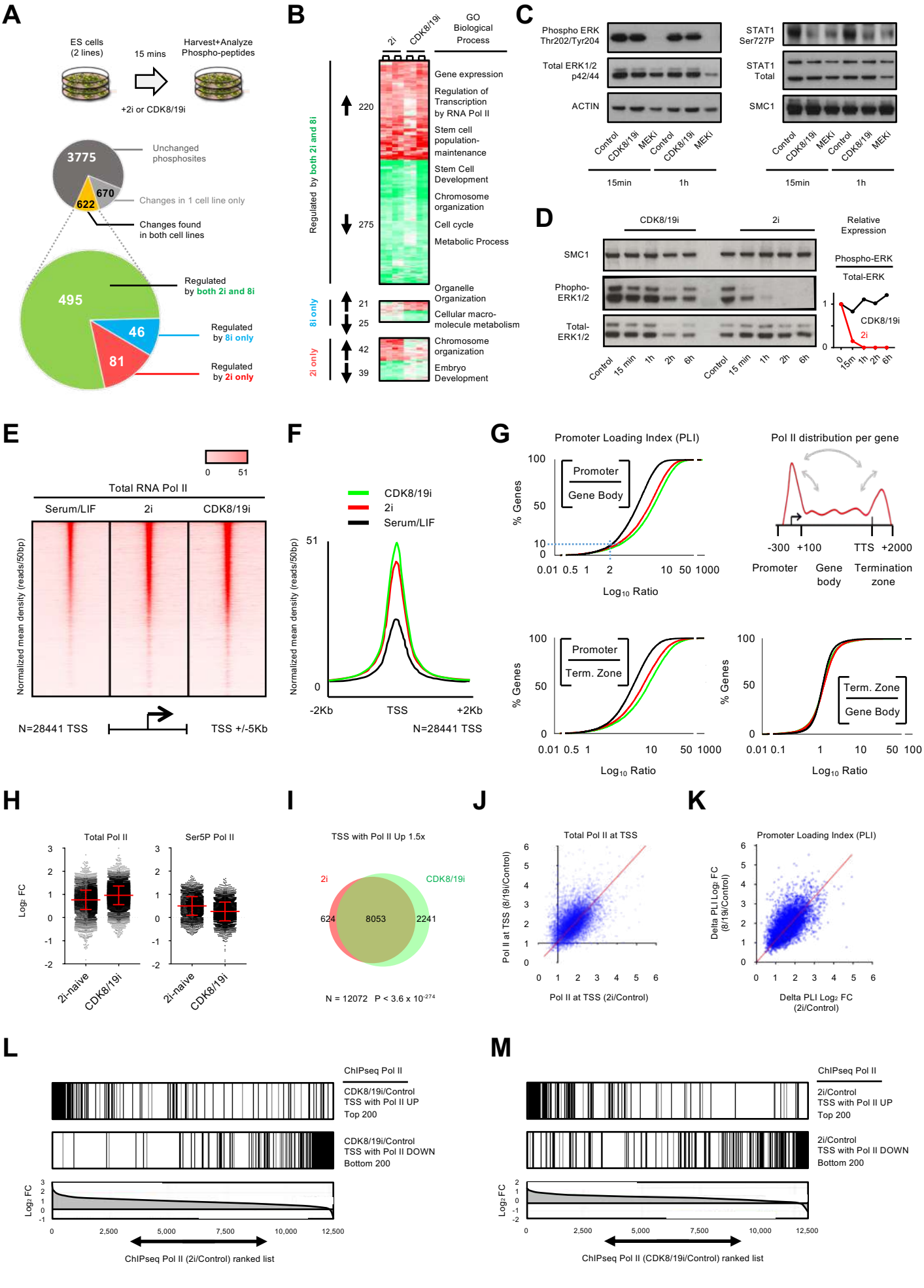

A
ES cells (2 lines)
15 mins
Harvest+Analyze Phospho-peptides
+2i or CDK8/19i
3775
Unchanged phosphosites
670
622
Changes in 1 cell line only
Changes found in both cell lines
495
Regulated by both 2i and 8i
46
Regulated by 8i only
81
Regulated by 2i only
B
2i
CDK8/19i
GO Biological Process
Regulated by both 2i and 8i
220
275
Gene expression
Regulation of Transcription by RNA Pol II
Stem cell population-maintenance
Stem Cell Development
Chromosome organization
Cell cycle
Metabolic Process
8i only
21
25
Organelle Organization
Cellular macro-molecule metabolism
2i only
42
39
Chromosome organization
Embryo Development
C
Phospho ERK Thr202/Tyr204
Total ERK1/2 p42/44
ACTIN
STAT1 Ser727P
STAT1 Total
SMC1
Control
CDK8/19i
MEKi
15min
1h
D
CDK8/19i
2i
Relative Expression
SMC1
Phopho-ERK1/2
Total-ERK1/2
Phospho-ERK
Total-ERK
CDK8/19i
2i
Control
15 min
1h
2h
6h
E
0
51
Total RNA Pol II
Serum/LIF
2i
CDK8/19i
Normalized mean density (reads/50bp)
N=28441 TSS
TSS +/-5Kb
F
CDK8/19i
2i
Serum/LIF
51
0
Normalized mean density (reads/50bp)
-2Kb
TSS
+2Kb
N=28441 TSS
G
Promoter Loading Index (PLI)
% Genes
Promoter
Gene Body
Log10 Ratio
Pol II distribution per gene
-300
+100
TTS
+2000
Promoter
Gene body
Termination zone
Promoter
Term. Zone
Term. Zone
Gene Body
H
Total Pol II
Ser5P Pol II
Log2 FC
2i-naive
CDK8/19i
I
TSS with Pol II Up 1.5x
2i
CDK8/19i
624
8053
2241
N = 12072 P < 3.6 x 10-274
J
Total Pol II at TSS
Pol II at TSS (8/19i/Control)
Pol II at TSS (2i/Control)
K
Promoter Loading Index (PLI)
Delta PLI Log2 FC (8/19i/Control)
Delta PLI Log2 FC (2i/Control)
L
ChIPseq Pol II
CDK8/19i/Control
TSS with Pol II UP
Top 200
CDK8/19i/Control
TSS with Pol II DOWN
Bottom 200
Log2 FC
ChIPseq Pol II (2i/Control) ranked list
M
ChIPseq Pol II
2i/Control
TSS with Pol II UP
Top 200
2i/Control
TSS with Pol II DOWN
Bottom 200
Log2 FC
ChIPseq Pol II (CDK8/19i/Control) ranked list

Figure 4: CDK8/19-repression regulates the phospho-proteome and global RNA Pol II loading similar to 2i-naïve pluripotency

(A) Overview of phosphoproteomic study. Two mouse ES cell lines were treated with 2i or CDK8/19i for 15 minutes. Pie-charts summarize the total and differential phosphosites detected. See Table S5 for protein list.

(B) Heatmap of phosphosites significantly altered by 15 minutes of 2i and/or CDK8/19i, as indicated in panel (A) above. Gene ontology analysis is summarised on right, for the function of the proteins on which differential phosphorylation was detected. See Table S5 for protein list.

(C) Western blot analysis of protein phosphorylation during short-term exposure of PSCs to 2i or CDK8/19i, as indicated. Left panel, ERK1/2 phosphorylation during short-term exposure of mouse ES cells to 2i or CDK8/19i. Right panel, STAT1 Ser727 phosphorylation during short-term exposure of human iPS cells to 2i or CDK8/19i.

(D) Western blot analysis of ERK1/2 phosphorylation during short-term exposure of mouse ES cells to 2i or CDK8/19i, as indicated. On right, plot of relative ERK1/2 phospho-levels, normalized by total ERK1/2 levels.

(E) Heatmaps of RNA Pol II density at all Refseq Transcription Start Sites (TSS; n = 28,441) +/-5 Kb, for mouse PSCs adapted to the indicated culture conditions, determined by ChIPseq (n= 3 pooled replicates; see **Methods**).

(F) Metagene average RNA Pol II density at all Refseq TSS (n = 28,441) +/-2Kb, for mouse PSCs adapted to the indicated culture conditions, as in (E) above.

(G) Cumulative plots of RNA Pol II loading ratios. RNA Pol II abundance at the Promoter region (-300 to +100bp around the TSS), Gene Body (TSS+100bp to Transcription Termination Site, TTS), and transcription termination zone (TTS+2Kb) was calculated. Regional abundance was then compared in the indicated ratios for each gene, as shown in the three cumulative plots (see schematic, also Table S6 for full list of Pol II abundance and regional ratios for each gene). Cumulative plots show the RNA Pol II ratios across n = 12,072 genes in mouse PSCs adapted to the indicated culture conditions, as in (E) above. For

the Promoter Loading Index (PLI; plot on left), the dotted line indicates that 90% of genes have a PLI > 2.0. See Table S6 for gene lists.

(H) Change in Total or Ser5P RNA Pol II abundance at TSS in 2i or CDK8/19i versus serum/LIF control cells, as in (E) above. Data is Mean +/- SD. TSS at which RNA Pol II was detected in serum/LIF, 2i and CDK8/19i conditions are shown (see **Methods**). Total Pol II, n = 12,693; Ser5P, n = 4,470.

(I) Overlap of genes where RNA Pol II abundance at the TSS increased > 1.5fold in 2i or CDK8/19i, n = 12,693.

(J) Fold-change in RNA Pol II abundance at TSS in mouse PSCs cultured in 2i or CDK8/19i versus control serum/LIF, n = 12,693.

(K) Fold-change in RNA Pol II Promoter Loading Index (PLI) on genes in 2i or CDK8/19i versus control serum/LIF cells. See also schematic in (G) above. Genes at which RNA Pol II was detected in serum/LIF, 2i and CDK8/19i conditions are shown, n = 12,693.

(L,M) Correlation between genes where the promoter-TSS have the greatest change in RNA Pol II abundance in CDK8/19i or 2i-naïve conditions, as in (E) above. In (L), individual TSS with RNA Pol II loading altered in CDK8/19i by the greatest (upper panel) or least amount (middle panel) versus control serum/LIF conditions (top 200 TSS in each case), are compared against a ranked list (bottom panel) of the differential Pol II loading on all TSS for 2i-naïve versus control serum/LIF conditions. (M) The reverse comparison is shown: top 200 altered TSS in 2i-naïve versus ranked list of Pol II changes in CDK8/19i.

Fig. 5

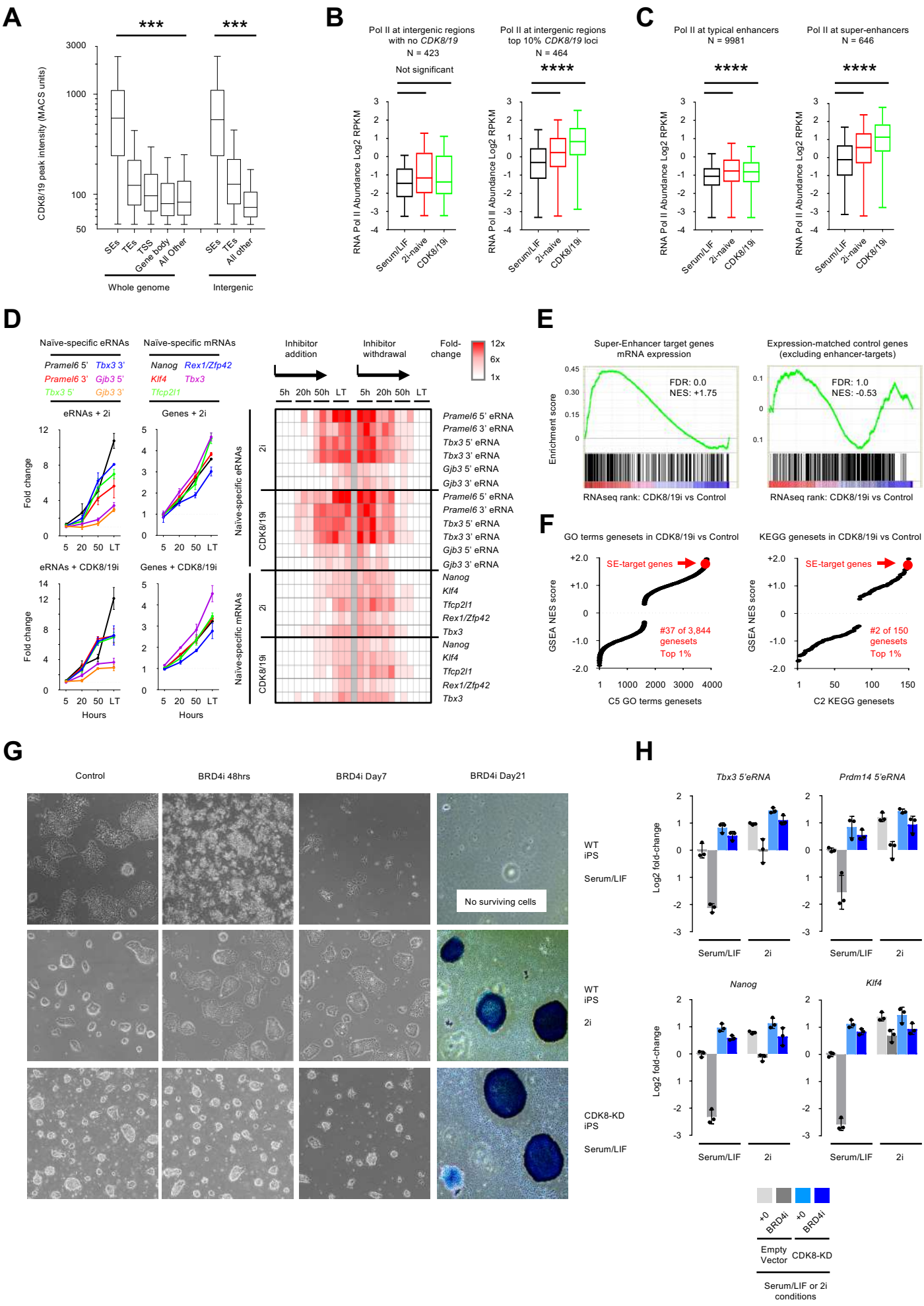

A
***
***
CDK8/19 peak intensity (MACS units)
3000
1000
100
50
SEs
TEs
TSS
Gene body
All Other
SEs
TEs
All other
Whole genome
Intergenic
B
Pol II at intergenic regions with no CDK8/19
N = 423
Not significant
Pol II at intergenic regions top 10% CDK8/19 loci
N = 464
****
RNA Pol II Abundance Log2 RPKM
Serum/LIF
2i-naive
CDK8/19i
C
Pol II at typical enhancers
N = 9981
****
Pol II at super-enhancers
N = 646
****
D
Naïve-specific eRNAs
Naïve-specific mRNAs
Pramel6 5' Tbx3 3'
Pramel6 3' Gjb3 5'
Tbx3 5' Gjb3 3'
Nanog Rex1/Zfp42
Klf4 Tbx3
Tfcp2l1
Inhibitor addition
Inhibitor withdrawal
Fold-change
12x
6x
1x
5h 20h 50h LT 5h 20h 50h LT
eRNAs + 2i
Genes + 2i
eRNAs + CDK8/19i
Genes + CDK8/19i
Fold change
5 20 50 LT
Hours
Naïve-specific eRNAs
Naïve-specific mRNAs
2i
CDK8/19i
Pramel6 5' eRNA
Pramel6 3' eRNA
Tbx3 5' eRNA
Tbx3 3' eRNA
Gjb3 5' eRNA
Gjb3 3' eRNA
Nanog
Klf4
Tfcp2l1
Rex1/Zfp42
Tbx3
E
Super-Enhancer target genes mRNA expression
Expression-matched control genes (excluding enhancer-targets)
Enrichment score
FDR: 0.0
NES: +1.75
FDR: 1.0
NES: -0.53
RNAseq rank: CDK8/19i vs Control
F
GO terms genesets in CDK8/19i vs Control
KEGG genesets in CDK8/19i vs Control
GSEA NES score
SE-target genes
#37 of 3,844 genesets Top 1%
#2 of 150 genesets Top 1%
C5 GO terms genesets
C2 KEGG genesets
G
Control
BRD4i 48hrs
BRD4i Day7
BRD4i Day21
No surviving cells
WT iPS Serum/LIF
WT iPS 2i
CDK8-KD iPS Serum/LIF
H
Tbx3 5'eRNA
Prdm14 5'eRNA
Nanog
Klf4
Log2 fold-change
Serum/LIF
2i
+0 BRD4i
+0 BRD4i
Empty Vector
CDK8-KD
Serum/LIF or 2i conditions

Figure 5: 2i and CDK8/19i hyperactivate naïve-state enhancer activity, conferring resistance to enhancer destabilization.

(A) Plot of CDK8/19 peak intensity in the indicated genomic regions. CDK8/19 binding loci were defined in mouse PSCs by ChIPseq[2,3] and MACS peak calling was performed. CDK8/19 binding loci were then grouped by localization in enhancer constituent regions, SE: Super-Enhancer; TE: Typical-Enhancer; as defined in mouse PSCs [2,3], or functional annotation of the region by HOMER (see Tables S3 and S7, and **Methods**). Promoter-TSS: TSS+/- 1Kb. Gene Body: Exons, Introns, and transcription termination site TTS +/- 1Kb. *** P<0.001.

(B) RNA Pol II abundance measured by ChIPseq (3 pooled replicates, see **Methods**) in mouse PSC regions without CDK8/19 binding (on left, n = 423), or with the top 10% strongest CDK8/19 binding signals (on right, n = 464), as defined in (A) above. **** P<0.0001.

(C) RNA Pol II abundance, measured in mouse PSCs as in (B) above, at typical enhancers (TE; on left, n = 9981), or super-enhancers (SE; on right, n = 646), as defined [2,3]. **** P<0.0001.

(D) Left panel: pluripotency marker genes and naïve-specific enhancer RNA (eRNA) abundance measured by qRT-PCR in mouse PSC at time intervals after addition of 2i or CDK8/19i to the culture, relative to the standard serum/LIF control. Naïve-specific eRNAs and primers, as defined [100]. Mean +/-SEM, of three independent cell experiments. Right panel: heatmap of pluripotency marker genes and naïve-specific eRNA abundance, showing the fold-change in expression relative to the standard serum/LIF control at time intervals after inhibitor addition, or inhibitor withdrawal. Heatmap data are the mean values of replicate experiments determined by qRT-PCR as in the left panel (see also Figure S7G for plots during inhibitor withdrawal).

(E) Left panel: GSEA analysis of super-enhancer target gene mRNAs in mouse PSCs adapted to CDK8/19i (3 replicates). SE-targets were defined as the single nearest gene by GREAT analysis (see: **Methods** and Table S7 for SE-target gene list, n = 189). SE-target genes are significantly up-regulated (left panel; FDR q-value P~0), while, expression-matched control genes show no significant change

in expression levels (right panel; FDR q-value P = 1.0. The expression-matched control geneset (see Table S2 for list) contains genes which are not predicted to be enhancer targets (by Standard GREAT analysis; see **Methods**) but which have similar high levels of expression to the SE-target genes.

(F) The relative specificity of SE-target gene up-regulation by CDK8/19i was determined by comparison with databases of other genesets by GSEA. C5 Gene Ontology (GO) terms (n = 3,844 genesets) and C2 KEGG (n = 150 genesets) as defined by the Broad Institute. SE-target genes lie within the top 1% most-significantly up-regulated genesets realtive to these GO term or KEGG databases.

(G) Brightfield images showing typical cell morphology following treatment with 500 nM BRD4i/JQ1 for 48 h (left) or 7 days (right), in WT iPS cells or in CDK8/19-double-knockout (CDK8/19-dKO) iPS cells stably expressing pMSCV-CDK8-Kinase Dead (CDK8-KD). Panels on right show brightfield image of colonies which have been fixed and stained for alkaline phosphatase at Day 21/passge 5 of treatment with 500 nM BRD4i/JQ1. Data are representative of three cell experiments.

(H) qRT-PCR expression of naïve eRNA and marker genes following treatment with 500 nM BRD4i/JQ1 for 48 h. CDK8/19-dKO iPS +/- CDK8-KD were cultured +/- 2i or standard serum/LIF, as indicated. Mean +/- SD of three clones. See also Figure S7J.

Fig.6

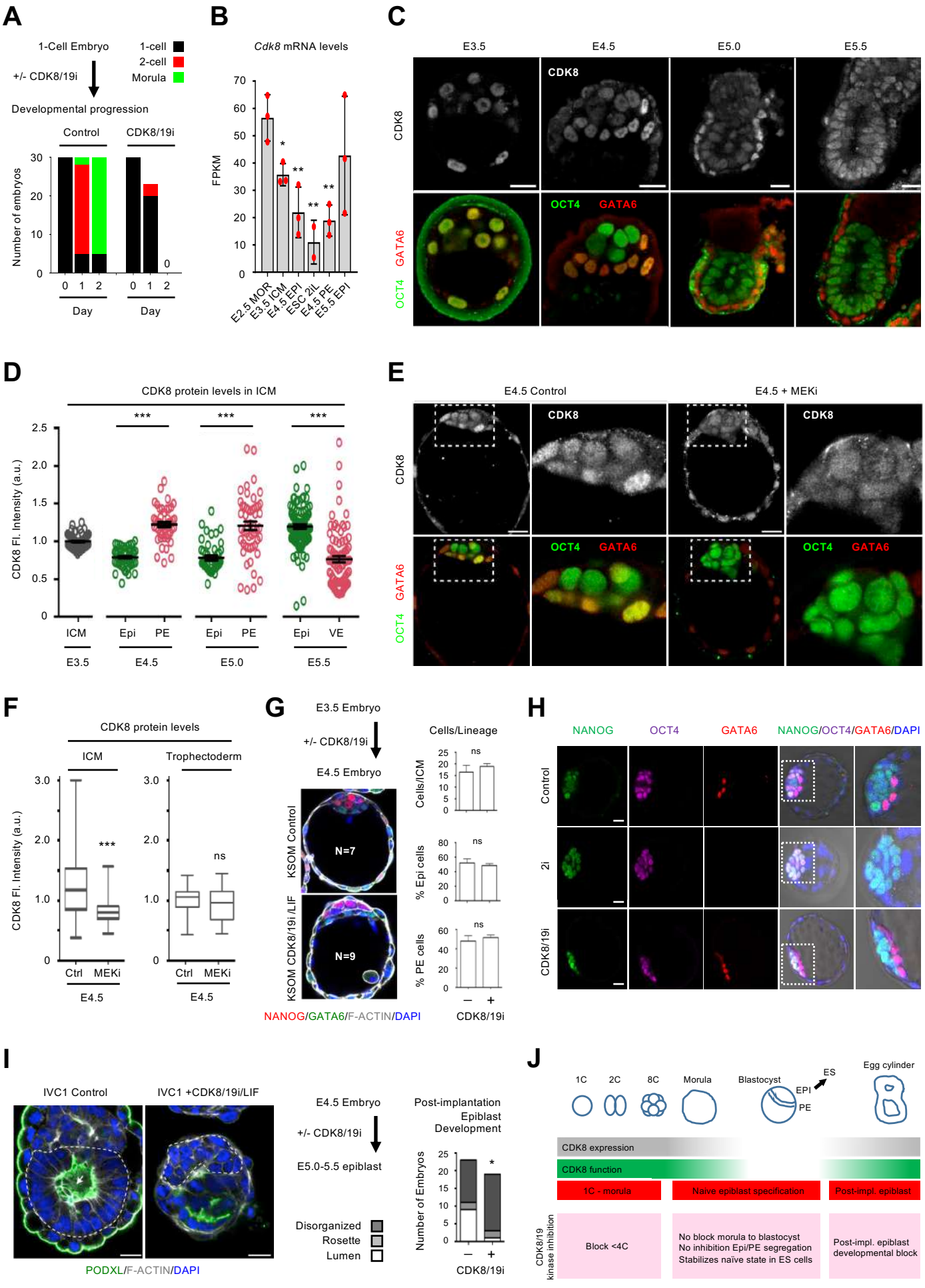

A
1-Cell Embryo
+/- CDK8/19i
Developmental progression
1-cell
2-cell
Morula
Control
CDK8/19i
Number of embryos
Day
B
Cdk8 mRNA levels
FPKM
E2.5 MOR
E3.5 ICM
E4.5 EPI
ESC 2iL
E4.5 PE
E5.5 EPI
C
E3.5
E4.5
E5.0
E5.5
CDK8
OCT4
GATA6
D
CDK8 protein levels in ICM
CDK8 Fl. Intensity (a.u.)
ICM
Epi
PE
VE
E3.5
E4.5
E5.0
E5.5
E
E4.5 Control
E4.5 + MEKi
F
CDK8 protein levels
ICM
Trophectoderm
Ctrl
MEKi
E4.5
G
E3.5 Embryo
+/- CDK8/19i
E4.5 Embryo
KSOM Control
KSOM CDK8/19i /LIF
N=7
N=9
NANOG/GATA6/F-ACTIN/DAPI
Cells/Lineage
Cells/ICM
% Epi cells
% PE cells
CDK8/19i
H
NANOG
OCT4
GATA6
NANOG/OCT4/GATA6/DAPI
Control
2i
CDK8/19i
I
IVC1 Control
IVC1 +CDK8/19i/LIF
PODXL/F-ACTIN/DAPI
E4.5 Embryo
+/- CDK8/19i
E5.0-5.5 epiblast
Disorganized
Rosette
Lumen
Post-implantation Epiblast Development
Number of Embryos
CDK8/19i
J
1C
2C
8C
Morula
Blastocyst
ES
EPI
PE
Egg cylinder
CDK8 expression
CDK8 function
1C - morula
Naive epiblast specification
Post-impl. epiblast
CDK8/19 kinase inhibition
Block <4C
No block morula to blastocyst
No inhibition Epi/PE segregation
Stabilizes naïve state in ES cells
Post-impl. epiblast developmental block

Figure 6: CDK8 expression in vivo and the role of Mediator during mouse preimplantation development

(A) 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 30 embryos per condition, across two independent experiments.

(B) CDK8 mRNA expression levels in specific embryo stages and lineages during mouse preimplantation development. Data from [78]. Significance assessed by one-way ANOVA unpaired T-test. * $P<0.05$; ** $P<0.01$.

(C, D) Immunofluorescence and quantification of CDK8 expression at the indicated timepoints during early mouse embryo development. In (C), single Z-section shown, scale bars = 20 μm. In (D), CDK8 protein levels per nucleus were quantified at each timepoint relative to internal controls. OCT4 and GATA6 are co-expressed in all cells of the inner cell mass (ICM) at E3.5, but from E4.5 to E5.5, they segregate, marking the epiblast (Epi, Oct4+) and the primitive endoderm (PE, GATA6+), which subsequently develops into visceral endoderm (VE) at E5.5. Embryo staging, CDK8 quantification and normalization: see **Methods**. Data from three independent experiments (images are representative). Significance was assessed by Kruskal-Wallis test with multiple comparisons.

(E,F) CDK8 expression is repressed by MEK inhibition in vivo. Embryos were incubated +/- MEKi PD0320591 for 48hrs, from E2.5 8-cell morula to E4.5 blastocyst stage. (E) Immunofluorescence for CDK8 protein expression in the E4.5 blastocyst. Scale bars = 20 μm. (F) CDK8 protein expression levels per cell were quantified in the ICM, or Trophectoderm, relative to internal controls, with significance assessed by T-test, *** $P<0.001$; ns = not significant. Data from three independent experiments (images are representative).

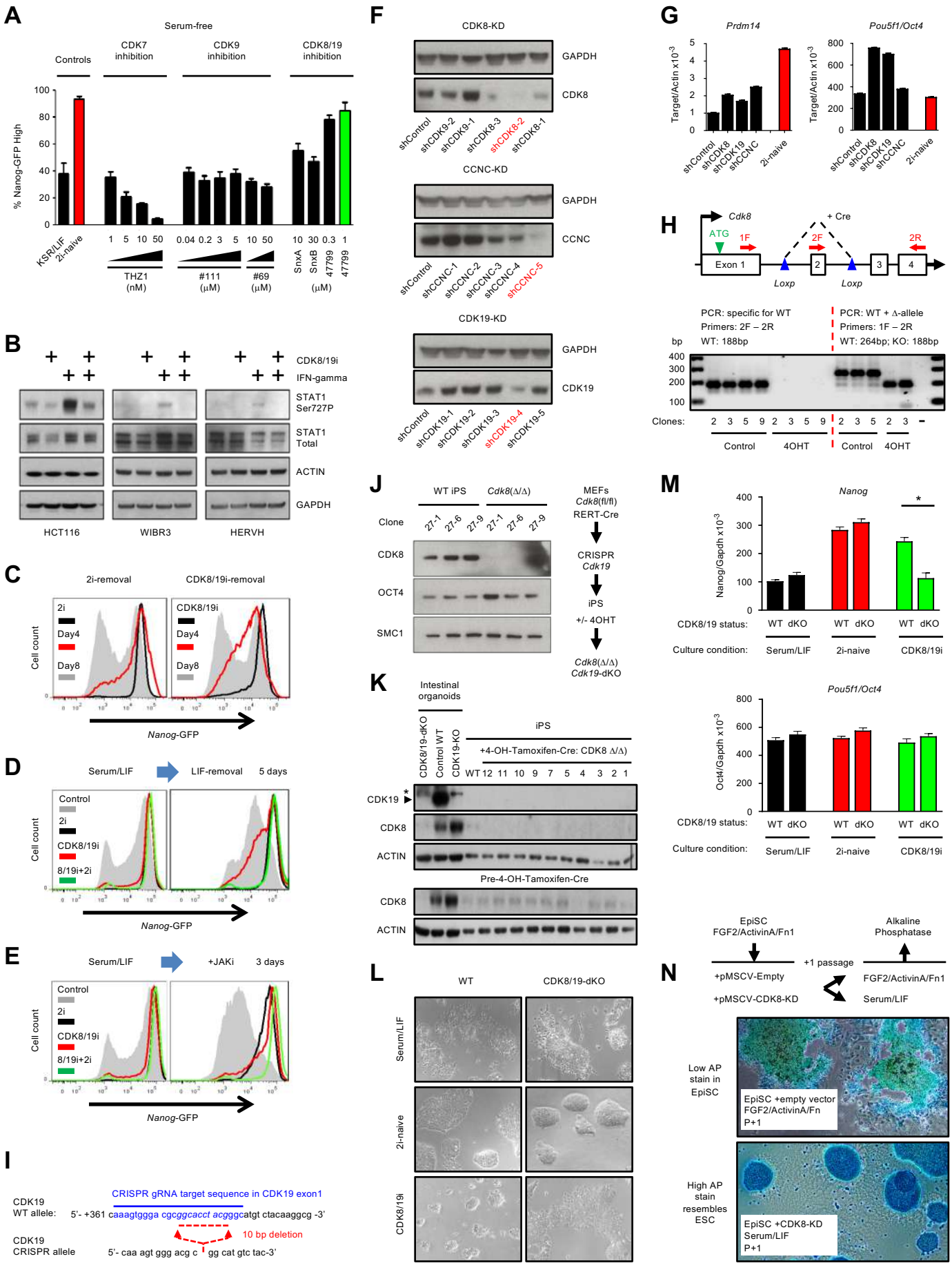
(G,H) CDK8/19-inhibition does not prevent Epi/PE segregation. E3.5 embryos were incubated +/- CDK8/19i for 24 hrs, during the emergence of Epi/PE segregation. Quantification of lineage allocation in the inner cell mass (ICM), Epi (Nanog+), PE (Gata6+) and ICM (Nanog+ or Gata6+). Significance was assessed

by T-test. Scale bar 20 μm. Data in plot are Mean +/- SEM, from two independent experiments (images are representative).

(I) CDK8/19-inhibition interrupts pre- to post-implantation morphogenic events. Pre-implantation E4.5 embryos were cultured until E5.0 in vitro as described [107,139] (see, **Methods**) +/- CDK8/19i. Embryos were stained with PODXL and F-ACTIN to determine the emergence of a luminal space (pro-amniotic cavity) within the epiblast (outlined in the images), scale bars = 20 μm. Morphogenesis was quantified on right, and assessed for significance by Chi-square test, * $P<0.05$. Data from two independent experiments (images are representative)

(J) Table summarizing the three periods of early embryo development studied in this figure. A requirement for CDK8 function appears to follow the pattern of CDK8 expression. Notably, maxima in CDK8 expression coincide with a requirement for development around the 1C-morula or post-implantation stages. Between these two periods, a minima in CDK8 expression occurs transiently during the emergence of the naïve epiblast, where CDK8 function appears dispensable.

Fig. S1


A
Serum-free
Controls
CDK7 inhibition
CDK9 inhibition
CDK8/19 inhibition
% Nanog-GFP High
KSR/LIF
2i-naive
THZ1 (nM)
#111 (μM)
#69 (μM)
SnxA
SnxB
47799 (μM)
B
CDK8/19i
IFN-gamma
STAT1 Ser727P
STAT1 Total
ACTIN
GAPDH
HCT116
WIBR3
HERVH
C
2i-removal
CDK8/19i-removal
Cell count
2i
Day4
Day8
CDK8/19i
Nanog-GFP
D
Serum/LIF
LIF-removal
5 days
Control
2i
CDK8/19i
8/19i+2i
Cell count
Nanog-GFP
E
Serum/LIF
+JAKi
3 days
Control
2i
CDK8/19i
8/19i+2i
Cell count
Nanog-GFP
I
CDK19 WT allele:
CRISPR gRNA target sequence in CDK19 exon1
5'- +361 caaagtggga cgcggcacct acgggcatgt ctacaaggcg -3'
10 bp deletion
CDK19 CRISPR allele
5'- caa agt ggg acg c gg cat gtc tac-3'
F
CDK8-KD
GAPDH
CDK8
shControl
shCDK9-2
shCDK9-1
shCDK8-3
shCDK8-2
shCDK8-1
CCNC-KD
GAPDH
CCNC
shControl
shCCNC-1
shCCNC-2
shCCNC-3
shCCNC-4
shCCNC-5
CDK19-KD
GAPDH
CDK19
shControl
shCDK19-1
shCDK19-2
shCDK19-3
shCDK19-4
shCDK19-5
G
Prdm14
Pou5f1/Oct4
Target/Actin x10-3
shControl
shCDK8
shCDK19
shCCNC
2i-naive
H
Cdk8
+ Cre
ATG
1F
2F
2R
Exon 1
Loxp
Loxp
PCR: specific for WT
Primers: 2F – 2R
WT: 188bp
PCR: WT + Δ-allele
Primers: 1F – 2R
WT: 264bp; KO: 188bp
bp
Clones:
Control
4OHT
Control
4OHT
J
WT iPS
Cdk8(Δ/Δ)
Clone
CDK8
OCT4
SMC1
MEFs Cdk8(fl/fl) RERT-Cre
CRISPR Cdk19
iPS
+/- 4OHT
Cdk8(Δ/Δ) Cdk19-dKO
K
Intestinal organoids
CDK8/19-dKO
Control WT
CDK19-KO
iPS
+4-OH-Tamoxifen-Cre: CDK8 Δ/Δ
WT 12 11 10 9 7 5 4 3 2 1
CDK19
CDK8
ACTIN
Pre-4-OH-Tamoxifen-Cre
CDK8
ACTIN
L
WT
CDK8/19-dKO
Serum/LIF
2i-naive
CDK8/19i
M
Nanog
Nanog/Gapdh x10-3
CDK8/19 status:
WT dKO
Culture condition:
Serum/LIF
2i-naive
CDK8/19i
Pou5f1/Oct4
Oct4/Gapdh x10-3
N
EpiSC FGF2/ActivinA/Fn1
+1 passage
Alkaline Phosphatase
+pMSCV-Empty
+pMSCV-CDK8-KD
FGF2/ActivinA/Fn1
Serum/LIF
Low AP stain in EpiSC
EpiSC +empty vector FGF2/ActivinA/Fn P+1
High AP stain resembles ESC
EpiSC +CDK8-KD Serum/LIF P+1

Supplemental Figure S1.

An Inhibitor Screen for Promotion of ES Naïve State identifies a distinct role for Mediator kinase activity.

(A) Effect of indicated treatments on Nanog expression per cell, using a mouse *Nanog*-GFP knockin reporter ES cell line in serum-free KSR/LIF (knockout-serum replacement) base media. Percentage of Nanog-GFP-high determined by FACS. Data representative of 3 independent experiments.

(B) Phosphorylation levels of the CDK8-target STAT1 Ser727 by Western blot in the indicated three cell lines. Cells were treated with 1,1 μM CDK8/19i for 3 hours, +/- simultaneous induction of STAT1-Ser727P by Interferon-γ for 3 hours.

(C) Effect of inhibitor-removal from mouse PSC previously adapted to 2i or CDK8/19i. FACS histogram with *Nanog*-GFP knockin reporter ES cells showing changes to *Nanog*-GFP at intervals following inhibitor removal. A decrease in the proportion of *Nanog*-GFP^{High} cells indicates loss of the naïve state over several days, in a similar manner for 2i-removal or CDK8/19i-removal. Data representative of 2 independent experiments.

(D) Effect of LIF-removal from mouse PSC previously adapted to 2i or CDK8/19i. FACS histogram with *Nanog*-GFP knockin reporter ES cells showing changes to *Nanog*-GFP at intervals following LIF removal. A decrease in the proportion of *Nanog*-GFP^{High} cells indicates loss of the naïve state in some cells, while others appear protected, over several days following LIF removal in CDK8/19i. 2i protects *Nanog*-GFP^{High} cells to a better extent than CDK8/19i following LIF removal. Data representative of 2 independent experiments.

(E) Effect of inhibition of LIF-STAT3 signaling with a JAK-STAT inhibitor (JAKi) in mouse PSC previously adapted to 2i or CDK8/19i. FACS histogram with *Nanog*-GFP knockin reporter ES cells showing changes to *Nanog*-GFP at intervals after addition of JAKi. A decrease in the proportion of *Nanog*-GFP^{High} cells indicates loss of the naïve state in some cells, while others appear protected, over several days following LIF removal in CDK8/19i or 2i. Only the combination of CDK8/19i plus 2i protects *Nanog*-GFP^{High} cells completely following JAK-STAT inhibition. Data representative of 2 independent experiments.

(F) Western blots indicating lentiviral shRNA-mediated knockdown of CDK8, CDK19 or Cyclin C in mouse PSC. The most efficient shRNAs (red) were selected for use in subsequent experiments. Data representative of 2 independent experiments.

(G) Pluripotency marker mRNA expression measured by qRT-PCR in Mouse ES cells following 7 days of lentiviral shRNA-mediated knockdown of CDK8, CDK19, or CyclinC (CCNC), as indicated. Mean +/-SD of 3 replicates.

(H) Upper: schematic of inducible CDK8 knockout. 4-hydroxy-tamoxifen (4OHT)-inducible Cre drives excision of Exon2. Lower: PCR confirmation of CDK8 Exon2 deletion using the indicated primers. Mouse Cdk8(fl/fl) RERT-Cre iPS cells were treated with 0,5 µM 4OHT for 6 days.

(I) Schematic example of indel mutation in one mouse CDK19-KO clone. Indel was induced by the indicated CRISPR guide RNA against CDK19 Exon1, using a lentiviral CRISPR-Cas9 system (see: **Methods**). This indel consists of 10bp deletion at the predicted CRIPSR target site, generating a frameshift immediately downstream of the ATG start codon of CDK19.

(J) Western blots of 4-hydroxy-tamoxifen (4-OHT) inducible CDK8-knockout mouse iPS cells as in (H) above. Schematic summarizes the generation of these cells (see: **Methods**).

(K) Western blots of 4-hydroxy-tamoxifen (4-OHT) inducible CDK8/19-double-knockout mouse iPS clones, generated as in (J) above. Loss of CDK8 is confirmed at the protein level after 4-hydroxy-tamoxifen-inducible treatment (4-OHT-Cre). CDK19 was undetectable at the protein level in PSCs, but readily detectable in the control intestinal organoid samples. Arrow indicates CDK19, confirmed by CRISPR-knockout of CDK19 as indicated, while asterisk indicates a non-specific band.

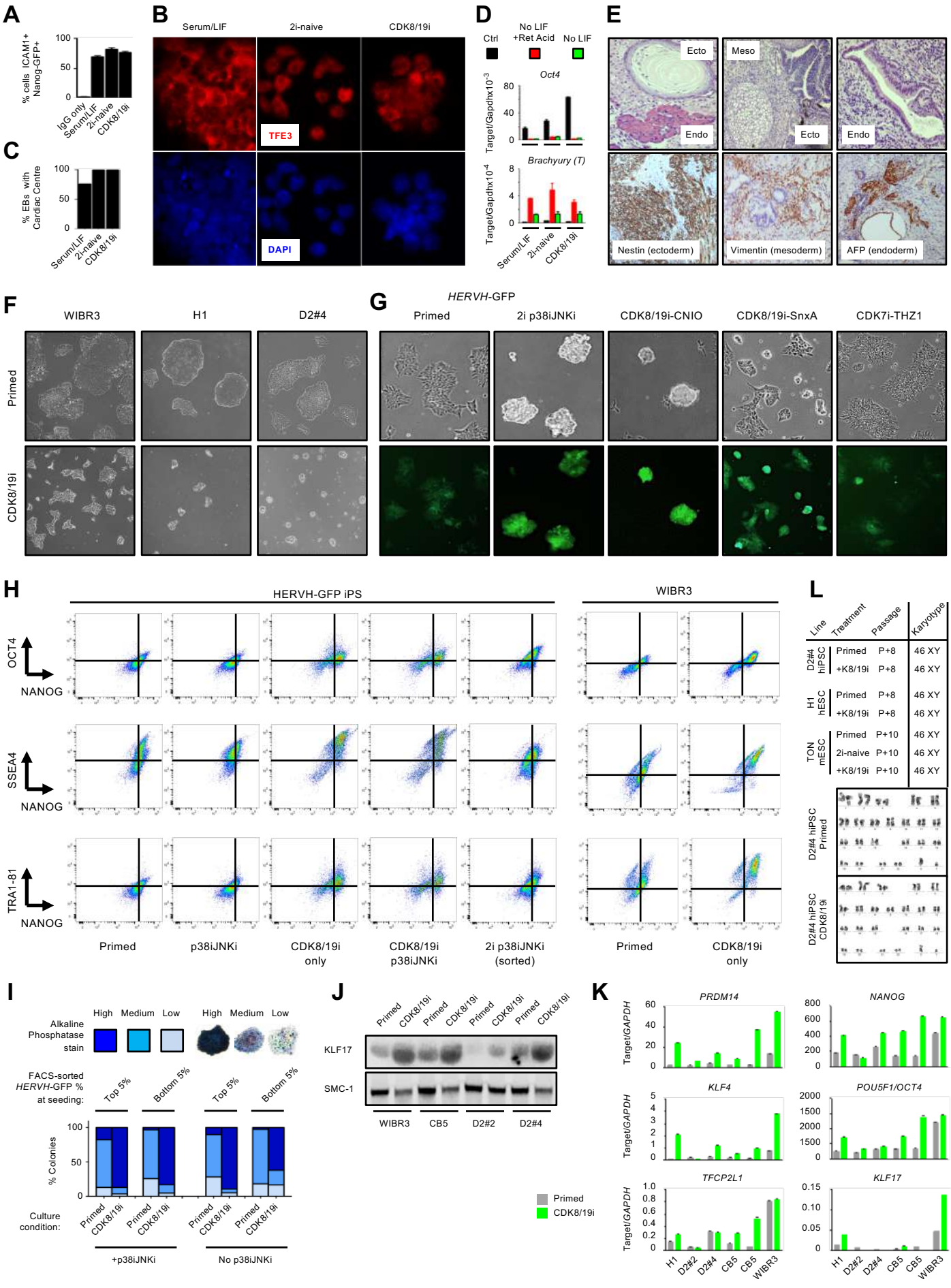
(L) Brightfield images of mouse iPS lines which express CDK8 and CDK19 (CDK8 fl/fl; CDK19 +/-, named in the figure as WT) versus CDK8/19 double-knockout (dKO). CDK8/19-dKO mouse iPS cells proliferate in definitely (at least 25 passages) in 2i-naïve or serum/LIF conditions. Importantly, CDK8/19-dKO iPS no

longer respond to CDK8/19-inhibitors. Images are representative of 3 independent clones.

(M) Expression of *Nanog* and *Pou5f1/Oct4* mRNA levels by qRT-PCR in mouse iPS clones (n=3) which express CDK8 and CDK19 (CDK8 fl/fl; CDK19+/-, abbreviated in the figure as WT) versus iPS clones (n=6) which are CDK8/19-dKO (CDK8 Δ/Δ; CDK19-/-). Mean +/-SD of replicates. Cells were cultured in the indicated condition for 7 days. In the iPS cells expressing CDK8/19, both 2i and CDK8/19i treatments show the expected increase in *Nanog* mRNA levels. However, the CDK8/19-dKO iPS cells no longer increase *Nanog* mRNA levels upon exposure to the CDK8/19 inhibitor (* P-value < 0.0001). This suggests that increased *Nanog* expression is induced only by kinase-inhibition of CDK8/19, and not when the two proteins are physically absent, consistent with the known kinase-independent roles for CDK8/19 within the Mediator complex (see Introduction).

(N) Related to main Figure 1N and 1O. Cell morphology and alkaline phosphatase staining intensity of WT mouse EpiSC infected with pMSCV-Empty or pMSCV-CDK8-KD, which were then maintained for 1 passage in EpiSC media (Fgf2/ActivinA/fibronectin; see **Methods**) or 1 passage in standard serum/LIF ES media (see schematic, upper panel, and also Figures 1N and 1O). Images are representative of day 7 after pMSCV infection and selection, in two cell experiments.

Fig. S2


A
% cells ICAM1+ Nanog-GFP+
IgG only
Serum/LIF
2i-naive
CDK8/19i
B
Serum/LIF
2i-naive
CDK8/19i
TFE3
DAPI
C
% EBs with Cardiac Centre
D
Ctrl
No LIF +Ret Acid
No LIF
Oct4
Brachyury (T)
Target/Gapdhx10-3
Target/Gapdhx10-4
E
Ecto
Meso
Endo
Nestin (ectoderm)
Vimentin (mesoderm)
AFP (endoderm)
F
WIBR3
H1
D2#4
Primed
CDK8/19i
G
HERVH-GFP
Primed
2i p38iJNKi
CDK8/19i-CNIO
CDK8/19i-SnxA
CDK7i-THZ1
H
HERVH-GFP iPS
WIBR3
OCT4
SSEA4
TRA1-81
NANOG
Primed
p38iJNKi
CDK8/19i only
CDK8/19i p38iJNKi
2i p38iJNKi (sorted)
L
Line
Treatment
Passage
Karyotype
D2#4 hiPSC
Primed P+8 46 XY
+K8/19i P+8 46 XY
H1 hESC
Primed P+8 46 XY
+K8/19i P+8 46 XY
TON mESC
Primed P+10 46 XY
2i-naive P+10 46 XY
+K8/19i P+10 46 XY
D2#4 hiPSC Primed
D2#4 hiPSC CDK8/19i
I
High Medium Low
Alkaline Phosphatase stain
FACS-sorted HERVH-GFP % at seeding:
Top 5%
Bottom 5%
% Colonies
Culture condition:
+p38iJNKi
No p38iJNKi
J
KLF17
SMC-1
WIBR3
CB5
D2#2
D2#4
K
PRDM14
NANOG
KLF4
POU5F1/OCT4
TFCP2L1
KLF17
Target/GAPDH
H1
D2#2
D2#4
CB5
CB5
WIBR3

Supplemental Figure S2.

Positive effect of long-term CDK8/19i on mammalian ES self-renewal and pluripotency

(A) FACS analysis to confirm retention of pluripotency markers in mouse PSCs. The percentage of double-positive Nanog-GFP+/ICAM1+ PSCs is shown following long-term adaption (3 weeks) to control (serum/LIF), 2i-naïve, or CDK8/19i culture. Data representative of 2 cell experiements.

(B) Immunofluorescence for TFE3 expression and localization in mouse PSCs adapted to control (serum/LIF), 2i-naïve, or CDK8/19i conditions, as in (A) above.

(C) Embryoid body differentiation assay in vitro to confirm developmental capacity of mouse PSCs previously adapted to control (serum/LIF), 2i-naïve, or CDK8/19i conditions. EB differentiation (see **Methods**) was assessed visually by observation of beating cardiac centres.

(D) Differentiation assays in vitro to confirm 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 (see **Methods**). Differentiation was assessed by qRT-PCR to show loss of pluripotency (*Oct4*) and induction of differentiation (*Brachyury, T*). Related to main Figure 2C. Data representative of 2 independent experiments, showing Mean +/- SD of 3 technical replicates.

(E) Teratoma developmental assay in vivo to determine developmental capacity of mouse PSCs previously adapted to CDK8/19i conditions. The development of three germ layers was confirmed in teratomas using histology by H+E staining (upper panels), and by immunohistochemical staining for germ layer markers: NESTIN (ectoderm), VIMENTIN (mesoderm), and Alpha-feto-protein (AFP, endoderm).

(F) Brightfield images showing colony morphology in 3 human PSC lines in primed state (upper panels), or following 14 days treatment with CDK8/19i. Images representative of at least 5 independent cell experiments.

(G) Brightfield and live-cell GFP-fluorescence images of human iPS cells (HERVH-GFP reporter) in primed conditions, or following 14 days treatment with the indicated media cocktails including: CDK7i, CDK8/19i, or 2i p38iJNKi. Cytometric sorting to select at each passage for the cells with the top 10% of HERVH-GFP was required to derive and maintain the 2i p38iJNKi condition. Images representative of at least 5 independent cell experiments.

(H) FACS analysis of pluripotency markers in human PSCs (HERVH iPS or WIBR3 ES), following 3 weeks adaption to the indicated culture conditions, as in (G) above. Cells in Primed or CDK8/19i conditions were routinely passaged in bulk using collagenase. However in contrast, for the 2i p38iJNKi condition, at each passage, cytometric-sorting was required to select for the cells with highest 10% of HERVH-GFP, and this was performed for 3 passages, before fixing the cells 4 days after the third passage/selection-round.

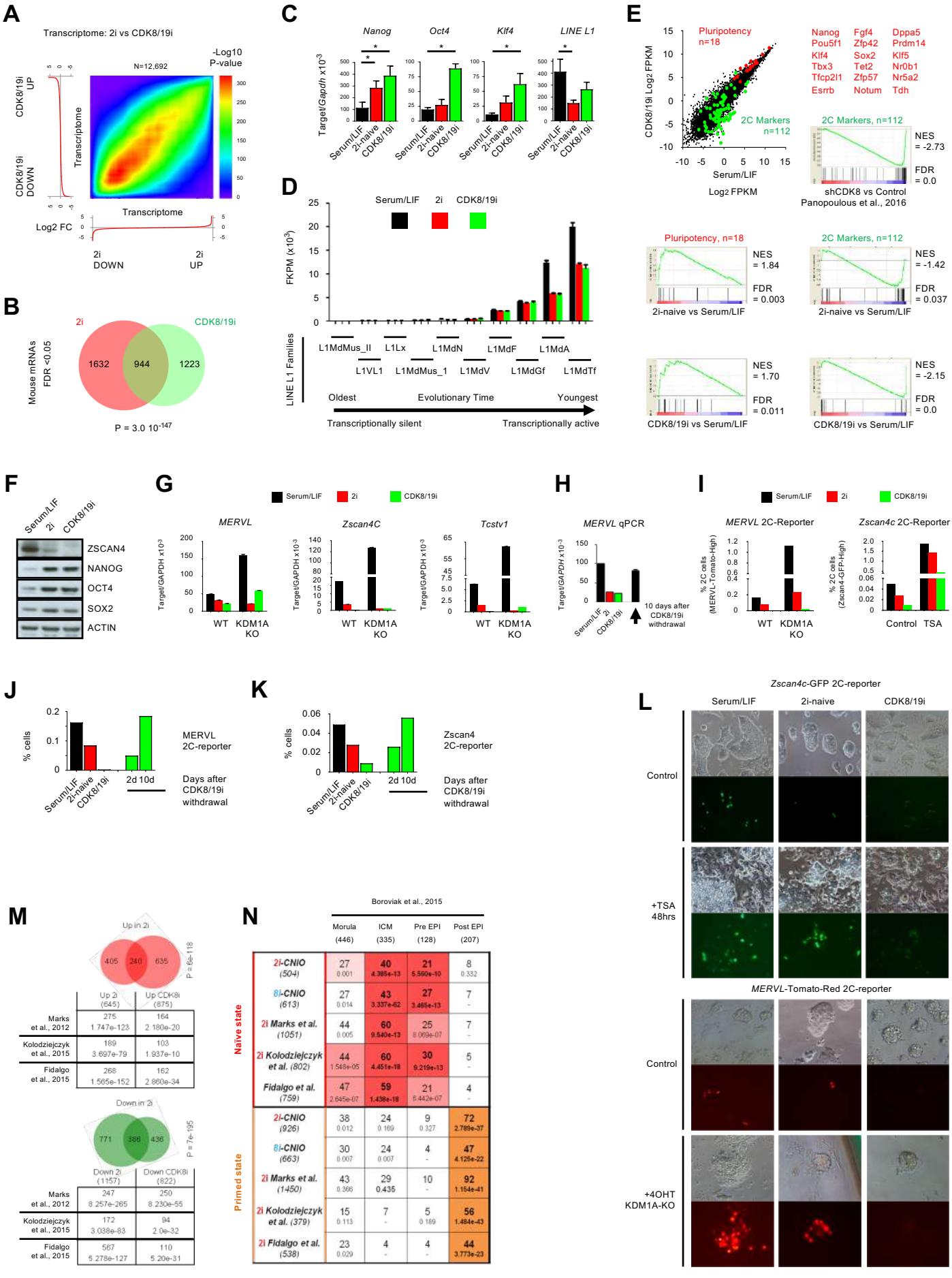
(I) Assay of clonogenicity of human PSCs cultured in primed or CDK8/19i conditions as in (G) above. Cells were FACS sorted according to their *HERVH*-GFP intensity to collect the top or bottom 5% GFP intensity cells (indicated above graph). These top or bottom 5% were then seeded at clonal density in the control primed, or CDK8/19i, media (indicated below graph) for seven days, to allow individual colonies to arise separately. At day 7, cells were fixed and stained for Alkaline phosphatase as an indicator of their pluripotent status. The alkaline phosphatase staining intensity was recorded visually for each colony in ten fields of view (a guide of colony staining score is shown on left of graph). Graph shows the overall ability of the seeded cells to retain or enhance pluripotency (medium or homogenously high staining, respectively) in either primed or CDK8/19i culture conditions. Also,the effect of CDK8/19i was independent of the inclusion of p38i and JNKi inhibitors.

(J) Western blots for protein expression levels of the naïve pluripotency marker KLF17 in four human PSC lines, adapted to primed or CDK8/19i conditions (see **Methods**). SMC1 is a nuclear protein as internal loading control.

(K) qRT-PCR for mRNA expression levels of pluripotency markers in five human PSC lines, adapted to primed or CDK8/19i conditions for at least 14 days. Mean +/- SD of three replicates. See also Table S1 for all differentiation markers tested.

(L) Karyotyping to determine genomic stbility of PSC lines adapted for at least 8 passages in control primed, or CDK8/19i, media conditions.

Fig. S3


A
Transcriptome: 2i vs CDK8/19i
-Log10 P-value
N=12,692
CDK8/19i UP
CDK8/19i DOWN
Transcriptome
Log2 FC
2i DOWN
2i UP
B
2i
CDK8/19i
Mouse mRNAs FDR <0.05
1632
944
1223
P = 3.0 10-147
C
Nanog
Oct4
Klf4
LINE L1
Target/Gapdh x10-3
Serum/LIF
2i-naive
CDK8/19i
D
Serum/LIF
2i
CDK8/19i
FKPM (x10-3)
LINE L1 Families
L1MdMus_II
L1Lx
L1MdN
L1MdF
L1MdA
L1VL1
L1MdMus_1
L1MdV
L1MdGf
L1MdTf
Oldest
Evolutionary Time
Youngest
Transcriptionally silent
Transcriptionally active
E
Pluripotency n=18
2C Markers n=112
CDK8/19i Log2 FPKM
Serum/LIF
Log2 FPKM
Nanog
Pou5f1
Klf4
Tbx3
Tfcp2l1
Esrrb
Fgf4
Zfp42
Sox2
Tet2
Zfp57
Notum
Dppa5
Prdm14
Klf5
Nr0b1
Nr5a2
Tdh
2C Markers, n=112
NES = -2.73
FDR = 0.0
shCDK8 vs Control
Panopoulous et al., 2016
Pluripotency, n=18
NES = 1.84
FDR = 0.003
2i-naive vs Serum/LIF
2C Markers, n=112
NES = -1.42
FDR = 0.037
2i-naive vs Serum/LIF
NES = 1.70
FDR = 0.011
CDK8/19i vs Serum/LIF
NES = -2.15
FDR = 0.0
CDK8/19i vs Serum/LIF
F
Serum/LIF
2i
CDK8/19i
ZSCAN4
NANOG
OCT4
SOX2
ACTIN
G
Serum/LIF
2i
CDK8/19i
MERVL
Zscan4C
Tcstv1
Target/GAPDH x10-3
WT
KDM1A KO
H
MERVL qPCR
Target/GAPDH x10-3
Serum/LIF
2i
CDK8/19i
10 days after CDK8/19i withdrawal
I
Serum/LIF
2i
CDK8/19i
MERVL 2C-Reporter
Zscan4c 2C-Reporter
% 2C cells (MERVL-Tomato-High)
% 2C cells (Zscan4-GFP-High)
WT
KDM1A KO
Control
TSA
J
% cells
Serum/LIF
2i-naive
CDK8/19i
2d
10d
MERVL 2C-reporter
Days after CDK8/19i withdrawal
K
% cells
Serum/LIF
2i-naive
CDK8/19i
2d
10d
Zscan4 2C-reporter
Days after CDK8/19i withdrawal
L
Zscan4c-GFP 2C-reporter
Serum/LIF
2i-naive
CDK8/19i
Control
+TSA 48hrs
MERVL-Tomato-Red 2C-reporter
Control
+4OHT KDM1A-KO
M
Up in 2i
405
240
635
P = 6e-118
Up 2i (645)
Up CDK8i (875)
Marks et al., 2012
275
1.747e-123
164
2.180e-20
Kolodziejczyk et al., 2015
189
3.697e-79
103
1.937e-10
Fidalgo et al., 2015
268
1.565e-152
162
2.860e-34
Down in 2i
771
386
436
P = 7e-195
Down 2i (1157)
Down CDK8i (822)
Marks et al., 2012
247
8.257e-265
250
8.230e-55
Kolodziejczyk et al., 2015
172
3.038e-83
94
2.0e-32
Fidalgo et al., 2015
567
5.278e-127
110
5.20e-31
N
Boroviak et al., 2015
Morula (446)
ICM (335)
Pre EPI (128)
Post EPI (207)
Naïve state
2i-CNIO (504)
27 0.001
40 4.385e-13
21 5.560e-10
8 0.332
8i-CNIO (613)
27 0.014
43 3.337e-62
27 3.465e-13
7 -
2i Marks et al. (1051)
44 0.005
60 9.540e-13
25 8.069e-07
7 -
2i Kolodziejczyk et al. (802)
44 1.548e-05
60 4.451e-18
30 9.219e-13
5 -
Fidalgo et al. (759)
47 2.645e-07
59 1.438e-18
21 5.442e-07
4 -
Primed state
2i-CNIO (926)
38 0.012
24 0.169
9 0.327
72 2.789e-37
8i-CNIO (663)
30 0.007
24 0.007
4 -
47 4.125e-22
2i Marks et al. (1450)
43 0.368
29 0.435
10 -
92 1.154e-41
2i Kolodziejczyk et al. (379)
15 0.113
7 -
5 0.189
56 1.484e-43
2i Fidalgo et al. (538)
23 0.029
4 -
4 -
44 3.773e-23

Supplemental Figure S3.

Gene expression analysis in mouse and human PSCs adapted to 2i or CDK8/19i

(A) Rank-Rank Hypergeometric Overlap (RRHO)[140] analysis of mRNA expression changes in mouse PSCs adapted to 2i-naïve or CDK8/19i conditions versus serum/LIF (3 replicates; N = 12,629 genes). RNAseq-detected mRNA expression changes are arranged according to their magnitude in 2i-naïve versus control serum/LIF (X-axis), and then assessed for overlap by RRHO compared to the same set of mRNAs ranked by change in their abundance in CDK8/19i versus control serum/LIF (Y-axis) (see also[140], and **Methods**). Colour intensity indicates the -log10 p-value after Benjamini-Yekutieli correction of the hypergeometric overlap.

(B) Overlap of differentially expressed mRNAs in mouse PSCs adapted to 2i or CDK8/19i, versus control KSR/LIF, in serum-free conditions (RNAseq; 3 replicates; FDR<0.05).

(C) Expression levels of pluripotency markers, or LINE L1 repeat element family, determined by qRT-PCR in mouse PSCs cultured as in (A) above. Mean +/- Std Dev, * P < 0,05.

(D) Effect of 2i or CDK8/19i on the LINE L1 super-family expression levels determined by RNAseq in mouse PSCs cultured as in (A) above. LINE L1 families are arranged according to their evolutionary age, which also reflects their transcriptional activity, as adapted from previously described studies [141–144]. LINE L1 elements of different evolutionary age are regulated by different mechanisms [141,142]. These data indicate that 2i and CDK8/19i exert highly similar effects, selectively regulating RNA expression of the youngest and most transcriptionally active families. This is consistent with previous reports of different mechanisms repressing LINE L1 families of different evolutionary ages, and moreover, the data suggest that LINEL1 regulation is highly similar in 2i-naïve and CDK8/19i treatments.

(E) Dot plot of RNAseq FPKM values in mouse PSCs adapted to control serum/LIF versus CDK8/19i, cultured as in (A) above. Critical regulators of pluripotency are

indicated in red, and listed on the right (n=18). Markers of the 2C state are indicated in green (n=112; see: Table S2)[74,145]. Results of GSEA analysis of these two genesets is shown in panels below, versus the effect of 2i or CDK8/19i in RNAseq gene expression data from the current study, or, versus the effect of shRNA knockdown of CDK8 on the transcriptome of ES cells from a recently published report[146].

(F) Western blots of protein levels in mouse PSCs adapted to the three indicated culture conditions, cultured as in (A) above. Markers of pluripotency (NANOG, OCT4, SOX2) and the 2C-fluctuation (ZSCAN4) are shown.

(G) RNA expression levels of three markers of the 2C-fluctuation determined by qRT-PCR, in mouse PSCs adapted to 2i-naïve, CDK8/19i, or control serum/LIF, cultured as in (A) above. Mean +/- SD of 3 techincal replicates.

(H) RNA expression levels of MERVL, determined by qRT-PCR, in mouse PSCs adapted to the indicated conditions, or following withdrawal of CDK8/19i. Mean +/- SD of 3 techincal replicates.

(I) Two ES reporter lines for the 2-Cell (2C)-state fluctuation. Plots show cytometric quantification of high fluorescence cells, specifically marking the population subset in the 2C-state in the three indicated culture conditions. Treatments previously reported to increase the proportion of cells in the 2C state are included: *Kdm1a*-KO (inducible knockout), or treatment with TSA for 48hrs[74].

(J,K) Two ES reporter lines for the 2C-state fluctuation. Plots show cytometric analysis of high fluorescence cells specifically marking the active 2C-state fluctuation in mouse PSCs adapted to the three indicated culture conditions. Both 2i and CDK8/19i repress the 2C state. However this is reversible within 48hrs upon removal of the inhibitor. (J) MERVL-Tomato-red reporter; (K) Zscan4c-eGFP reporter. Data is representative of 3 independent cell experiments.

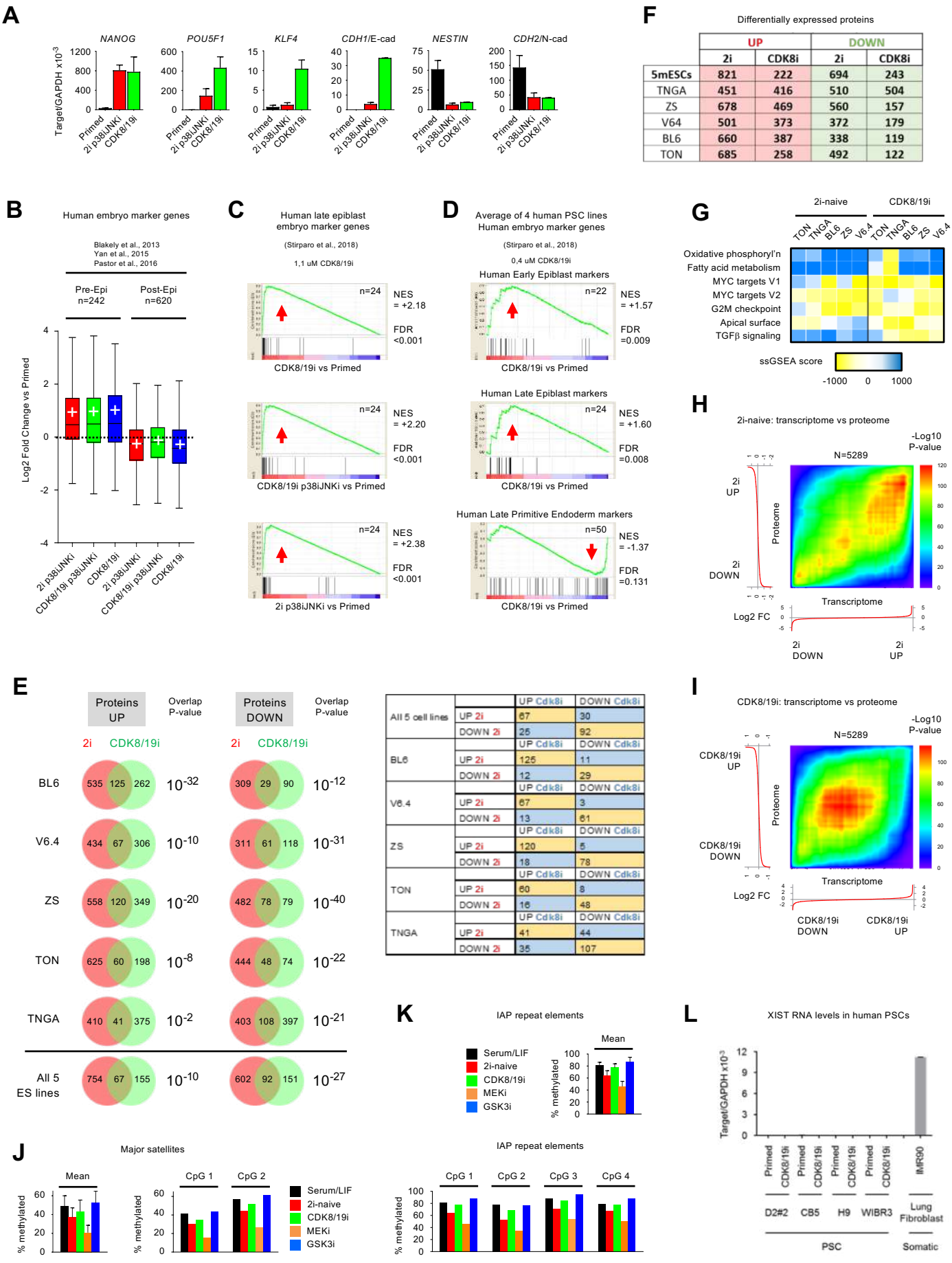
(L) Two ES reporter lines for the 2C-state fluctuation. Brightfield and fluorescence images of mouse Zscan4c-eGFP reporter (upper panel) or MERVL-Tomato-red (lower panel) ES reporter cell lines after 10 days of treatment with 2i or CDK8/19i. The effect of treatment with TSA (48 hours) is also shown for the Zscan4c-eGFP reporter cells. The effect of inducible Kdm1a-KO (7 days) is also shown for the

MERVL-Tomato-red ES reporter cells. TSA and *Kdm1a*-KO are both reported to increase the percentage of ES cells in the 2C-state [74].

(M) A comparison of the overlap between RNAseq changes in the current study versus published datasets in mouse PSCs in culture. Significantly differentially expressed mRNAs were identified in mouse PSCs adapted to 2i or CDK8/19i versus control serum/LIF in the current study, with a threshold of FDR<0.01 and a 2-fold change. These are compared to the three published studies indicated.

(N) A comparison of the overlap between RNAseq changes in the current study together with three published datasets in mouse PSCs in culture, versus developmental stage-specific marker genesets determined in vivo in preimplantation mouse embryos[78].

Fig. S4


A
NANOG
POU5F1
KLF4
CDH1/E-cad
NESTIN
CDH2/N-cad
Target/GAPDH x10-3
Primed
2i p38iJNKi
CDK8/19i
F
Differentially expressed proteins
UP
DOWN
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
B
Human embryo marker genes
Blakely et al., 2013
Yan et al., 2015
Pastor et al., 2016
Pre-Epi n=242
Post-Epi n=620
Log2 Fold Change vs Primed
2i p38iJNKi
CDK8/19i p38iJNKi
CDK8/19i
C
Human late epiblast embryo marker genes
(Stirparo et al., 2018)
1,1 uM CDK8/19i
n=24
NES = +2.18
FDR <0.001
CDK8/19i vs Primed
NES = +2.20
CDK8/19i p38iJNKi vs Primed
NES = +2.38
2i p38iJNKi vs Primed
D
Average of 4 human PSC lines
Human embryo marker genes
(Stirparo et al., 2018)
0,4 uM CDK8/19i
Human Early Epiblast markers
n=22
NES = +1.57
FDR =0.009
Human Late Epiblast markers
NES = +1.60
FDR =0.008
Human Late Primitive Endoderm markers
n=50
NES = -1.37
FDR =0.131
G
2i-naive
CDK8/19i
TON TNGA BL6 ZS V6.4
Oxidative phosphoryl'n
Fatty acid metabolism
MYC targets V1
MYC targets V2
G2M checkpoint
Apical surface
TGFβ signaling
ssGSEA score
-1000 0 1000
H
2i-naive: transcriptome vs proteome
-Log10 P-value
N=5289
2i UP
2i DOWN
Proteome
Transcriptome
Log2 FC
E
Proteins UP
Overlap P-value
Proteins DOWN
2i CDK8/19i
BL6 535 125 262 10-32 309 29 90 10-12
V6.4 434 67 306 10-10 311 61 118 10-31
ZS 558 120 349 10-20 482 78 79 10-40
TON 625 60 198 10-8 444 48 74 10-22
TNGA 410 41 375 10-2 403 108 397 10-21
All 5 ES lines 754 67 155 10-10 602 92 151 10-27
All 5 cell lines
UP 2i 67 30
DOWN 2i 25 92
BL6
UP 2i 125 11
DOWN 2i 12 29
V6.4
UP 2i 67 3
DOWN 2i 13 61
ZS
UP 2i 120 5
DOWN 2i 18 78
TON
UP 2i 60 8
DOWN 2i 16 48
TNGA
UP 2i 41 44
DOWN 2i 35 107
I
CDK8/19i: transcriptome vs proteome
-Log10 P-value
N=5289
CDK8/19i UP
CDK8/19i DOWN
Proteome
Transcriptome
Log2 FC
K
IAP repeat elements
Serum/LIF
2i-naive
CDK8/19i
MEKi
GSK3i
% methylated
Mean
CpG 1
CpG 2
CpG 3
CpG 4
L
XIST RNA levels in human PSCs
Target/GAPDH x10-3
Primed
CDK8/19i
IMR90
D2#2
CB5
H9
WIBR3
Lung Fibroblast
PSC
Somatic
J
Major satellites

Supplemental Figure S4.

Gene expression and DNA methylation analysis in PSCs adapted to 2i or CDK8/19i

(A) Expression levels of markers of pluripotency (*NANOG*, *POU5F1/OCT4*, *KLF4*, *CDH1*/E-cadherin) or differentiation (*CDH2*, *NESTIN*) determined by qRT-PCR, in human PSCs adapted to the indicated conditions. Mean +/- SD of 3 replicates.

(B) Specific marker genesets of the pre-implantation naïve epiblast (n = 242 genes) and post-implantation primed epiblast (n = 620 genes) in human embryo early development were determined by embryo single-cell RNAseq, and described in three previous studies[59,79,81]. These genesets (listed in Table S3) were also previously used to distinguish human PSCs in vitro between naïve and primed pluripotent states[59]. Here, in the current study, the RNAseq expression data for these genesets are shown in human PSCs adapted to the indicated conditions. Tukey box plots are shown, where the box reflects the 25th -75th percentile, the horizontal line is the median value, and the white cross marks the mean value. In the current study, the indicated treatments significantly up-regulate the pre-implantation naïve epiblast markers, while they significantly decrease the post-implantation primed state epiblast markers.

(C) A published specific marker geneset of the human late pre-implantation naive epiblast (n = 24 genes) was recently identified in a meta-analysis of published data[84]. Here, GSEA was used to compare the expression of this geneset in our RNAseq of human PSC cultured in 2i-naïve or CDK8/19i conditions. The GSEA enrichment plots are shown for this comparison, indicating highly significant up-regulation of these markers (red arrow) in PSCs treated with 2i-naïve (2i p38iJNKi plus repeated selection at each passage), or CDK8/19i (1,1 μM) conditions.

(D) Similar to (C) above, the behaviour of published specific marker genesets of human embryo stages were assessed in RNAseq of four human PSC lines treated in this study adapted to primed or CDK8/19i conditions. The genesets[84] for early (n = 22 genes) or late (n = 24 genes) pre-implantation naive epiblast, or late primitive endoderm (n = 50 genes) were tested here by GSEA, in order to assess the expression of each geneset in the RNAseq of four human PSC lines adapted

to primed versus CDK8/19i (0,4 μM) conditions. The GSEA enrichment plots are shown for these comparisons, indicating highly significant up-regulation of the markers in PSCs for early and late pre-implantation naive epiblast, while late primitive endoderm markers are significantly down-regulated (red arrows indicute up- or down-regulation).

(E) Overlap and hypergeometric significance test (P-value) of differentially expressed proteins in each of five mouse PSC lines adapted to 2i-naïve or in CDK8/19i, versus standard serum/LIF mouse PSCs (FDR<0.05). Below the Venn diagrams, the table compares the overlap in proteins up- or down-regulated. The values represented in yellow in the table correspond to the positive overlap between conditions, that is, the proteins up-regulated in one treatment that are also up-regulated in the other treatment. In contrast, in blue, there are the values from the negative overlap, that is, for example, the proteins up-regulated in 2i that are regulated in the opposite direction in Cdk8/19i. These data highlight the simularity in protein expression changes in 2i or CDK8/19i. See also Table S4 for full list of differentially expressed proteins.

(F) Table of proteomic changes per mouse PSC line adapted to 2i-naïve or CDK8/19i conditions, versus control serum/LIF, as cultured in (E) above.

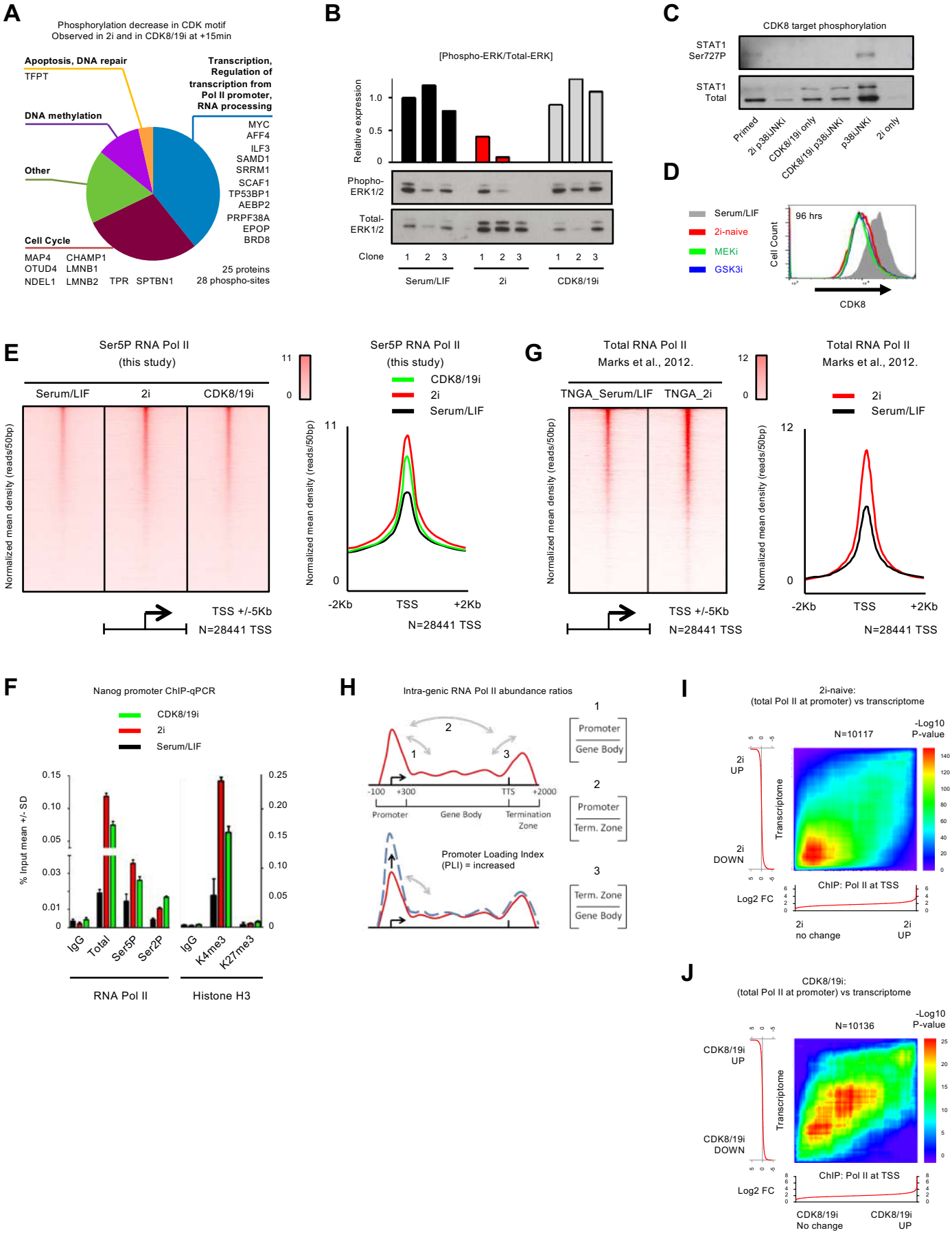
(G) Heatmap of biological pathways identified as significantly up-regulated (blue), or down-regulated (yellow), by Gene Set Enrichment Analysis (GSEA) of the proteome of mouse PSCs in 2i-naïve or CDK8/19i conditions, compared to control serum/LIF, as cultured in (E) above. PSC cell lines indicated above heatmap.

(H,I) Rank-Rank Hypergeometric Overlap (RRHO)[140] analysis of mRNA expression changes in mouse PSCs adapted to 2i-naïve conditions (H) or CDK8/19i (I), versus, change in abundance of the same set of proteins (n = 5289). Genes defined as up or down in inhibitor-treated cells were determined versus control serum/LIF conditions. RNAseq-detected mRNA expression changes are arranged according to their magnitude (X-axis), and then assessed for overlap by RRHO compared to the same set of proteins also ranked by change in their abundance detected by Mass spectrometry (Y-axis) (see[140], and **Methods**). Colour intensity indicates the -log10 p-value after Benjamini-Yekutieli correction of the hypergeometric overlap.

(J,K) CpG methylation status of specific loci in repeat elements in mouse PSC adapted to the indicated conditions. Major Satellites (J) and IAP elements (K) were assessed by pyrosequencing (see **Methods**)[147]. (J) Data is shown for two CpG methylation sites located in Major Satellite repeats independently, or the Mean +/- SD of the methylation levels across the two CpG loci. (K) Data is shown for four CpG methylation sites located in IAP repeats independently, or the Mean +/-SD of the methylation levels across the four CpG loci.

(L) XIST RNA levels in human PSC lines, determined by qPCR. Human PSC lines in this study show extremely low levels of XIST compared to the control adult female human somatic cells (lung fibroblasts). Since the human PSCs display very low XIST expression even in the primed state, this indicates a loss of XIST expression, and suggests that erosion of X-silencing may have already occurred in the parental cells. Note that CDK8/19i treatment did not reactivate XIST expression, a phenomenon which has been recently reported to occur in specific media cocktails which also induce several aspects of the naïve human pluripotent state [90–92]. In summary therefore, CDK8/19i treatment does not recapitulate this reported reactivation of XIST RNA expression after X-silencing erosion, indicating a distinction with the media cocktails based on MEK-inhibition.

Fig. S5


A
Phosphorylation decrease in CDK motif
Observed in 2i and in CDK8/19i at +15min
Apoptosis, DNA repair
TFPT
DNA methylation
Other
Cell Cycle
MAP4 CHAMP1
OTUD4 LMNB1
NDEL1 LMNB2 TPR SPTBN1
Transcription, Regulation of transcription from Pol II promoter, RNA processing
MYC
AFF4
ILF3
SAMD1
SRRM1
SCAF1
TP53BP1
AEBP2
PRPF38A
EPOP
BRD8
25 proteins
28 phospho-sites
B
[Phospho-ERK/Total-ERK]
Relative expression
Phopho-ERK1/2
Total-ERK1/2
Clone
Serum/LIF
2i
CDK8/19i
C
CDK8 target phosphorylation
STAT1 Ser727P
STAT1 Total
Primed
2i p38iJNKi
CDK8/19i only
CDK8/19i p38iJNKi
p38iJNKi
2i only
D
Serum/LIF
2i-naive
MEKi
GSK3i
96 hrs
Cell Count
CDK8
E
Ser5P RNA Pol II
(this study)
Serum/LIF
2i
CDK8/19i
Normalized mean density (reads/50bp)
TSS +/-5Kb
N=28441 TSS
-2Kb
TSS
+2Kb
G
Total RNA Pol II
Marks et al., 2012.
TNGA_Serum/LIF
TNGA_2i
F
Nanog promoter ChIP-qPCR
% Input mean +/- SD
IgG
Total
Ser5P
Ser2P
K4me3
K27me3
RNA Pol II
Histone H3
H
Intra-genic RNA Pol II abundance ratios
Promoter
Gene Body
Term. Zone
Promoter Loading Index
(PLI) = increased
I
2i-naive:
(total Pol II at promoter) vs transcriptome
N=10117
-Log10
P-value
Transcriptome
2i UP
2i DOWN
ChIP: Pol II at TSS
Log2 FC
2i no change
2i UP
J
CDK8/19i:
(total Pol II at promoter) vs transcriptome
N=10136
CDK8/19i UP
CDK8/19i DOWN
CDK8/19i No change
CDK8/19i UP

Supplemental Figure S5.

CDK8/19-repression regulates the phospho-proteome and global RNA Pol II loading similar to 2i-naïve pluripotency

(A) Functional analysis of proteins which displayed a phosphorylation decrease in their CDK motif sequence, 15 minutes after the addition of 2i or CDK8/19-inhibitor to mouse PSC cells in culture. In total, 28 phospho-sites on 25 proteins displayed a decrease in both 2i and in CDK8/19i treatments at +15mins. Data from the treatment of two mouse PSC lines.

(B) Western blot analysis of ERK1/2 phosphorylation after long-term adaption (3 weeks) of mouse ES cells to 2i or CDK8/19i, as indicated. Above, plot of relative ERK1/2 phospho-levels, normalized by total ERK1/2 levels.

(C) Western blot for levels of the known CDK8 kinase target STAT1 phospho-serine727, in human iPS cells treated with the indicated inhibitor cocktails.

(D) CDK8 protein levels per cell measured by cytometry in mouse ES cells treated with the indicated inhibitors.

(E) Left panel: heatmaps of RNA Pol II Serine 5 phosphorylation (Ser5P) density at all Refseq Transcription Start Sites (TSS; n = 28,441) +/-5Kb, for mouse PSCs adapted to the indicated culture conditions, determined by ChIPseq (3 pooled replicates, see **Methods**). Right panel: metagene average RNA Pol II Ser5P density at all Refseq Transcription Start Sites (TSS; n = 28,441) +/-2Kb, for mouse PSCs adapted to the indicated culture conditions.

(F) ChIP-qPCR for RNA Pol II and histone marks at the *Nanog* TSS. Abundance of RNA Pol II, total and phosphorylated forms, are increased. Also, the histone mark associated with active euchromatin, H3K4me3, is increased, while in contrast there is no change in the repressive mark H3K27me3. Data = Mean +/- SD of three ChIP replicate experiments.

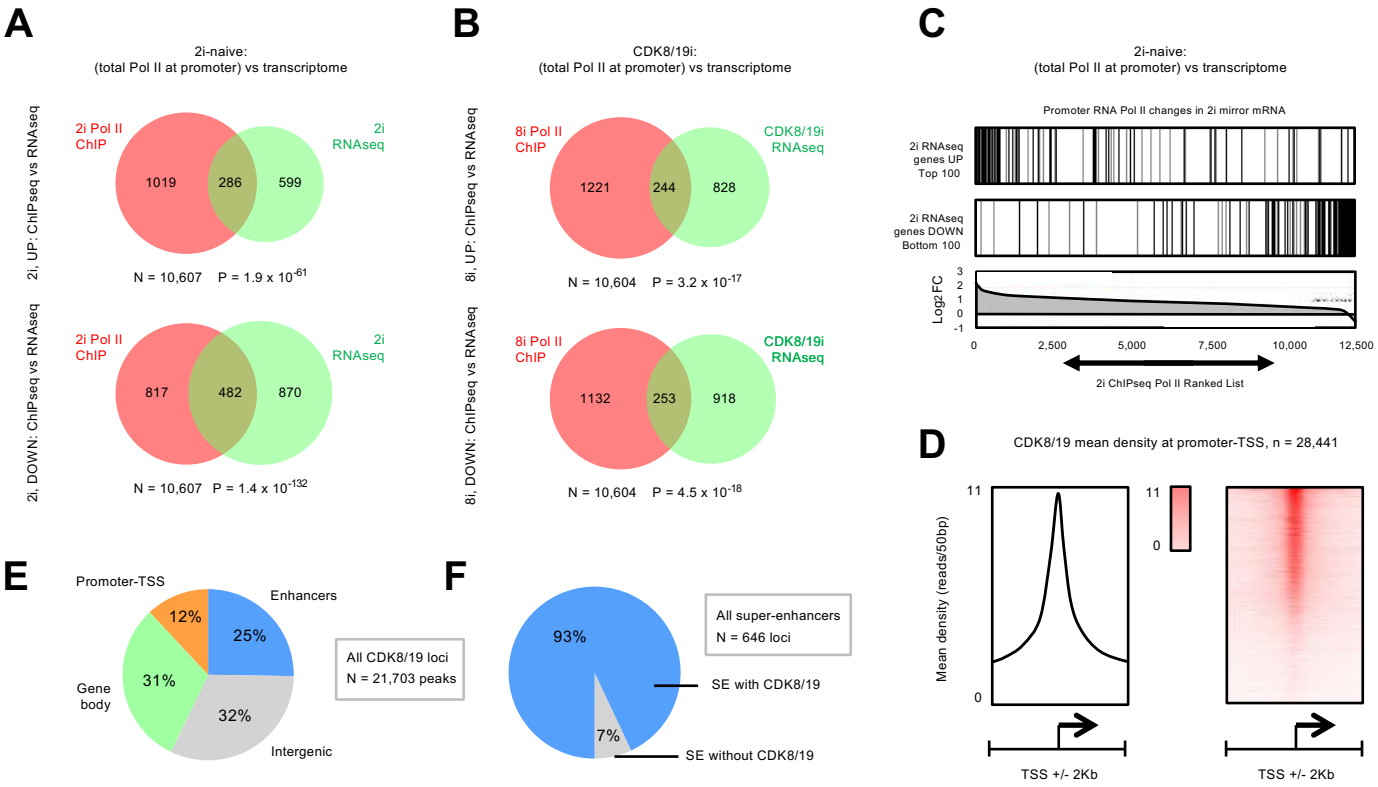
(G) Left panel: heatmaps of RNA Pol II density at all Refseq Transcription Start Sites (TSS; n = 28,441) +/- 5Kb, for mouse PSCs adapted to the indicated culture conditions, determined by ChIPseq as reported[38]. Right panel: metagene average RNA Pol II density at all Refseq TSS (n = 28,441) +/- 2Kb, for mouse PSCs adapted

to the indicated culture conditions. Note: this ChIPseq dataset[38], represents total RNA Pol II ChIPseq in mouse ES cells cultured in similar conditions to the current study, and is shown for comparison to the data from the current study in Figure 4E and 4F.

(H) Schematic defining the gene regions and Pol II loading ratios used in this study, which are similar to previous reports[95,97]. Lower panel depicts a schematic summary of the results in Figures 4E-4G, where Promoter Loading Index is increased (Promoter/Body). Note how the (Promoter/Termination-Zone) ratio is also increased, but the (Termination-zone/Body) ratio is constant. This reflects a preferential increase in Pol II loading at the promoter.

(I,J) Rank-Rank Hypergeometric Overlap (RRHO)[140] analysis of mRNA expression changes in mouse PSCs adapted to 2i-naïve conditions (I, n = 10,117) or CDK8/19i (J, n = 10,136), versus, change in abundance of RNA Pol II at the promoter of same set of genes. RNAseq-detected mRNA expression changes are arranged according to their magnitude (Y-axis), and then assessed for overlap by RRHO compared to the same set of genes ranked by change in RNA Pol II abundance detected by ChIPseq (X-axis) (see[140] and **Methods**). Colour intensity indicates the -log10 p-value after Benjamini-Yekutieli correction of the hypergeometric overlap. Genes defined as up or down in inhibitor-treated cells were determined versus control serum/LIF conditions.

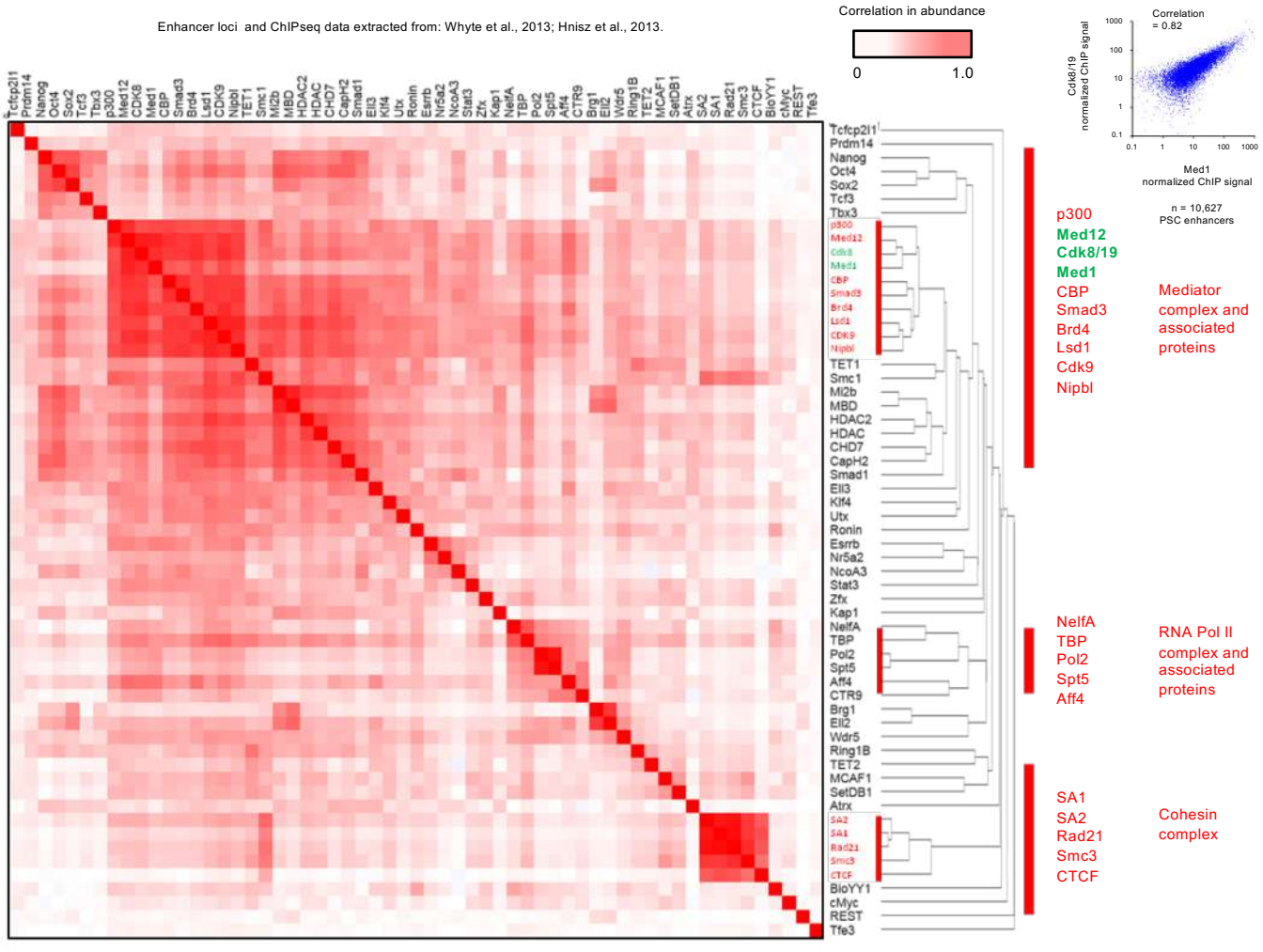
Fig. S6



G

Correlation Matrix: a comparison of ChIPseq signal intensity of 59 chromatin factors at 10,627 ES cell enhancers

Enhancer loci and ChIPseq data extracted from: Whyte et al., 2013; Hnisz et al., 2013.



Supplemental Figure S6.

ChIPseq for RNA Pol II and CDK8 and analysis.

(A,B) Venn diagrams showing how mRNA expression levels overlap significantly with genes where the promoter has the greatest change in RNA Pol II abundance in 2i-naïve conditions (A), or CDK8/19i (B). Genes most affected in terms of RNA Pol II abundance at promoter (red circles; promoter with fold change more than one standard deviation from the mean), versus, differential transcript expression (green circles; FDR<0.01). Genes up (top Venn diagram), and genes down (lower Venn diagram) are in inhibitor-treated cells versus control serum/LIF conditions.

(C) Change in mRNA expression levels at genes where the promoters have the greatest change in RNA Pol II abundance in 2i-naïve conditions. The top 100 most differentially expressed mRNAs up- or down-regulated in 2i-naïve conditions (top two panels) are arranged according to their rank in terms of differential Pol II loading on all TSS (bottom panel). The RNAseq and Pol II ChIPseq data represents the comparison of 2i-naïve conditions versus control-serum/LIF. Changes in Pol II loading at the promoter correlate with mRNA changes in the same genes.

(D) CDK8/19 average ChIPseq enrichment[2,3] density in mouse ES cells at Promoter-TSS regions +/- 2 Kb, n = 28,441 TSS (Refseq).

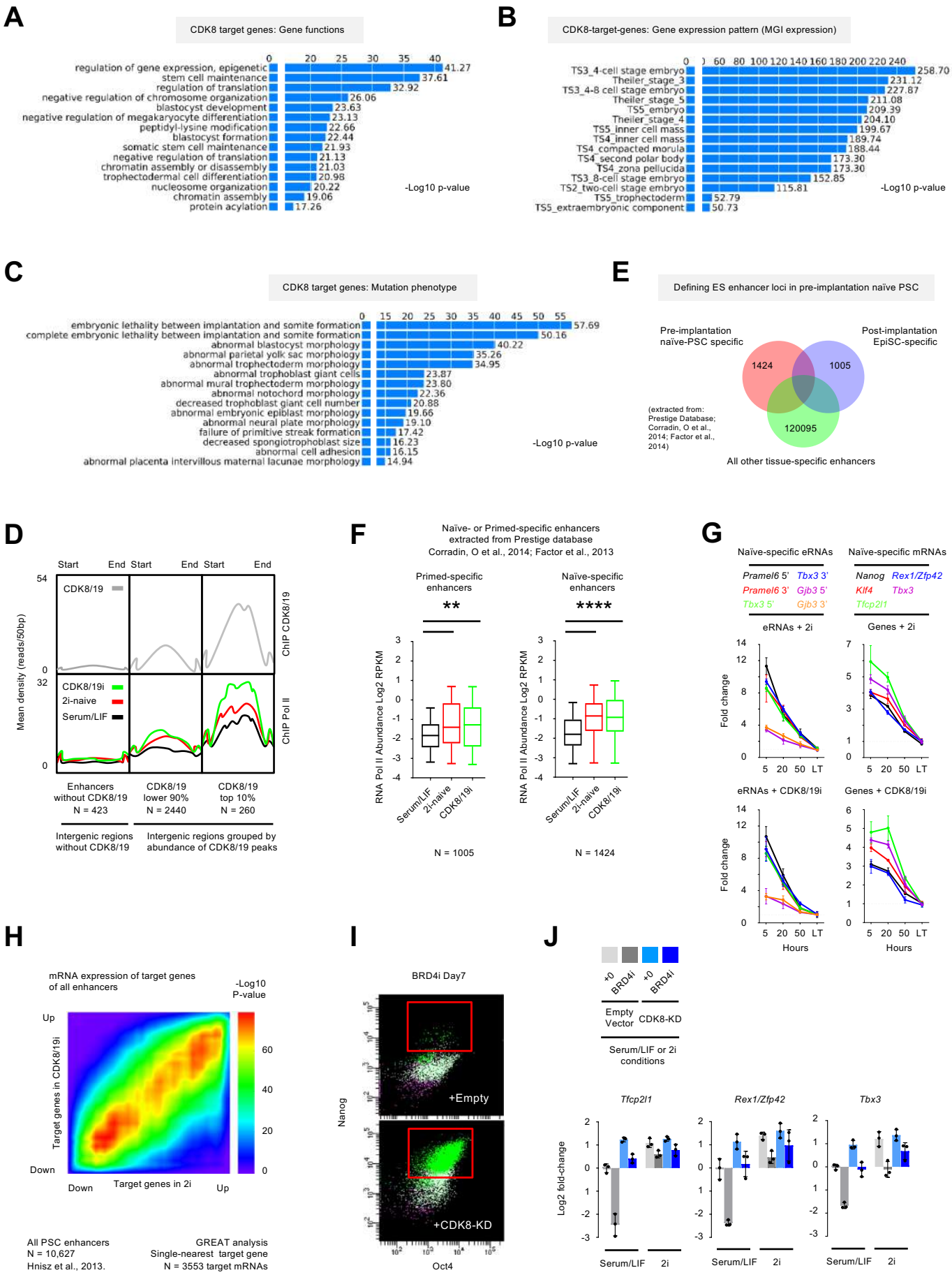
(E) CDK8/19 binding loci defined in mouse PSCs by ChIPseq[2,3], MACS peak calling, and categorized by functional annotation of the region by HOMER (see Table S7). Note: ChIP antibody binds both CDK8 and CDK19, see **Methods**. Promoter-TSS: TSS+/- 1Kb. Gene Body: Exons, Introns, and transcription termination site TTS +/- 1Kb. Enhancer constituent regions as defined[2,3].

(F) Percentage of SE-constituent regions enriched for CDK8/19 binding (see also Table S7).

(G) Correlation Matrix based on comparison of ChIP-seq signal intensity of 59 factors at 10,627 ES cell enhancers. 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 stem cell 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, consistent with their reported combinatorial roles at enhancers. CDK8/19 clusters most-closely within the Mediator complex and other critical regulators of enhancer function. See **Methods** for analysis of the published ChIP datasets and enhancer loci defined by [2,3].

Fig. S7


A
CDK8 target genes: Gene functions
regulation of gene expression, epigenetic 41.27
stem cell maintenance 37.61
regulation of translation 32.92
negative regulation of chromosome organization 26.06
blastocyst development 23.63
negative regulation of megakaryocyte differentiation 23.13
peptidyl-lysine modification 22.66
blastocyst formation 22.44
somatic stem cell maintenance 21.93
negative regulation of translation 21.13
chromatin assembly or disassembly 21.03
trophectodermal cell differentiation 20.98
nucleosome organization 20.22
chromatin assembly 19.06
protein acylation 17.26
-Log10 p-value
B
CDK8-target-genes: Gene expression pattern (MGI expression)
TS3_4-cell stage embryo 258.70
Theiler_stage_3 231.12
TS3_4-8 cell stage embryo 227.87
Theiler_stage_5 211.08
TS5_embryo 209.39
Theiler_stage_4 204.10
TS5_inner cell mass 199.67
TS4_inner cell mass 189.74
TS4_compacted morula 188.44
TS4_second polar body 173.30
TS4_zona pellucida 173.30
TS3_8-cell stage embryo 152.85
TS2_two-cell stage embryo 115.81
TS5_trophectoderm 52.79
TS5_extraembryonic component 50.73
-Log10 p-value
C
CDK8 target genes: Mutation phenotype
embryonic lethality between implantation and somite formation 57.69
complete embryonic lethality between implantation and somite formation 50.16
abnormal blastocyst morphology 40.22
abnormal parietal yolk sac morphology 35.26
abnormal trophectoderm morphology 34.95
abnormal trophoblast giant cells 23.87
abnormal mural trophectoderm morphology 23.80
abnormal notochord morphology 22.36
decreased trophoblast giant cell number 20.88
abnormal embryonic epiblast morphology 19.66
abnormal neural plate morphology 19.10
failure of primitive streak formation 17.42
decreased spongiotrophoblast size 16.23
abnormal cell adhesion 16.15
abnormal placenta intervillous maternal lacunae morphology 14.94
-Log10 p-value
E
Defining ES enhancer loci in pre-implantation naïve PSC
Pre-implantation naïve-PSC specific
Post-implantation EpiSC-specific
1424
1005
120095
(extracted from: Prestige Database; Corradin, O et al., 2014; Factor et al., 2014)
All other tissue-specific enhancers
D
Start End Start End Start End
CDK8/19
CDK8/19i
2i-naive
Serum/LIF
ChIP CDK8/19
ChIP Pol II
Mean density (reads/50bp)
Enhancers without CDK8/19 N = 423
CDK8/19 lower 90% N = 2440
CDK8/19 top 10% N = 260
Intergenic regions without CDK8/19
Intergenic regions grouped by abundance of CDK8/19 peaks
F
Naïve- or Primed-specific enhancers extracted from Prestige database
Corradin, O et al., 2014; Factor et al., 2013
Primed-specific enhancers
Naïve-specific enhancers
**
****
RNA Pol II Abundance Log2 RPKM
Serum/LIF
2i-naive
CDK8/19i
N = 1005
N = 1424
G
Naïve-specific eRNAs
Naïve-specific mRNAs
Pramel6 5' Tbx3 3'
Pramel6 3' Gjb3 5'
Tbx3 5' Gjb3 3'
Nanog Rex1/Zfp42
Klf4 Tbx3
Tfcp2l1
eRNAs + 2i
Genes + 2i
eRNAs + CDK8/19i
Genes + CDK8/19i
Fold change
5 20 50 LT
Hours
H
mRNA expression of target genes of all enhancers
-Log10 P-value
Up
Down
Target genes in CDK8/19i
Target genes in 2i
All PSC enhancers N = 10,627 Hnisz et al., 2013.
GREAT analysis Single-nearest target gene N = 3553 target mRNAs
I
BRD4i Day7
+Empty
+CDK8-KD
Nanog
Oct4
J
+0 BRD4i
+0 BRD4i
Empty Vector
CDK8-KD
Serum/LIF or 2i conditions
Tfcp2l1
Rex1/Zfp42
Tbx3
Log2 fold-change
Serum/LIF
2i

Supplemental Figure S7.

ChIPseq for RNA Pol II and CDK8 and analysis. 2i and CDK8/19i hyper-actívate naïve-state enhancer activity

(A,B,C) Gene ontology enrichment and functional annotation of CDK8/19-target genes. CDK8/19 target genes were defined by the single-nearest gene to each CDK8/19 binding site identified by ChIPseq and MACS peak calling (related to Figure 5A; see also Methods). Gene ontology and functional annotation were derived for all CDK8/19-target genes by GREAT analysis.

(D) Average ChIPseq enrichment in the indicated genomic regions is shown for CDK8/19 or RNA Pol II binding, as determined in mouse PSCs. Genomic regions were defined in groups by the relative level of CDK8/19. CDK8/19 peak intensity as defined in (A) above.

(E) Derivation of mouse ES super-enhancer loci specific to pre-implantation naïve epiblast, or post-implantation primed epiblast. Enhancer loci were extracted from the Prestige Database[108,148]. The SEs in naïve or primed epiblast were first selected (see, Methods), and then any SEs common (overlapping locus) to a panel of 16 somatic tissues were subtracted (see Methods).

(F) RNA Pol II abundance in mouse primed-specific super-enhancers (on left), or naïve-specific super-enhancers (on right), as defined in (D) above, using the PREStige database[108,148], see also Methods. ** P<0.01; **** P<0.0001.

(G) Pluripotency marker genes and naïve-specific eRNA abundance measured by qRT-PCR in mouse PSC at time intervals after withdrawal of 2i or CDK8/19i from the culture. Naïve-specific eRNAs and primers, as defined [100]. Mean +/- SEM, of three independent cell experiments.. Related to Figure 5D.

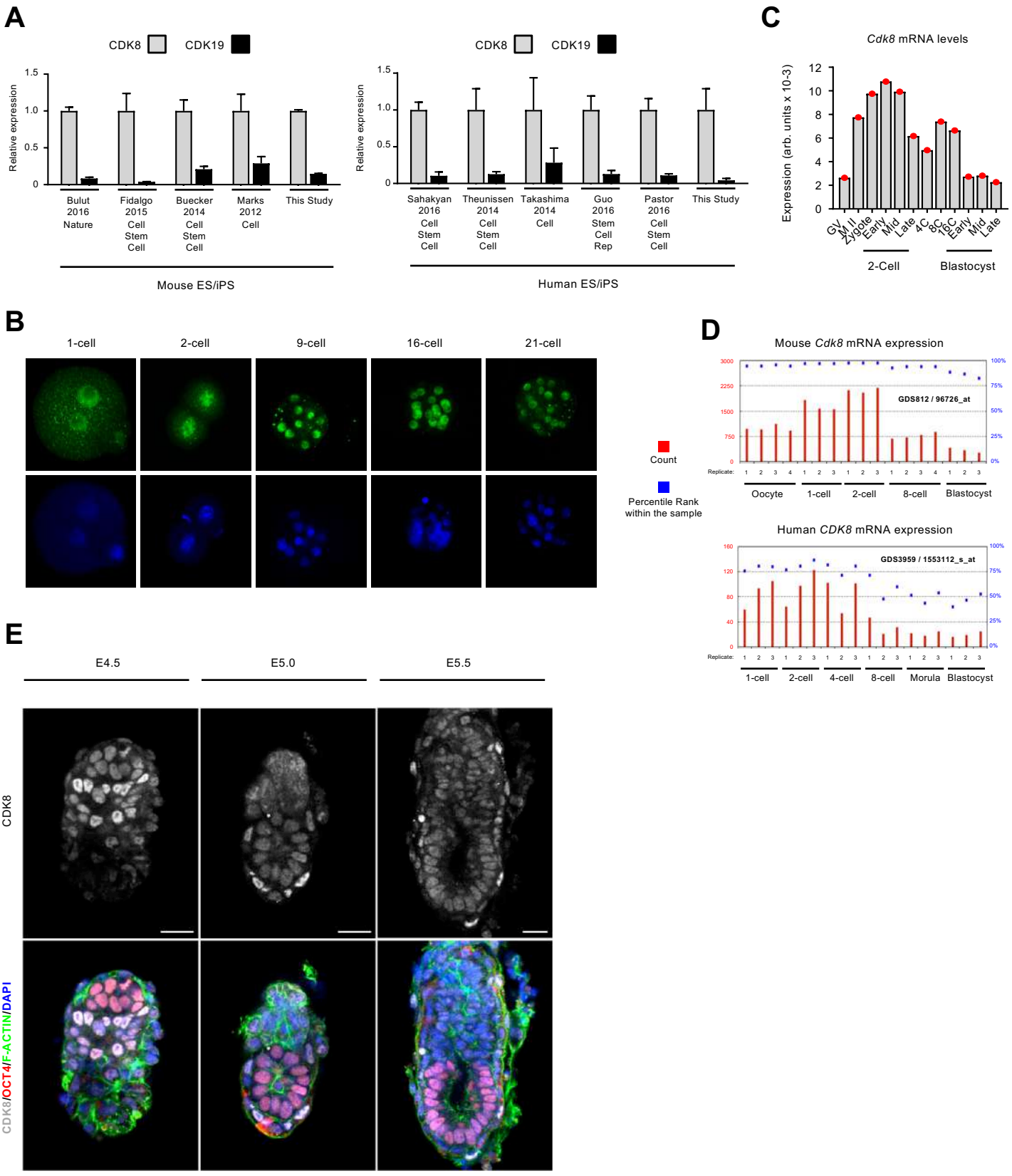
(H) Rank-Rank Hypergeometric Overlap (RRHO)[140] analysis of mRNA expression changes for the single-nearest target genes (N = 3,553) identified for all PSC enhancers (N = 10,627), as defined[2,3]. RNA expression of enhancer-target genes in 2i-naïve conditions (X-axis) or CDK8/19i (Y-axis), is compared to control serum/LIF conditions. RNAseq-detected mRNA expression changes are arranged according to their magnitude, and then assessed for overlap by RRHO (see also:

Methods). Colour intensity indicates the -log10 p-value after Benjamini-Yekutieli correction of the hypergeometric overlap. Highly significant overlap along the diagonal indicates a similar regulation of enhancer-target gene mRNA expression in 2i and CDK8/19i.

(I) FACS measurement of NANOG and OCT4 protein expression following treatment with 500 nM BRD4i/JQ1for 7 days, in CDK8/19-double-knockout (CDK8/19-dKO) iPS cells stably expressing pMSCV-Empty or pMSCV-CDK8-Kinase Dead (CDK8-KD). Representative of three independent cell experiments.

(J) qRT-PCR expression of naïve marker genes following treatment with 500 nM BRD4i/JQ1 for 48 h. CDK8/19-dKO iPS +/- CDK8-KD were cultured +/- 2i or standard serum/LIF, as indicated. Mean +/- SD of three clones. See also Figure 5H.

Fig. S8


A
CDK8
CDK19
Relative expression
1.5
1.0
0.5
0
Bulut 2016 Nature
Fidalgo 2015 Cell Stem Cell
Buecker 2014 Cell Stem Cell
Marks 2012 Cell
This Study
Mouse ES/iPS
Sahakyan 2016 Cell Stem Cell
Theunissen 2014 Cell Stem Cell
Takashima 2014 Cell
Guo 2016 Stem Cell Rep
Pastor 2016 Cell Stem Cell
This Study
Human ES/iPS
C
Cdk8 mRNA levels
Expression (arb. units x 10-3)
12
10
8
6
4
2
0
GV
M II
Zygote
Early
Mid
Late
4C
8C
16C
Early
Mid
Late
2-Cell
Blastocyst
B
1-cell
2-cell
9-cell
16-cell
21-cell
D
Mouse Cdk8 mRNA expression
GDS812 / 96726_at
Count
Percentile Rank within the sample
Oocyte
1-cell
2-cell
8-cell
Blastocyst
Human CDK8 mRNA expression
GDS3959 / 1553112_s_at
1-cell
2-cell
4-cell
8-cell
Morula
Blastocyst
E
E4.5
E5.0
E5.5
CDK8
CDK8/OCT4/F-ACTIN/DAPI

Supplemental Figure S8.

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 five mouse datasets and six human datasets (including the current study).

(B) Immunofluorescence for CDK8 protein levels during mouse preimplantation development from 1-Cell to early blastocyst stage (E3.5).

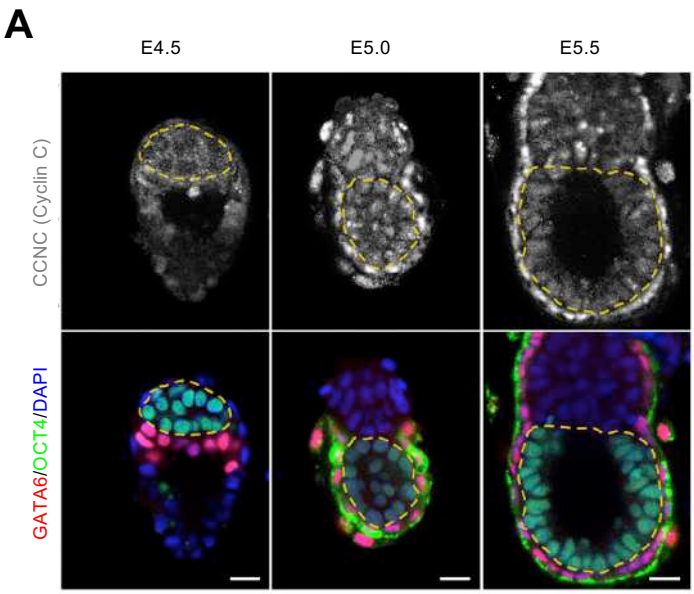
(C) CDK8 mRNA expression levels in specific embryo stages and lineages during mouse preimplantation development. Data from published studies [102].

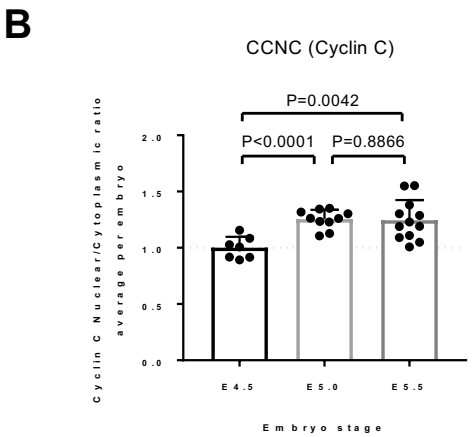
(D) CDK8 mRNA expression levels during mouse or human embryo pre-implantation development, as detected by microarray in published datasets.

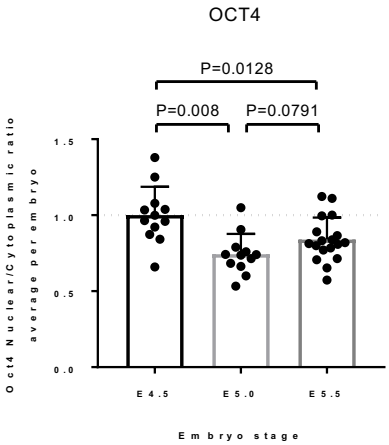
Mouse:
https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS812:96726_at
Human:
https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS3959:1553112_s_at

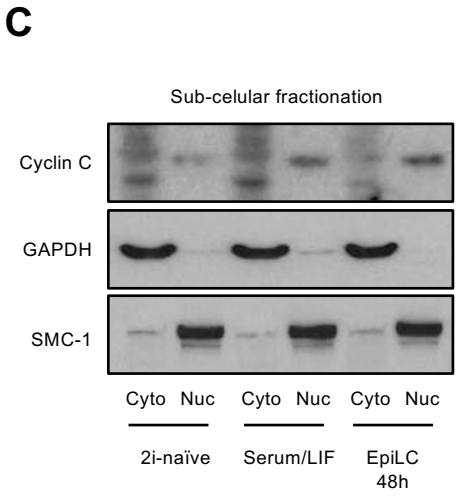
(E) Immunofluorescence for CDK8, OCT4, and F-ACTIN in mouse early embryos from E4.5 to E5.5. Scale bars = 20 μm. Images representative of three experiments.

Fig. S9


A
E4.5
E5.0
E5.5
CCNC (Cyclin C)
GATA6/OCT4/DAPI


B
CCNC (Cyclin C)
P=0.0042
P<0.0001
P=0.8866
Cyclin C Nuclear/Cytoplasmic ratio average per embryo
2.0
1.5
1.0
0.5
0.0
E4.5
E5.0
E5.5
Embryo stage


OCT4
P=0.0128
P=0.008
P=0.0791
Oct4 Nuclear/Cytoplasmic ratio average per embryo
1.5
1.0
0.5
0.0
E4.5
E5.0
E5.5
Embryo stage


C
Sub-celular fractionation
Cyclin C
GAPDH
SMC-1
Cyto Nuc
Cyto Nuc
Cyto Nuc
2i-naïve
Serum/LIF
EpiLC
48h

Supplemental Figure S9.

Cyclin C expression localization during mouse preimplantation development

(A,B) In (A), representative examples are shown of two independent imunofluorescence stainings for cyclin C protein levels during mouse development from preimplantation blastocyst stage (E4.5) to post-implantation cylinder stage (E5.5). Co-staining was performed with OCT4 to mark the epiblast, and GATA6, to mark the primitive endoderm at E4.5 and its maturation into post-implantation visceral endoderm. In (B), Cyclin C nuclear-cytoplasmic ratio was quantified and plotted, where each data point represents the mean Nuc-Cyto ratio for the epiblast cells of one embryo. As an internal control, the Nuc-Cyto ratio for OCT4 was also quantified. We observe a nuclear-cytoplasmic pattern, where 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, suggesting that the pattern of cyclin C is not related to artefacts of staining or imaging.

(C) Western blot analysis of cyclin C localization by sub-cellular localization. Nuclear and cytoplasmic fractions were prepared from mouse cells across the developmental spectrum from naïve (adapted to 2i), metastable and primed-like (adapted to serum/LIF), or primed (derived by 48 h treatment of PSCs with EpiSC media, forming EpiLC cells, as described [149], and see **Methods**. The relative abundnace of nuclear cyclin C is greater in primed state EpiLC and in serum/LIF conditions, compared to 2i-naïve.

MATERIALS AND METHODS

KEY REAGENT TABLES

PRIMERS

Mouse qRT-PCR primers used in this study		
Target gene (mouse)	**Forward 5'-3'**	**Reverse 5'-3'**
β-Actin	GGCACCACACCTTCTACAATG	GTGGTGGTGAAGCTGTAGCC
Gapdh	TTCACCACCATGGAGAAGGC	CCCTTTTGGCTCCACCCT
Pou5f1/Oct4	TCTTTCCACCAGGCCCCCGGCTC	TGCGGGCGGACATGGGGAGATCC
Sox2	TAGAGCTAGACTCCGGGCGATGA	TTGCCTTAAACAAGACCACGAAA
Klf4	GCGAACTCACACAGGCGAGAAACC	TCGCTTCCTCTTCCTCCGACACA
Nanog	CAGGTGTTTGAGGGTAGCTC	CGGTTCATCATGGTACAGTC
MERVL	CCCATCATGAGCTGGGTACT	CGTGCAGAGCCATCAGTAAA
MERVL (LTR – Int)	CTTCCATTCACAGCTGCGACTG	CTAGAACCACTCCTGGTACCAAC
Zscan4	GAGATTCATGGAGAGTCTGAC TGATGAGTG	GCTGTTGTTTCAAAAGCTTGA TGACTTC
Tcstv3	AGAAAGGGCTGGAACTTGT GACCT	AAAGCTCTTTGAAGCCATG CCCAG
LINE L1	TGGCTTGTGCTGTAAGATCG	TCTGTTGGTGGTCTTTTTGTC

Mouse eRNA qRT-PCR primers used in this study (as published [1])		
Target gene (mouse)	**Forward 5'-3'**	**Reverse 5'-3'**
Tbx3 5' 9.4 eRNA	GATTGTCCCACCACGAAACT	GGGGAGGATTTGTTTGGAAT
Tbx3 3' 9.4 eRNA	TCCTCCCCAAGATCTGTGTC	GCCGTAGTGGTGGAAATCTT
Gjb3 5' 16.1 eRNA	TTTTGCCACAAAACCCTACC	TGGGGCTACACAAAGAAACC
Gjb3 3' 16.1 eRNA	AGCAAGTCCCCAGAATCCTT	TATTCAGCCTGGGAAGATGG
Pramel6 5' 7.5 eRNA	AGTGATTCTCCCCGAGTGTG	TGGCCTCGAACTCAGAAATC
Pramel6 3' 7.5 eRNA	GCTTAGGCGGTTAAGTGTGC	TTACAAGCCTACCCCACTGC
Prdm14 52.3 5' eRNA	TGTGCTCAGTTGTGGAGGAG	GAAGCATTTGTGGGGTTGTT

Mouse ChIP-qPCR primers (locus indicated in brackets)		
Target gene	**Forward 5'-3'**	**Reverse 5'-3'**
Nanog (promoter 1F-1R)	AAATCTATCGCCTTGAGCCGT	CACCAACCAAATCAGCCTATCTG

Mouse DNA methylation primers		
Target gene	**Forward 5'-3'**	**Reverse 5'-3'**
Line1	AGTTTTTGGAATAG GTAGAAGTATAGAG	[Btn]ACAATTCCCAA ATAATACAAACTCT
Line1 sequencing primer **5'-3'**	TGAGGTAGTATTTTGTGT	
IAP	AGGGTGGTTTTTTAT TTTATGTGT	[Btn]TATCACTCCCTA ATTAACTACAACC
IAP sequencing primer **5'-3'**	TGAGGTAGTATTTTGTGT	
Major Satellite	GGAATATGGTAAGA AAATTGAAAATTATGG	[Btn]ACATATTCCAA ATCCTACAATATACAT
Major Satellite sequencing primer **5'-3'**	AATTATGGAAAATGAGAAATATTTA	

Human qRT-PCR primers used in this study		
Target gene (human)	**Forward 5'-3'**	**Reverse 5'-3'**
GAPDH	GGACTCATGACCACAGTCCATGCC	TCAGGGATGACCTTGCCCACAG
ACTIN	GGACTTCGAGCAAGAGATGG	AGCACTGTGTTGGCGTACAG
NANOG	TTTCAGAGACAGAAATACCTCAGC	TCACACCATTGCTATTCTTCG
OCT4	GACAGGGGGAGGGGAGGAGCTAGG	CTTCCCTCCAACCAGTTGCCCCAAAC
SOX2	GGGAAATGGGAGGGGTGCAAAAGAGG	TTGCGTGAGTGTGGATGGGATTGGTG
KLF4	ACGATCGTGGCCCCGGAAAAGGACC	TGATTGTAGTGCTTTCTGGCTGGGCTCC
CMYC	GCCCCTCAACGTTAGCTTCAC	GTAGAAATACGGCTGCACCGA

CDH1/E-CAD	CGAGAGCTACACGTTCACGG	GTGTCGAGGGAAAAATAGGCTG
CDH2/N-CAD	GAGGAGTCAGTGAAGGAGTCA	GGCAAGTTGATTGGAGGGATG
HERVH-POL	CGCCCTTCTTCCCAATCCAA	GCCAAGGAGGGAGTAGAGGT
HERVK/HML2	GGCCATCAGAGTCTAAACCACG	CTGACTTTCTGGGGGTGGCCG
SOX17	CGCTTTCATGGTGTGGGCTAAGGACG	TAGTTGGGGTGGTCCTGCATGTGCTG
GATA6	GAGGGTGAACCCGTGTGCAATG	TGGAAGTTGGAGTCATGGGAATGG
GATA4	ACACCCCAATCTCGATATGTTTG	GTTGCACAGATAGTGACCCGT
VIMENTIN	AGTCCACTGAGTACCGGAGAC	GGTTCCTTTAAGGGCATCCAC
SNAIL	CGAGCTGCAGGACTCTAAT	CCACTGTCCTCATCTGACA
NEUROD1	CCAAAAAGAAGAAGATGACTAAGG	AGCTGTCCATGGTACCGTAA
NESTIN	TTGCCTGCTACCCTTGAGAC	GGGCTCTGATCTCTGCATCTAC
KLF17	GCTGCCCAGGATAACGAGAAC	ATCTCTGCGCTGTGAGGAAAG
AFP	TTGGGCTGCTCGCTATG	TTTGTAACTGTTGCTGCCTTTG
FOXA2	GTGAAGATGGAAGGGCACG	CATGTTGCTCACGGAGGAGTAG
EOMES	TGCAGGGCAACAAAATGTATG	GTCTCATCCAGTGGGAACCAGTA
PAX6	GTCCATCTTTGCTTGGGAAA	TAGCCAGGTTGCGAAGAACT
MAP2	TTGGTGCCGAGTGAGAAGA	GTCTGGCAGTGGTTGGTTAA
β3-TUBULIN	TTCTGGGAAGTCATCAGTGATGA	CGAGTCGCCCACGTAGTTG
TUJ1	GGCCTTTGGACATCTCTTCA	CCTCCGTGTAGTGACCCTTG
GFAP	ACATCGAGATCGCCACCTAC	TCTGCAGGTTGGAGAAGGTC
NKX2.5	AAGTGTGCGTCTGCCTTT	GTTGTCCGCCTCTGTCTTC
TNNI3	CCAACTACCGCGCTTATGC	CTCGCTCCAGCTCTTGCTTT
MYH6	GACTGTTGTGGCCCTGTACC	GGAAGGATGAGCCCTTTTTC
MYL2	CGCCAACTCCAACGTGTTCT	CCATCCCTGTTCTGGTCCAT
MYL7	CCCATCAACTTCACCGTCTTCCT	AGAGAACTTGTCTGCCTGGGTCA

ANTIBODIES

Antibodies used in this study		
Target	**Company**	**Code**
Nanog (Western, IF)	Chemicon/Millipore	#AB5731
Nanog (FACS, IF)	eBiosciences	51-5761
Pou5f1/Oct4	BD Biosciences/ Pharmingen	611203
Total RNA Pol II (RPB1)	Santa Cruz	sc-899x (N-20)
RNA Pol II Ser-5P	Abcam	ab5131
RNA Pol II Ser-2P	Abcam	ab5095
SMC1	Bethyl Laboratories	A300-055A
Gapdh	Sigma	G8795
β-Actin	Sigma	A5441
γ-Tubulin	Sigma	#T6557, CLONE GTU-88 ascites fluid
Lamin A/C	Santa Cruz	sc-6215 (N-18)
CDK8	Cell Signaling	#4106 (P455) Target epitope is at P455 which is not present in CDK19. See also Figures 1 and S1 for validation of this antibody by knockout and knockdown strategies.
CDK19	Atlas Antibodies	HPA007053
CDK8/19 (both)	Santa Cruz	SC-1521 specificity for both [2] (see Methods for ChIP-Seq analysis)
Cyclin C	Santa Cruz	Sc-1061
STAT1 total	Cell Signaling	#9172
STAT1 Ser727-phospho	Cell Signaling	#9177
ERK1/2 total (p44/42)	Cell Signaling	#9102
ERK1/2 phospho (Thr202/Tyr204)	Cell Signaling	#9101
SSEA4	Stem Cell Technologies	#60062 Clone MC-813-70
Tra1-81	Millipore	MAB4381
cMyc	Santa Cruz	sc-40 (9E10)
Sox2	Chemicon/Millipore	#AB5603
Zscan4c	Millipore	AB4340
GATA6	R+D Systems	AF1700
Tfe3	Atlas Antibodies	HPA023881
ICAM1/CD54	eBiosciences	13-0541
H3K9me3	Upstate/Millipore	07-442
Alpha-fetoprotein (AFP)	Abcam	Ab46799
Vimentin	Santa Cruz	sc-6260

Nestin	ThermoFisher Scientific	MA1-110 (10C2)

Antibodies used for mouse embryo/embryoid immunofluorescence in this study			
Target	**Code**	**Company**	**Dilution**
Rabbit pAb anti-Nanog	ab80892	Abcam	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	ThermoFisher Scientific	1:500
Alexa Fluor 594 Donkey anti-Rat	A21209	ThermoFisher Scientific	1:500

Antibodies used for human teratoma/embryoid immunofluorescence			
Target	**Code**	**Company**	**Dilution**
Alpha-fetoprotein	A0008, Lot# 20016625	Dako	1:200
FOXA2	AF2400, Lot# ULB0414041	R&D Systems	1:50
Troponin T cardiac isoform	MS-295-P1ABX, Clone 13-11; Lot# 295X1601A	Thermo Scientific	1:200
GATA4	sc-9053, Lot# L1014	Santa Cruz	1:25
TUJ1	MMS-435P, Clone TUJ1, Lot# D14JF02140	Biolegend	1:500
GFAP	Z0334, Lot# 20019134	Dako	1:1000
Vimentin	ab92547, Clone EPR3776, Lot# GR219216-33	Abcam	1:200

α-Smooth muscle actin (SMA)	A5228, Clone 1A4, Lot# 074M4814V	Sigma	1:400

shRNAs

shRNAs used in this study	
From: Open Biosytems (TRC Mission Library) with a pLKO.1 lentiviral backbone.	
shRNA	**Clone**
Mouse Cyclin C (CCNC) Open Biosystems #RMM4534-EG51813	
shRNA Cyclin C #1	TRCN0000077828
shRNA Cyclin C #2	TRCN0000077829
shRNA Cyclin C #3	TRCN0000077830
shRNA Cyclin C #4	TRCN0000077831
shRNA Cyclin C #5	TRCN0000077832
Mouse Cdk8 Open Biosystems #RMM3981	
shRNA Cdk8 #1	TRC0000023104
shRNA Cdk8 #2	TRC0000023105
Mouse CDK19 Open Biosystems #RMM4534-EG78334	
shRNA Cdk19 #1	TRCN0000023279
shRNA Cdk19 #2	TRCN0000023280
shRNA Cdk19 #3	TRCN0000023281
shRNA Cdk19 #4	TRCN0000023282
shRNA Cdk19 #5	TRCN0000023283

shRNA shSCR non-targetting	scramble shRNA was acquired from Addgene (plasmid 1864)	Non-targeting
pMD2.G	Lentiviral envelope packaging plasmid Addgene plasmid #12259	
pCMV-dR8.91	Lentiviral envelope packaging plasmid Harvard Medical School, plasmid #516	

CRISPR-Cas9 guide RNAs and vector systems

Guide RNA name	System	Target sequence	Location in Cdk19	Predicted %AA's deleted
Cdk19 CRISPR #2	pLentiCRISPRv2 (Addgene #52961)	5'- 3' AAAGTGGGACGCGGCACCTA At +341 to +361 of mouse Cdk19 gene.	Mid Exon 1; +76 bp downstream of ATG start	94%

CONTACT FOR REAGENT AND RESOURCE SHARING

Please contact Manuel Serrano. Manuel.serrano@irbbarcelona.org

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse and human work

Animal experimentation at the Spanish National Cancer Research Centre CNIO (from the name in Spanish: Centro Nacional de Investigaciones Oncológicas) was performed according to protocols 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 the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012, and reviewed by the University of Cambidge Animal Welfare and Ethical Review Body (AWERB). Cdk8 flox/flox RERT-Cre mice were generated by the laboratory of Daniel Fisher (IGMM, Montpellier).

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

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)26Sor*$^{tm1(CAG-EGFP)Luo}$/J mice (Jackson 006053) and from Tg.CAG-Katushka mice [3], respectively. *Nanog*-GFP knock-in mouse ES cells (TNGA, TON) were previously described [4] 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 [5]; The ZS mouse ES line was a 2C-reporter shared by the laboratory of Minoru Ko [6]. Mouse ES cells and iPS cells, were routinely cultured on gelatin-coated plates in a base media of either "Serum/LIF" (15% FBS), or Knockout Serum Replacement (KSR, Invitrogen) "KSR/LIF" (15% KSR), in DMEM (high glucose) basal media, with LIF (1000 Units/mL), non-essential amino acids, glutamax and β-mercaptoethanol plus antibiotics. Where used with mouse PSC, the "2i" two-inhibitor cocktail comprised 1 μM MEK-inhibitor (PD0325901, Axon Medchem, #1408) plus 3 μM GSK3β-inhibitor (CHIR 99021, Axon Medchem #1386) as described [7]. Cultures were routinely tested for mycoplasma. Primary mouse embryo fibroblasts (wild-type, MEFs, passage 2) were obtained at E13.5 from pure inbred C57BL6 background mice as described previously [8], or from CDK8 flox/flox RERT-Cre mice. Human 293T cells were from ATCC. All the above-mentioned cells were maintained in DMEM medium with 10% FBS (Gibco) with antibiotics (penicillin/streptomycin 100 U/ml). Reprogrammed iPS cells were initially derived and expanded on mitomycin-C inactivated feeder cells on gelatin-coated plates, before transfer to gelatin-only.

Human PSC cell resources

HERVH iPS were shared by the laboratory of Zsussanna Izsvak (Max Delbruck Centre for Molecular Medicine) [9]. WIBR3 ES cells were shared by the laboratory of Jacob Hanna (Weizmann Institute of Science). OSCAR ES cells carrying inducible STAT3 were shared by the laboratory of Pierre Savatier (SBRI, Stem Cell and Brian Research Institute) [10]. H1 and H9 human ES cells, and CB5, D2#2, and D2#4 human iPS cells, were shared by the laboratory of Nuria Montserrat (IBEC, Institute for Bioengineering).

Human PSC cell culture in primed state

Human PSC (H1, H9, WIBR3, HERVH, CB5, D2#2, D2#4, OSCAR) were maintained in conventional primed conditions as described [9–11], specifically, by culture on growth factor-reduced phenol red-free matrigel (BD Biosciences #356231) with mTeSR1 media (Stem Cell Technologies). Cultures were passage every 5-7 days manually using either 2mg/ml dispase (Gibco), 0,5 μM EDTA/1xPBS, or accutase (Gibco).

Resetting human PSC from primed to naïve state using 2i-based media cocktail

The naïve human pluripotent state was obtained by two methods. OSCAR cells were reset to the naïve state with 2i (TL2i) or CDK8/19i (1,1 μM or 0,4 μM) plus rhLIF and STAT3 transgene induction, essentially as described [10]. In a transgene-free approach, the human PSCs were cultured in a 2i-based chemical cocktail [11] referred to here in the text as "2i p38iJNKi". Cells were maintained on matrigel (BD Biosciences #356231) using mTeSR1 (Stem Cell Technologies), and the media was supplemented with 10 μg/ml of recombinant human LIF (Peprotech, as described [11], 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). In order to obtain and maintain the naïve state using the 2i p38iJNKi media cocktail, cells were selected at each passage, by sorting in cytometry for the top 10% HERVH-GFP levels, or by repeated manual picking for selection of colonies with dome-shaped morphology. Initial conversion of

the human PSC from primed to naïve required 3 passages/rounds of selection, over 14-18 days.

Resetting human PSC from primed to naïve pluripotent state using CDK8/19i

To adapt and maintain human PSC to CDK8/19i culture (CDK8/19i-adapted), cells were maintained on matrigel (BD Biosciences #356231) using mTeSR1 (Stem Cell Technologies), and the media was supplemented with 10 μg/ml of recombinant human LIF (Peprotech), as described [11], plus 0,4 or 1.1 μM of CNIO-CDK8/19 inhibitor, or plus 10 μM of SenexinA-CDK8/19 inhibitor [12]. This adaption process can also include 10 μM SP600125 (JNKi, TOCRIS) plus 2 μM BIRB796 (p38i, Axon Medchem), but they are not required. Following background cell death in the first passage, colonies gradually become dome-shaped within 10-14 days without additional selection, and could be expanded using 3–5 min treatment with 0,5 μM EDTA/1xPBS when necessary to avoid confluency, usually every 5-7 days due to a slowdown in proliferation. The optimal CDK8/19i concentration was 1.1μM for HERVH-GFP hiPSC, while it was 0,4μM for all other human cell lines.

mESC derivation

To test mouse ES cell derivation in the presence of 2i or CDK8/19i, 8-cell stage mouse embryos were recovered from the oviducts of pregnant females and cultured in serum/LIF on mitomycin C-inactivated MEF feeders plus 2i or CDK8/19i, in order to derive ES cell lines by standard methods [13]. Inhibitors were added fresh every 2 days until emergence of colonies from hatched blastocysts. We noted that feeders were not compatible with several days of CDK8/19i, therefore the cells were passed every 2 days to fresh feeders, and then moved to culture on 0,1% gelatin. Colonies were confirmed as ES cells by immunofluorescence and PCR methods (see below).

EpiSC Derivation

E14 WT mES cells in 2i/LIF cultured on gelatin were firstly induced to differentiate into Epiblast-Like Cells (EpiLCs) over a 48 h period as described[14–16]. Briefly, the mES cells were seeded on fibronectin-coated plates (10ng/ml) and switched to media containing 1%KSR, N2B27, FGF2 (12ng/ml) and Activin A (20ng/ml). In 48h, the cells were in a flat EpiLC state. After 48h, the media was switched to include 20% KSR and expanded for 5 passages to stabilize the cells in the EpiSC primed state, confirmed by typical flat colony morphology and Fgf5 expression. Flattened EpiSC colonies were passaged as clumps, as described[14–16].

Analysis of PSC self-renewal and developmental pluripotency

Mouse or human PSC self-renewal and pluripotency was scored by colony morphology, by cytometry (mouse: *Nanog*-GFP heterogeneity and overall intensity, and co-staining for ICAM1; human: HERVH-GFP intensity, and assessing expression of NANOG, OCT4, SSEA4, TRA1-81), by immunofluorescence for pluripotency markers indicated in the Figures, by alkaline phosphatase staining of fixed cells (Promega #S3771), and by qRT-PCR for pluripotency markers, including *NANOG*, *OCT4*, *KLF4*, *TFCP2L1*, and *SOX2* (See: **Figures 1 and 2)** (see also, protocols below). Alkaline phosphatase staining intensity was quantified by scoring colonies observed by brightfield microscopy in 10 random fields of view per well. Mouse PSC developmental capacity was assessed by differentiation in adherent 2D-culture +/- retinoic acid (protocol below). Differentiation in 3D culture was by embryoid body cardiac centre development, or matrigel-stimulated spheroid formation and lumenogenesis (protocol below). Human PSC developmental

capacity was assessed by differentiation toward embryoid bodies in vitro (protocol below) or by human-rabbit interspecies chimerism (protocol below).

Mouse PSC differentiation with retinoic acid

Differentiation of ES cells with retinoic acid (RA) was performed essentially as described [17]. LIF was first removed for 24hrs by culture in LIF-free Differentiation medium (that is DMEM (high glucose) supplemented with serum 15%, non-essential amino acids, glutamax and beta-mercaptoethanol; hereinafter referred as "differentiation medium"). Next, LIF-free differentiation media was supplemented with Retinoic Acid at 10 μM from +24 to +72 hrs, followed by LIF-free differentiation medium alone from +72 to +96 hrs. P19EC cell differentiation was by Retinoic Acid addition at 10 μM. Differentiation was also assessed by the same protocol of LIF-withdrawal except without adding Retinoic Acid.

Mouse PSC differentiation by hanging-drop and Embryoid Bodies

This was performed essentially as described [18]. ES cells were transferred to Differentiation medium (that is DMEM (high glucose) supplemented with serum 15%, non-essential amino acids, glutamax and beta-mercaptoethanol; hereinafter referred as "differentiation medium"), and suspended in hanging drop culture at a cell density of 1000-5000 cells/20 μLs. ES cells were allowed to form spherical aggregates known as Embryoid Bodies (EBs) for 48h in the hanging drops before transfer to suspension culture in low-adherence petri-dishes. In suspension culture, fresh Differentiation medium was added every 3 days, and the percent of EBs was scored daily for the development of beating cells in cardiac centres.

Mouse Chimera assays

Mouse ES cells labelled constitutively with Rosa26-GFP or Tg.CAG-Katushka [3] were treated for 10 passages in serum/LIF, 2i, or CDK8/19i conditions before testing in mouse chimera formation assays by morula aggregation at E2.5, or blastocyst micro-injection at E3.5, as described [19]. To assess incorporation and contribution of the ES cells to the host embryo epiblast, the extent of GFP+ or Katushka-red+ cells was assessed by confocal fluorescence. Embryos were cultured in KSOM to assess chimerism in E4.5 late blastocysts. Alternatively, micro-injected embryos were re-introduced into CD1 pseudo-pregnant females for implantation, and harvested at post-implantation timepoints: E6.5, E14.5, or E19.5 for direct fluorescence or immunohistochemistry analysis. Some chimeras were allowed to develop to adulthood to assess coat colour contribution and capacity for germline transmission.

Cardiac and endoderm directed differentiation of EBs derived from hPSCs.

The formation of embryoid bodies from hPSCs and their differentiation into endoderm and mesoderm (cardiac) derivatives was carried out as reported [20]. Briefly, hPSC colonies were dissociated using Accumax (Stem cell technologies), and cultured in suspension for 3 days with EB medium composed of DMEM/F12 supplemented with 15% fetal bovine serum, 2 mM L-glutamine, nonessential amino acids, and penicillin/streptomycin. For endoderm differentiation, EBs were plated on 0.1% gelatin coated plates and maintained with differentiation medium (DMEM supplemented with 20% fetal bovine serum, 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol, nonessential amino acids, and penicillin/streptomycin) for 2 weeks. For cardiac differentiation, EBs were transferred to 0.1% gelatin coated plates and differentiated in differentiation medium supplemented with 100 μM ascorbic acid (Sigma) for 2 weeks. A specific neural-directed differentiation

protocol was not performed, since EBs maintained in the same differentiation conditions used for endoderm or mesoderm also spontaneously gave rise to neural cell clusters.

Teratoma assays
Two injections of PSCs were performed per mouse. For mouse PSCs, 1 million cells in 100µl were injected sub-cutaneously in nude mice. For human PSCs, 2 million cells in 30µl were injected into the testis of male SCID beige mice.

Immunohistochemistry of embryoid body and teratoma samples.
Antigen retrieval was performed with citrate buffer (pH6) at 95ºC. Blocking for 1h with TBS + 1% Triton X-100 (Sigma) + 3% donkey serum (Millipore) at RT. Incubation with primary antibodies diluted in TBS + 0.5% Triton X-100 + 3% donkey serum. Primary antibodies are: Alpha-fetoprotein (Dako, A0008, 1:200, rabbit polyclonal, Lot# 20016625); FOXA2 (R&D Systems, AF2400, 1:50, goat polyclonal, Lot# ULB0414041); Troponin T, cardiac isoform (Thermo Scientific, MS-295-P1ABX, 1:200, mouse monoclonal, Clone 13-11, Lot# 295X1601A); GATA4 (Santa Cruz, sc-9053, 1:25; rabbit polyclonal, Lot# L1014); TUJ1 (Biolegend, MMS-435P, 1:500, mouse monoclonal, Clone TUJ1, Lot# D14JF02140); GFAP (Dako, Z0334, 1:1000; rabbit polyclonal, Lot# 20019134); Vimentin (Abcam, ab92547, 1:200, rabbit monoclonal, Clone EPR3776, Lot# GR219216-33); α-Smooth muscle actin (SMA) (Sigma, A5228, 1:400, mouse monoclonal, Clone 1A4, Lot# 074M4814V). Incubation with the appropriate Alexa Fluor 488- and Alexa Fluor 555- conjugated secondary antibodies (Fischer Scientific; all 1:200) for 2h at RT. Nuclei were counterstained with DAPI (Life Technologies, D1306; 1:5000). Confocal images were acquired using a SP5 Leica microscope.

Labeling of human PSCs for human/rabbit interspecies chimera assay
For transfection of human iPS cells, primed cells were cultured with ROCK inhibitor 24h before electroporation. Primed human PSC were collected and prepared as a unicellular suspension with accutase (Gibco). Cells were resuspended in Buffer R and electroporated (Neon Transfection System; Invitrogen; 1pulse/1400V/20ms) with 10 µg of DNA constructs for constitutive tdTomato expression (PB-Hygro-PGK-CAG-tdTomato 1752 and PBase pCMV-Transposase 1459). Cells were subsequently plated on matrigel in mTeSR1 medium supplemented with ROCK inhibitor for the first 24 h, then antibiotic selection with 20 µg/ml hygromycin was applied for 12 days, before a final step of cytometric sorting for 100% tdTomato constitutively-labelled cells.

Rabbit embryo production
Sexually mature NZW rabbits were purchased from HyPharm (Roussay, France). Female rabbits were superovulated as described previously [21]. Sixty hours after artificial insemination, fertilized embryos at the 8-cell stage (E1.5) were flushed from the explanted oviducts by using Euroflush® (IMV Technologies) and were 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 in 5% CO2.

Human PSC micro-injection and rabbit embryo development in vitro
For microinjection, human PSCs were dissociated into single cell suspensions with trypsin, and 5–10 cells were microinjected under the mucus coat and zona pellucida of morula-cell stage rabbit embryos, the day after collection. After microinjection, the embryos were sequentially cultured in the CDK8/19i media for 4 hours, followed by 20 hours incubation with a 1:1 mixture of RDH:CDK8i media and finally in RDH medium

for extended in vitro culturing. After 24 hours of in vitro culturing, early blastocyst stage embryos (E3.5) were rinsed 3 times in an embryo-holding medium (IMV Technologies) and were treated with 5 mg/ml protease E (Sigma) for 3 min at 37°C to digest the mucus coat and weaken the zona pellucida. The embryos were then rinsed 3 times in 199 HEPES medium (Sigma) and were cultured in the RDH medium for 3 days until they reached the late-blastocyst stage (E5.5).

Rabbit embryo immunofluorescence

Rabbit embryos were fixed in 2% paraformaldehyde for 20 minutes at room temperature, washed in PBS + 0.1% Tween-20, and permeabilized in PBS + 1% Tween-20 overnight at 4°C. Embryos were then incubated for 1 h at room temperature in PBS + 0.1% Tween-20 + 5% donkey serum and incubated with primary antibodies (anti-Oct4, SC-9081, Santa-Cruz; anti-GFP, A10262, Thermo Scientific) diluted in blocking solution at a concentration of 1:300 overnight at 4°C on a rotating shaker. After three washes of 5 min and one wash of 30 min in PBS + 0.1% Tween-20, embryos were then placed in secondary antibodies diluted in blocking solution at 1:300 for 1 hour at room temperature on a rotating shaker, transferred through several washes of PBS + 0.1% Tween-20 before stainning the nuclei with DAPI. Embryos were analyzed by confocal imaging (DM 6000 CS SP5; Leica). Acquisitions were performed using a water immersion objective (25Å~/1.25 0.75, PL APO HCX; Leica). Tiled scans were automatically acquired using LAS AF software (Leica).

Production of retrovirus and lentivirus, and infection of recipient cells

Briefly, retroviral and lentiviral supernatants were produced in HEK-293T cells ($5x10^6$ cells per 100 mm diameter dish). Vector transfections were performed using Fugene-6 transfection reagent (Roche) according to the manufacturer's protocol. Two days later, viral supernatants (10 ml) were collected serially during the subsequent 48 hours, at 12-hour intervals, each time adding fresh medium to the cells (10 ml). The recipient cells were seeded the previous day ($1.5x10^5$ cells per well in a 6-well plate) and each well received 1.0 ml of the corresponding retroviral and/or lentiviral supernatants as indicated in each Figure. This procedure was repeated every 12 hours for 2 days (a total of 4 additions).

For lentiviral shRNA production, per dish, 293T cells were transfected with 3 plasmids: (i) the ecotropic lentiviral envelope packaging plasmid pMD2.G (0.3 µg; Addgene, plasmid #12259; containing the VsVg gene); (ii) the lentiviral packaging plasmid pCMV-dR8.91 (3.0 µg); (from: Harvard Medical School, plasmid #516); (iii) plus one of the following lentiviral shRNA constructs (3.0 µg) expressing mouse shRNAs against CDK8, CDK19, Cyclin C, or the corresponding non-targeting control (Scramble, shSCR) vector. (see: **Resources Tables**, for sequences and plasmid details). After lentiviral infection was completed, lentiviral shRNA-knockdown cells were selected with puromycin (1 µg/ml). A panel of lentiviral shRNA against CDK8, CDK19 or CyclinC were tested for knockdown of their respective target. From these shRNAs, we identified that the best knockdown of CDK8, CDK19, or CyclinC expression by Western blot (see **Figure S1F**). See shRNA clone details in **Resource Tables.**

For retrovirus, per dish, 293T cells were transfected with the ecotropic packaging plasmid pCL-Eco (4 µg) together with one of the following retroviral constructs (4 µg): pMXs-Oct4, pMXs-Sox2, pMXs-Klf4, pMXs-cMyc, or pMXs-Nanog (obtained from Addgene and previously described [22] -the backbone is pMXs plasmid in all cases and the expression

of the coding sequences of the reprogramming factors are driven by the MMLV LTR promoter.

Generation of iPS cells from primary MEFs

For retroviral-mediated iPS reprogramming of primary (passage 2-5) mouse embryo fibroblasts was performed by a previous protocol [23]. Briefly, after infection of primary MEFs with retrovirus expressing the four Yamanaka transcription factors (OSKM), as outlined above, MEF media was replaced by KSR/LIF medium (see above). Cultures were maintained in the absence of drug selection with medium changes every 48 hrs [24]. Colonies were individually picked, and expanded clonally in the presence of 2i on feeders using standard procedures, before derivation onto 0.1% gelatin for feeder-free conditions. Status of iPS cells was confirmed by colony morphology, proliferation in 2i, Alkaline Phosphatase staining according to manufacturer's protocol (AP detection kit, Chemicon International, or, Promega #S3771); and qRT-PCR for multiple pluripotency markers including Nanog, Oct4, Tfcp2l1, Zfp42/Rex1 and Esrrb.

Generation of CDK8/19-double knockout iPS cells

To target mouse CDK19, we designed an sgRNA against CDK19 exon1, targeting 76 bp downstream of the ATG start of translation to generate indels (see: **Resources Tables**, for sgRNA sequences, and plasmid details) (see also, schematic: **Figure S1I**). Primary CDK8 flox/flox RERT-Cre MEFs of passage P+1 to P+4 were infected with lenti-CRISPR-Cas9 containing the CDK19 sgRNA (pLenti-CRISPRV2; Addgene #52961) followed by selection with puromycin (1μg/ml). CDK19-knockout was assessed by Western blot. The MEFs were immediately reprogrammed to iPS (see below), where single clones were picked, expanded and CRISPR-induced indels characterized by sequencing the CDK19 target region for frameshift mutations. Clones of iPS which were knockout for CDK19 were compared versus iPS clones which retained wild-type CDK19 expression, and no effect of CDK19-knockout was observed in MEFs or in iPS cells. CDK8-knockout was induced by 6 days of culture in the presence of 0.5 μM 4-hydroxy-tamoxifen to induce Cre-mediated deletion of CDK8 exon2 (see schematic: **Figure S1J**). CDK8 knockout was confirmed by allele-specific PCR (to demonstrate exon 2 deletion; see **Figure S1H**) and by western blot (to demonstrate complete loss of CDK8 protein; see **Figures S1J and S1K**).

Transcriptional CDK inhibitors

Structure and characterization of the CNIO CDK8/19 inhibitor (CDK8/19i-47799) are detailed in Table S1, Sheet#2. CDK8 inhibitors Senexin A (Tocris #4875) and Senexin B (Biocrick #BCC3990), as described[12]. CDK8 inhibitor CCT251545, as described [25]. CDK9 inhibitors #69 and #111, as described [26]. CDK7 inhibitors THZ-1 (Merck #532372) and BS-181 (Tocris #5608) were used as described [27,28].

Cytometry

FACS was performed as described [29]. Briefly, for SSEA1 or ICAM1 analysis, cells were collected by scraping and pipetting to unicellularize, before resuspension in 500 μLs 1xPBS and incubation with antibody conjugated to allophycocyanin (anti-SSEA1: R+D Systems, #FAB2155A; anti-ICAM1/CD54: eBiosciences ICAM-1-biotin, #13-0541) for 15 mins at room temperature. Data were analyzed with FlowJo 9.6.2 software. Live cell analysis of the *Nanog*-GFP used 2i-adapted mouse PSCs to define the threshold for the homogenous *Nanog*-GFPhigh population, against which other treatment were compared

(see **Figure 1A and 1B**). Live cell sorting for human PSC carrying HERVH-GFP selected the top 10% GFP-expressing cells, as previously described [9].

Cell lysis and Western blot

Whole cell extracts were prepared using 50 mM TrisHCl pH8; 1 mM EDTA; 150 mM NaCl; 1% NP40; 0.5% Triton X-100; 1.0% SDS, with freshly added protease inhibitors (Roche #11873580001). A total protein of 10 µg was loaded per lane and resolved on NuPAGE 4-12% gradient Bis-Tris gels, transferred to nitrocellulose and hybridized using antibodies as described in **Key Reagent Tables**.

Cell Fractionation

Nuclear and cytoplasmic fractionation was performed with the NE-PER kit (Thermofisher #78833).

Histopathology and immunohistochemistry

Mouse tissues were fixed in formalin at 4ºC, embedded in paraffin block, and sectioned at a thickness of 5 µm. Sections were stained with hematoxylin and eosin for pathological examination or processed for immunohistochemical analysis (for a list of the antibodies used, see **Resource Tables**). Embryos at E0.5, E1.5, E2.5, E3.0 morulae or E4.5 blastocyst embryos were flushed using M2 media. Embryos were then washed in KSOM media (Chemicon #3699) and either fixed immediately in 4% paraFormaldehyde for 10 minutes, or cultured in vitro in 70µl drops of KSOM +/- inhibitors under glycerol, as described [30] (and see below), before immunofluorescence.

Cell immunofluorescence

PSCs were grown on chamber slides using the same protocols as for the rest of the experiments. Cells were fixed with 4% paraformaldehyde for 2 minutes at room temperature, washed with PBS and permeabilized with PBS containing 0.02% Tween-20 for 20 minutes. Cells were blocked in PBS with 10% Australian FBS/1xPBS for 1h and incubated with antibodies (for a list of the antibodies used, see **Key Resource Tables**) at 1:200 to 1:1000 in PBS-4%BSA, for 3 hrs or overnight, washed with PBS and further incubated with secondary anti-rabbit antibodies conjugated with Alexa-488, Alex-555 and/or Alexa-647 (1:500 in PBS-4%BSA). Nuclei were counter-stained with DAPI. Confocal immunofluorescence cell images were captured using a Leica SP5, equipped with white light laser and hybrid detection.

DNA methylation

DNA was purified (Qiagen #69504) and global DNA methylation status was quantified by mass spectrometry, as described [31]. CpG methylation status at individual CpG sites of repeat DNA regions was assessed by DNA bisulphite-conversion and pyrosequencing, as described [32]. Bisulfite modification of DNA was performed with the EZ DNA methylation-gold kit (Zymo Research) following the manufacturer's instructions. The set of primers for PCR amplification and sequencing were designed using the specific software PyroMark assay design (version 2.0.01.15) (see primer tables above). After PCR amplification, pyrosequencing was performed using PyroMark Q24 reagents and a vacuum prep workstation, equipment, and software (Qiagen).

Image analysis

All image analysis was done using Fiji software [33] (http://fiji.sc).

Cloning
Mouse CDK8 D173A was cloned in to pMSCV-Puro-IRES-GFP (Addgene #21654), using BglII and HpaI restriction enzymes.
Mouse embryo manipulation and analysis

Embryo collection and fixation
CD1 females (4-5 weeks) were superovulated by intraperitoneal injection of 5 IU of pregnant mare serum gonadotropin (PMSG) followed 48h later by injection of 5IU of human chorionic gonadotropin (hCG). E0.5, E1.5 and E2.5 embryos were collected from the oviducts. For chimera contribution assays, E3.5 blastocysts were collected from the uterus by flushing with M2 culture medium. For immunofluorescence of CDK8 expression, E3.5 and E4.5 mouse embryos were flushed from the uterus of naturally mated MF1 females using M2 medium. E5.0, E5.5, and E6.5 mouse embryos were manually dissected from the decidual tissue of naturally mated females in M2 medium. Embryos were fixed in 4% PBS/paraformaldehyde (PFA) (15710, Electron Microscopy Sciences) for 20 min at RT and subsequently washed twice with 0.1% Tween/PBS. Permeabilization was in PBS 0.3% containing Triton X-100 and 0.1M glycine for 30 minutes at RT. Primary and secondary antibodies were diluted in blocking buffer (PBS containing 1% BSA and 0.1% Tween).

Pre-implantation mouse embryo culture
8-cell stage (E2.5) and early blastocyst stage (E3.5) mouse embryos were recovered from superovulated F1 females as described above. Superovulations were done by injecting F1 females with 7.5 IU of pregnant mares' serum gonadotropin (PMSG, Intervet), followed by injection of 7.5 IU of human chorionic gonadotropin (hCG, Intervet) and mating with F1 males. Embryos were cultured in KSOM medium (MR-020P-5F, Millipore) (control group), KSOM medium supplemented with MEK-inhibitor PD0325901 (Stem Cell Institute, Cambridge) or KSOM medium supplemented with CDK8/19i-LIF up to the late blastocyst stage (E4.5), in a drop culture covered with mineral oil (9305, Irvine Scientific) at 37°C in 21% O_2, 5% CO_2. Embryo allocation to control and experimental group was random. Fixation, permeabilization and immunostaining were performed as described above.

Pre- to post-implantation mouse embryo culture
To culture mouse embryos beyond implantation *in vitro* E4.5 mouse embryos were recovered from naturally mated MF1 females as described above. Embryos were cultured as previously described [34]. Briefly, the mural trophectoderm was manually removed using a finely pulled glass needle. Embryos were cultured in IbiTreat µ-plates (IB-80826, Ibidi GmbH) in IVC1 (control group) or IVC1 supplemented with CDK8/19 inhibitor and LIF (experimental group) for 24 hours at 37°C in 21% O_2, 5% CO_2. Embryo allocation to control and experimental group was random. IVC1 comprised: Advanced DMEM/F12 (12634010, ThermoFisher Scientific), 20% v/v heat-inactivated FBS (Stem Cell Institute, Cambridge), GlutaMAX (35050061, ThermoFisher Scientific), 25 units/ml Penicillin/25 µg/ml Streptomycin (15140122, ThermoFisher Scientific), 1X ITS-X (10 mg/ml insulin, 5.5 mg/L transferrin, 0.0067 mg/L sodium selenite, 2 mg/L etholamine) (51500056, ThermoFisher Scientific), 8 nM β-estradiol (E8875, Sigma), 200 ng/ml Progesterone (P0130, Sigma), and 25 µM N-aceyl-L-cysteine (A7250, Sigma).

Analysis of CDK8 levels across different developmental stages

To determine the expression of CDK8 at the single cell level across different stages in epiblast and primitive endoderm cells, a representative Z plane was selected. The Oct4 and Gata6 channels were binarized and two masks were created to segment the epiblast and primitive endoderm nuclei. For each nucleus a Regions of Interest (ROI) was generated, which was used to measure the nuclear levels of CDK8 in individual epiblast and primitive endoderm cells. To account for changes in fluorescence in the Z-axis and to determine the intercellular heterogeneity of CDK8 levels in each embryo, data was normalized to the average CDK8 expression per embryo (including both epiblast and primitive endoderm cells).

Analysis of CDK8 and Cyclin C levels in pre-implantation mouse embryos

To compare the levels of CDK8 Cyclin C in mouse embryos cultured in control conditions or in the presence of MEK-inhibitor PD0325901, a single representative Z plane was used to generate epiblast (Oct4+), primitive endoderm (Gata6+) and trophectoderm (Oct4- Gata6- DAPI+) masks, as described above. For the trophectoderm a single ROI was defined, whereas for the epiblast and primitive endoderm individual nucleus were saved as individual ROIs. These were used to measure the nuclear levels of CDK8 or Cycln C in individual epiblast and primitive endoderm cells. To account for changes in fluorescence in the Z-axis, the CDK8 fluorescence intensity in epiblast and primitive endoderm cells was normalized to the CDK8 or Cyclin C fluorescence intensity in trophectoderm cells.

Lumenogenesis by mouse PSC embryoid formation in matrigel

mESCs were cultured in a 3D matrix of matrigel to induce polarization and lumen formation as previously described [35,36]. Briefly, mESCs were trypsinized, centrifuged and washed twice with PBS to obtain a single cell suspension. 20,000 mESCs were centrifuged and the pellet was resuspended in 20 μL of ice-cold matrigel. The matrigel/cell suspension was placed as drop on the centre of a well of an IbiTreat μ-plates (IB-80826, Ibidi GmbH) and incubated at 37°C for 4 minutes to allow the matrigel to solidify. Next, the matrigel was covered with 300 μL of N2B27 or N2B27 with CDK8/19i +LIF, and cultures were fixed after 48 hours. Fixation, permeabilization and immunostainings were done as described above in this "mouse embryo manipulation and analysis" section.

Small molecule inhibitor characterization assays

Data from the small molecule inhibitor characterization assays is summarized in **Table S1**.

CDK8/Cyclin C and CDK9/CyclinT binding assays

The binding assay relies on the LanthaScreen ™ Eu-Kinase Binding Assay (Invitrogen). This is a kinase assay platform based on measuring the binding and displacement of an Alexa Fluor® 647 conjugate of an ATP-competitive kinase inhibitor (Kinase Tracer 236, PV5592) at a kinase active site. Binding of the tracer to the kinase is detected by addition of a europium (Eu)-labeled anti-GST antibody (Invitrogen PV 5594) for CDK8 or Eu-labeled anti-His antibody (Invitrogen PV5596) for CDK9, which specifically labels the kinase of interest. This binding results in a high degree of fluorescence resonance energy transfer (FRET), whereas displacement of the tracer with a kinase inhibitor results in a loss of FRET.

The enzymes were purchased from Invitrogen (CDK8/CycC: PV4402; CDK9/CycT: PV4131), as a dimer of full length GST or His-tagged recombinant human proteins,

respectively. Assay conditions were as indicated by the kit manufacturers. Assays were performed in 384-well plates. The final read out was generated using an EnVision plate reader (Perkin-Elmer). The emission ratio was calculated by dividing the acceptor/tracer emission (665 nm) by the antibody/donor emission (615 nm). Values were plotted against the inhibitor concentration and fit to a sigmoid dose-response curve using GraphPad software.

CDKs, DYRK1A, GSK3β and mTOR biochemical assays

The biochemical assay to measure kinase activity of CDKs, DYRK1A and GSK3β**,** relies on the LanthaScreen™ kinase activity assay (Invitrogen), where the kinase, a GFP-labeled substrate, and ATP are allowed to react. Then EDTA (to stop the reaction) and terbium-labeled antibody (to detect phosphorylated product) are added. In a LanthaScreen™ kinase reaction, the antibody associates with the phosphorylated GFP-labeled substrate resulting in an increased TR-FRET value. The TR-FRET value is a dimensionless number that is calculated as the ratio of the acceptor (GFP) signal to the donor (terbium) signal. The amount of antibody that is bound to the tracer is directly proportional to the amount of phosphorylated substrate present, and in this manner, kinase activity can be detected and measured by an increase in the TR-FRET value.

The enzymes, together with their partner proteins where necessary, were purchased from Invitrogen (CDK1/CycB: PV3292; CDK2/CycA: PV3267; CDK4/CycD1: PV4400; CDK5/p25: PV4676; CDK6/CycD1: PV4401; CDK7/CyclinH/MNAT1: PV3868), as a dimer of full-length His-tagged recombinant human proteins, or as full-length His-tagged recombinant human proteins (DYRK1A: PV3785; GSK3β: PV3365 and mTOR: PV4754), as well as the GFP-labelled substrate (4EBP1-GFP; PV4759) and the Tb-anti-p4EBP1 (phospho-Thr46) antibody (PV4757). Assay conditions were as indicated by the kit manufacturers. Assays were performed in 96-well plates. The final read out was generated using an EnVision plate reader (Perkin-Elmer). The TR-FRET value (a dimensionless number) was calculated as the ratio of the acceptor signal (GFP, emission at 520 nm) to the donor signal (terbium, emission at 495 nm). Values were plotted against the inhibitor concentration and fit to a sigmoid dose-response curve using GraphPad software.

PI3K, PIM1/2 and FLT3 biochemical assays

The kinase activity was measured by using the commercial ADP HunterTM Plus assay available from DiscoveRx (#33-016), which is an homogeneous assay to measure the accumulation of ADP, a universal product of kinase activity. Enzyme PI3K (p110α) was purchased from Carna Biosciences (#07CBS-0402A). PIM1 and PIM2 have been expressed and purified in-house as a recombinant human protein with a C-terminal histidine tag, following the manufacturer recommendations. FLT3 was purchased from Invitrogen (PV3182). Fluorescence counts were read in a Victor instrument (Perkin Elmer) with the recommended settings (544 and 580 nm as excitation and emission wavelengths, respectively). Values were normalized against the control activity included for each enzyme (eg: 100 % PI3K kinase activity, without compound). These values were plotted against the inhibitor concentration and fit to a sigmoid dose-response curve by using the GraphPad software.

KDR, KIT, PDGR-α and SRC biochemical assays

The biochemical assay to measure KDR, KIT, PDGR-α and SRC activities relies on the LANCE® technology (Perkin Elmer). This technology uses the europium-based chelate as a donor dye (narrow-banded emission at ~615 nm) and the acceptor dye ULight™,

which receives the energy from irradiated Eu chelate molecules in close proximity and in turn emits light at 665 nm. In the presence of kinase and ATP, the ULight-peptide substrate is phosphorylated. It is then captured by a Eu-anti-phospho-substrate antibody, which brings the Eu chelate donor and ULight acceptor dyes into close proximity. Upon excitation at 320 or 340 nm, the Eu chelate transfers its energy to the ULight dye, resulting in a fluorescent light emission at 665 nm. The enzymes were purchased from Invitrogen (KDR: PR5992C; c-KIT: P3081; PDGFR- α: PR7346A, SRC: PR4336E). Ulight-polyGT (TRF0110-D) and the Eu-W1024 anti-phosphoY66 antibody (AD0069) from Perkin Elmer. The assay was done following the manufacturer recommendations. The ratio between the acceptor signal (Ulight, emission at 615 nm) and the donor signal (Europium, emission at 665 nm) was calculated.. Values were plotted against the inhibitor concentration and fit to a sigmoid dose-response curve using GraphPad software.

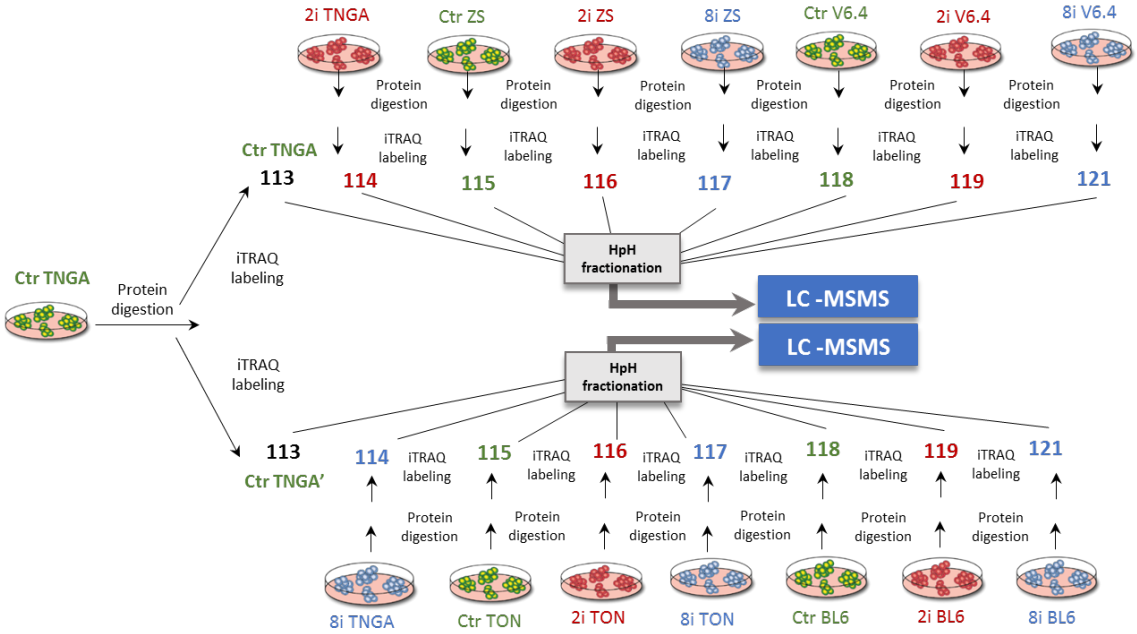
Proteomics: cell lysis and protein digestion

The experimental design consisted of five mouse ES cell lines: ZS, TNGA, TON, BL6 and V6.4. Cells were cultured in the presence of Serum/LIF (here referred as "control"), or additionally, with either 2i or CDK8/19i. Cell pellets were collected by trypsinization, washed with cold 1xPBS and preserved at -80ºC for further analysis. Cells were lysed using 7 M urea, 2 M thiourea, 50 mM Hepes, 1:1000 (v/v) of benzonase and 1:100 (v/v) of Halt™ phosphatase and protease inhibitor cocktail 100x. Cell lysates were homogenized by vortex plus sonication, and pre-cleared by centrifugation (20,000 g, 10 min, 4ºC). Protein concentration was measured with the Qubit® Protein Assay Kit. 110 µg of each lysate (except for Control TNGA for which 220 µg were used) were digested using the filter aided sample preparation (FASP) method (Wiśniewski et al., 2009). Samples were dissolved in 8M urea and 0.1M TEAB (UTEAB). Proteins were reduced in 15 mM TCEP for 30 min at room temperature with shaking (300 rpm) and alkylated in 50 mM of IAA for 20 min in darkness, with shaking at 450 rpm. Sample was cleaned twice with UTEAB. First digestion with endoproteinase Lys-C (1:50 w/w, Wako Pure Chemical Industries) was performed for 4 hours at room temperature in a wet chamber, followed by dilution 8-fold in 50 mM TEAB to reduce urea concentration. Second digestion with trypsin (1:100 w/w, Promega) was carried out overnight at 37ºC.

Proteomics: isobaric labelling with iTRAQ8plex

To quantify multiple samples using iTRAQ 8-plex, two multi-plex experiments were performed in parallel. Samples were organized in a labelling scheme such that each treatment (2i or CDK8/19i) was in the same analysis run as their control samples. TNGA control sample was added to both iTRAQ experiments. The complete labelling scheme is shown in Illustration 1, below.

Illustration 1: iTRAQ proteomic labelling scheme



For labelling, 110 µg of digested peptides were used for each channel using the iTRAQ® Reagent 8-plex kit (AB Sciex). The clean-up was performed with C18 Sep-Pack. Sample was loaded in 1% TFA, washed with 0.2% TFA and finally eluted with 1ml of 70% CH_3CN and 0.1% of TFA. Eluate was dried in vacuum and dissolved in 10 mM of NH_4OH for subsequent fractionation by high pH reversed phase chromatography.

Proteomics: high pH reverse phase fractionation

Each iTRAQ labelled sample was fractionated by high pH reverse phase chromatography. Peptides were dissolved in 100 µl of phase A (10 mM NH_4OH). Peptides were eluted at a flow rate of 500 µl/min onto a XBridge BEH130 C18 (3.5 µm, 4.6 x 250 mm) column (Waters) over 60 minutes, using the following gradient of phase B (10 mM NH_4OH, 90% CH_3CN): 0-50 min 25% B, 50-54 min 60% B and 54-61 min 70% B. Samples were collected every minute from minute 10 to minute 55, and concatenated into 15 fractions. Fractions were dried in vacuum and dissolved in 50 µl of 1% FA for subsequent LC-MS/MS analysis.

Proteomics: whole proteome LC-MS/MS

The Impact (Bruker Daltonics) was coupled online to a nanoLC Ultra system (Eksigent), equipped with a CaptiveSpray nanoelectrospray ion source supplemented with a CaptiveSpray nanoBooster operated at 0.2 bar/minute with isopropanol as dopant. 7.5 µl of each fraction were loaded onto a reversed-phase C18, 5 µm, 0.1 x 20 mm trapping column (NanoSeparations) and washed for 15 min at 2.5 µl/min with 0.1% FA. The peptides were eluted at a flow rate of 300 nl/min onto a home-made analytical column packed with ReproSil-Pur C18-AQ beads, 3 µm, 75 µm x 50 cm, heated to 45 °C. Solvent A was 4% ACN in 0.1% FA and Solvent B CH_3CN in 0.1% FA. The following gradient was used: 0–2 min 2% B, 2–119 min 2–20% B, 119–129 min 20-34% B, 129-140 min 98% B, 140–145 min 2% B. The MS acquisition time used for each sample was 145 min. The Q-q-TOF Impact was operated in a data dependent mode. The spray voltage was set to 1.35 kV (1868 nA) and the temperature of the source was set to 180°C. The MS survey scan was performed at a spectra rate of 2.5 Hz in the TOF analyzer scanning a window between 80 and 1600 m/z. The minimum MS signal for triggering MS/MS was set to a

normalized threshold of 500 counts. The 30 most abundant isotope patterns with charge ≥2 and m/z > 350 from the survey scan were sequentially isolated and fragmented in the collision cell by collision induced dissociation (CID) using a collision energy of 23 – 56 eV as function of the *m/z* value. The *m/z* values triggering MS/MS with a repeat count of 1 were put on an exclusion list for 30 s using the *rethinking* option: the precursor intensities were re-evaluated in the scan (n) regarding their values in the previous scan (n-1). Any *m/z* with intensity exceeding 5 times the measured value in the preceding survey scan was reconsidered for MS/MS. Data acquired were transformed to MGF format using the Compass DataAnalysis program. For each MS/MS spectra, the 400 most abundant non-deconvoluted ions exceeding a threshold of 100 counts were exported and recorded.

Proteomics: whole proteome data analysis

Raw files were analyzed using MaxQuant 1.5.3.30 [37] with Andromeda [38] as the search engine against a *Mus musculus* database (UniProtKB/Swiss-Prot, 43,539 sequences). Sample quantification type was set to iTRAQ8-plex. Carbamidomethylation of cysteine was included as fixed modification and oxidation of methionine, acetylation of protein N-terminal were included as variable modifications. Precursor mass tolerance was 35 ppm for the first search, and 7 ppm for the main search. Fragment mass tolerance was set to 40 ppm. Minimal peptide length was set to 7 amino acids and a maximum of two missed-cleavages were allowed. Peptides were filtered at 1% FDR. For protein assessment (FDR <1%) in MaxQuant, at least one unique peptide was required for identification. Other parameters were set as default. Proteins that did not have reporter intensity in all eight channels were discarded from further analysis.

In the full proteome, a total of 5,920 proteins were identified and quantified. Afterwards, the "protein-group" file was processed with R (v 3.3.0) and Perseus (v1.5.5.2) [39] for further statistical analysis. Reporter intensities were normalized using Loess function from Limma [40] package in R. Also, in order to integrate the two iTRAQ experiments, the common pool (Ctrl TNGA) was used to normalize between experiments. Statistically significant changes in protein expression between conditions (2i vs Control, or CDK8i vs Control) were determined using the Limma test implemented within Prostar [41] package in R, using a fold change threshold of 0.3 (in log2 scale) and a FDR of 5%.

Phospho-proteome: cell lysis, protein digestion and isobaric labelling

The experimental design consisted of two mouse ES cell lines: TON and ZS. Cells were cultured in the presence of Serum/LIF (here referred as "control"), or additionally, with either 2i or CDK8/19i. Inhibitor treatment of the cells was for precisely 15 minutes, after which, cells were collected rapidly by scraping in ice cold PBS, washed with ice-cold PBS, snap-frozen on dry ice, and preserved at -80ºC for further analysis. Cells pellets were lysed as described for the full proteome experiment above. 240 µg of 2i_TON, CDK8/19i_TON, 2i_ZS and 8i_ZS; and 480 µg of Control_TON or Control_ZS were digested using the filter aided sample preparation (FASP) method as described above. In the case of 480 µg of Control_TON and Control_ZS, samples were divided in two aliquots of 240 µg and digested separately and in parallel, to assess for technical variability that could arise due to sample preparation. The following labelling scheme was used: 113 Control_TON, 114 Control_TON, 117 Control_ZS, 118 Control_ZS; 115 15 min 2i_TON, 116 15 min CDK8/19i_TON, 119 15 min 2i_ZS, 121 15 min CDK8/19i_ZS. Two units of iTRAQ® Reagent 8-plex kit (AB Sciex) per sample were used following the manufacturer instructions. Sample clean-up was performed with C18 Sep-Pack.

Phospho-proteome: phosphopeptide enrichment and micro high pH reverse phase fractionation

To perform phosphopeptide enrichment, peptides were dissolved in 80% CH_3CN and 6% TFA. Titanium dioxide (TiO_2) beads were pre-conditioned once with 1ml of CH_3CN and twice with 500 µl of DHB solution (20 mg/ml DHB in 80% CH_3CN 6% TFA). Finally, beads were resuspended to a final concentration of 60 µg beads/µl of DHB solution. TiO_2 beads were added to the sample in a ratio 1:2 (Sample: TiO_2). Sample was incubated in rotation for 15 minutes and centrifuged 1 minute at 5000 g. Supernatant was used for a second TiO_2 binding with half the amount of TiO_2 beads. Next, beads from the first and second TiO_2 binding were transfer to separate C8-tips and washed with 100 ul pf 10% CH_3CN 6% TFA, 40% CH_3CN 6% TFA and 60% CH_3CN 6% TFA. Peptides were eluted with 25 µl of 5% NH_4OH and 25 µl of 10% NH_4OH 25% CH_3CN. Samples were dried in vacuum up to 5 µl.

Eluate from the second TiO_2 binding was resuspended in 22 µl 5% FA for subsequent LC-MS/MS analysis. Eluate from the first binding was fractionated with high pH reverse phase micro-columns. Briefly, 45 µl of phase A (20 mM NH4OH) was added to the sample obtained from the first TiO_2 incubation. 5 discs of C18 stage tip were used. Sample was loaded into the tips 3 times and the flow-through was collected to a vial. Next, 50 µl of phase A was loaded and collected in the same vial as the flow-through. Peptides were sequentially eluted increasing the percentage of Buffer B (20mM NH3 in CH3CN) (i.e. 4, 8, 12, 20, 60 and 80%). The last three fractions were pooled together. Samples were dissolved in 22 µl of 5% FA for subsequent LC-MS/MS analysis.

Phospho-proteome: LC-MS/MS

Each fraction were analysed twice (technical replicates) by LC-MS/MS using a 90 minute linear gradient of phase B (CH_3CN in 0.1% FA) using the same column settings and data acquisition method as described above.

Phospho-proteome: data analysis

Raw files were analyzed using MaxQuant 1.5.3.30 [37] with Andromeda [38] as the search engine against a *Mus musculus* database (UniProtKB/Swiss-Prot, 43,539 sequences). Sample quantification type was set to iTRAQ8-plex. Carbamidomethylation of cysteine was included as fixed modification and phosphorylation of serine, threonine and tyrosine, oxidation of methionine, acetylation of protein N-terminal were included as variable modifications. Precursor mass tolerance was 35 ppm for the first search, and 7 ppm for the main search. Fragment mass tolerance was set to 40 ppm. Minimal peptide length was set to 7 amino acids and a maximum of two missed-cleavages were allowed. Peptides were filtered at 1% FDR. For protein assessment (FDR <1%) in MaxQuant, at least one unique peptide was required for identification. Other parameters were set as default. Phosphorylation sites that were not quantified in all eight channels were discarded for further analysis. Afterwards, the "Phospho (STY) sites" file was processed with R (v 3.3.0) and Perseus (v1.5.5.2) [39] for further statistical analysis. Reporter intensities were normalized using Loess function from Limma package in R. Phosphorylation sites quantified with a median intensity below the intensity of the first quartile were tagged as 'noise'. Afterwards, ratios were calculated between 2i or CDK8/19i conditions versus Control conditions for each cell line and replicate. A phosphorylation site was considered up-regulated if at least 75% of the measured ratios were above a log2 fold change of 0.3. On the other hand, a phosphorylation site was considered down-regulated if 75% of the measured ratios were below a fold change of -0.3. The mass spectrometry proteomics

data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD009200.

Transcriptome: RNA isolation and quantitative real-time PCR (qRT-PCR)
Total RNA was extracted from cells on column by RNeasy kit with DNA digestion following provider's recommendations (Qiagen #74104, #79254), and retro-transcribed into cDNA following manufacturer´s protocol with Superscript Reverse Transcriptase (Life Technologies). Quantitative real time-PCR was performed using Syber Green Power PCR Master Mix (Applied Biosystems) in an ABI PRISM 7700 thermocycler (Applied Biosystem). Input normalization of all the qRT-PCR data was by the $2^{-\Delta\Delta Ct}$ method [42] using the housekeeping genes β-Actin or Gapdh as indicated in each Figure, and as described [43]. Primers used are in **Resource Tables**.

Transcriptome: RNA-seq transcriptomic analyses
For RNA-seq in mouse, samples of 1 μg of total RNA, with RIN numbers in the range 9.8 to 10 (Agilent 2100 Bioanalyzer), were used. PolyA+ fractions were processed using TruSeq Stranded mRNA Sample Preparation Kit (Agilent). Adapter-ligated library was completed by PCR with Illumina PE primers (8 cycles). The resulting directional cDNA libraries were sequenced for 40 bases in a single-read format (Genome Analyzer IIx, Illumina).

For RNA-seq in human, samples of total RNA with RIN numbers in the range 9.0 to 10 (Agilent 2100 Bioanalyzer), were used. For library construction 10 ng of total RNA samples were processed with the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech) by following manufacturer instructions. Resulting cDNA was sheared on a S220 Focused-ultrasonicator (Covaris) and subsequently processed with the ""NEBNext Ultra II DNA Library Prep Kit for Illumina"" (NEB #E7645). Briefly, oligo(dT)-primed reverse transcription was performed in presence of a template switching oligonucleotide, double stranded cDNA was produced by 11 cycles of PCR and submitted to acoustic shearing. Fragments were processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to Illumina adapters. Adapter-ligated libraries were completed by limited-cycle PCR (8 cycles). The resulting directional cDNA libraries were sequenced for 50 bases in a single-read format, instrument: Illumina HiSeq2500 Primary data processing: Image analysis, per-cycle base-calling and quality score assignment was performed with Illumina Real Time Analysis software. Conversion of Illumina BCL files to bam format was performed with the Illumina2bam tool (Wellcome Trust Sanger Institute - NPG).

The complete set of reads has been deposited in GEO (GSE112208). Sequencing quality for RNA-seq samples was analyzed with FastQC. Reads were aligned to the reference mouse genome (GRCm38/mm10) or the human genome (GRCh37/hg19) with TopHat-2.0.4 [44] (using Bowtie 0.12.7 [45] and Samtools 0.1.16 [46], allowing two mismatches and five multi-hits. Transcripts assembly, estimation of their abundance, and differential expression, were calculated with Cufflinks 1.3.0 [44], using the mouse genome annotation data set GRCm38/mm10, or the human genome annotation data set REF: GRCh37/hg19 from the UCSC Genome Browser [47]. When comparing samples, total read numbers were normalized, and visualized using SeqMiner 1.3.3e [48] or IGV (Integrated Genome Viewer) from the Broad Institute [49] available at: http://software.broadinstitute.org/software/igv/

Transcriptome: bioinformatics analysis of four human PSC lines by RNAseq

Reads were aligned to the hg19 human genome version using the STAR software[50] with default parameters. Number of reads per gene were calculated using the featureCounts function of the Rsubread[51] package of the R statistical software[52]. Gene anotations were performed using biomaRt [53] with the may2015.archive.ensembl.org version. Differential analysis was done with the R package DESeq2 [54] using the biological replicate as covariate.

Transcriptome: functional analyses of differential gene expression

For differential gene expression lists, see data in **Table S2** for mouse PSC cells adapted to Control, +2i, or +CDK8/19i. See also, data in **Table S3** for human PSC cells adapted to Control, +2i, or +CDK8/19i. Genes were ranked using the FDR q-value statistic to identify significant genes (FDR<0.05, as indicated in the Figures), then by fold change in expression. Selected differentially-expressed genes identified in the RNA-seq were validated by qPCR. Venn diagrams were generated by JVenn [55] and hypergeometric testing was performed to assess any significant overlaps. Gene Set Enrichment Analysis (GSEA) [56] with GSEA_Pre-ranked was used to perform a gene set enrichment analysis of annotations from the MsigDB Hallmarks, C5-Gene Ontology (GO) terms, C2-Curated, KEGG, Reactome and NCI databases, with standard GSEA and Leading Edge analysis settings. We used the RNA-seq gene list ranked by fold-change, setting 'gene set' as the permutation method and ran it with 1000 permutations for Kolmogorov-Smirnoff correction for multiple testing. We considered only those gene sets with significant enrichment levels (FDR q-value <0.25) [56]. GSEA Enrichment data were obtained and ranked according to their FDR q-value. Heatmaps of expression data were generated using Gene Pattern [57]. RRHO (Rank Rank Hypergeometric Overlap) was performed using the ranked list of Log_2_fold-change in gene expression or RNA Pol II abundance, using standard settings [58]. Colour intensity of RRHO heatmap indicates the -log10 p-value after Benjamini-Yekutieli correction of the hypergeometric overlap. RRHO available at: http://systems.crump.ucla.edu/rankrank/rankranksimple.php. Correlation matrix of ChIP-seq data in **Figure S6G** was produced using Morpheus software, available from the Broad Institute: https://software.broadinstitute.org/morpheus/. Analysis of Repeat sequences and Endogenous Retrovirus (ERV) expression was by using the Repbase datasets [59] for rodent or human repeat elements and FeatureCounts [51].

Differential gene expression comparing published mouse/human studies

Gene expression changes have been comprehensively characterized in several separate studies of mouse, primate, and human PSCs in response to over-expression of transcription factors, upon culture in various media cocktails, or in vivo, during development of the mouse or human embryos (see tables below and in: Table S2, Sheet#18; Table S3, Sheets#8 and #9). We used the marker genesets for each developmental stage, to perform GSEA on the ranked list of genes up/down in the cellular studies of mouse and human. We also performed the analysis in reverse, comparing the genesets of significantly differentially expressed mRNAs up- or down-regulated in our cell cells, versus, the complete ranked list of differential gene expression in other studies. GSEA results are shown in **Figure 3E** (mouse) **and Figure 3K** (human). The readout is the Normalized Enrichment score (NES). Data with P<0.05 and FDR q<0.05 are considered significant and marked with an asterisk (*) in the heatmaps of GSEA NES scores.

Table S2, Sheet#18 (re-printed here). Published mouse embryonic stem cells datasets used in this study for comparison.

Data set	Accession number	Processed data	PRIMED	NAÏVE
[14]	GSE56138	Table S1	1xEpiLC	1xESC
[60]	GSE23943	Table S2	3xSerum	3x2i
[61]	GSE81285	Table S2	2xSerum	2x2i
[62]	GSE81045	-	2xSerum/LIF	2x2i_48h
[62]	GSE81045	-	2xSerum/LIF	2ix2i_16d
[63]	E-MTAB-2600	http://www.ebi.ac.uk/teichmann-srv/espresso/	250x Serum/LIF	295x 2i and 159x a2i

Table S3, Sheet#8 (re-printed here). Published human embryonic stem cells datasets used in this study for comparison.

Data set	Accession number	Processed data	PRIMED	NAÏVE
[64]	GSE76970	TableS1	4xPrimed	4x5iLAF_SSE4neg
[64]	GSE76970	TableS1	4xPrimed	2x5iLAF_SSE4pos
[64]	GSE76970	TableS1	4xPrimed	2xUCLA20n
[65]	GSE87239	TableS2	2xPrimed	2xUCLA1_clone4
[65]	GSE87239	TableS2	2xPrimed	4xUCLA1_clone9_12
[66]	GSE60945	TableS3	3xH9	3xH9_reset
[67]	GSE59435	TableS1	2xhESM	5x6iLA_5iLA
[68]	E-MTAB-4461	TableS7	12xHES Primed	9xHNES
[69]	E-MTAB-2031	TableS3 (Takashima et al [66])	3xhESCs	3x3iL
[11]	-	TableS4 (Takashima et al [66])	1xH9,2xWIBR3	2xWIBR3,2xWIS1-2,1xH1, 1xH9

Table S3, Sheet#9 (re-printed here). Published human datasets of in vivo developmental stages used in this study for comparison.

Data set	Accession number	Processed data	DevStages
[70]	GSE44183	Available at GEO	Oocyte-Zygote-2cell-4cell-8cell-Morula
[71]	GSE36552	Table S1	Oocyte-Zygote-2cell-4cell-8cell-Morula-Blastocyst
[72]	E-MTAB-3929	Available at ArrayExpress	E3-E4-E5-E6-E7

Gene Set Enrichment Analysis (GSEA): *in vitro* comparisons

Ratios obtained from comparison of 2i vs Control, or CDK8/19i vs Control conditions in RNASeq in both mouse and human PSCs were used to perform the GSEA analysis [56,73]. Ratios were filtered to obtain a molecular signature of genes differentially up- or down-regulated in either 2i or CDK8/19i vs Serum/LIF as follows: log2 ratios larger than 1 and statistically significant (q.value < 0.05) were included in the molecular signature of up-

regulated genes in the corresponding condition. Ratios smaller than -1 and statistically significant (p.value < 0.05) were included in the molecular signature of down-regulated genes. These molecular signatures were used as gene sets to apply the GSEA algorithm against published datasets (**Table S2, Sheet#18: mouse; and Table S3, Sheets #8 and #9 -human**). In order to calculate all ratios, transcripts not measured in at least 75% of the samples of one condition were removed and data was normalized using 'normalizeCyclicLoess' function implemented in Limma [40]. Missing values were imputed from a normal distribution [39]. Ratios were calculated as the median of naïve (2i) conditions versus the median of primed conditions, and used as the ranked list input for the "GSEA_Preranked" analysis. The enrichment statistics used were 'classic', collapse data-sets was turned to FALSE, and maximum size for a gene set were increased up to 2500.

Gene Set Enrichment Analysis (GSEA): *in vivo* comparisons

Molecular signatures from developmental stages were obtained, when available, directly from published data, such as in mouse [74], or primate [75]. For those published data sets that did not provide a molecular signature for each developmental state, this was extracted as described [71]. Genes measured with low signal were discarded (FPKM or RPKM <0.1 in more than 75% of the measurements). Pairwise t-tests were performed against each developmental stage. Only transcripts which were significant (FDR q.value < 0.01) in at least one comparison with a fold change over 1 or below -1 (in log2 scale) were preserved. Expression levels were normalized by z-score and clustered using the 'sota'function in clValid R package into 25 categories or clusters. Finally, we chose the categories which comprised the most significant gene expression for each developmental stage.

Chromatin Immunoprecipitation (ChIP)

ChIP-qPCR was performed as described [29] with primers listed in **Resource Tables,** and antibodies for Total RNA Pol II (Santa Cruz N20, sc-899x) or Serine-5-phosphorylated Pol II (Abcam #5131). ChIP-seq for RNA Pol II was performed as described [76]. Briefly, cells were fixed using 1% formaldehyde, scrape-harvested, resuspended in ChIP lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.1) and sonicated using Covaris water bath sonicator to generate fragments of 150 to 500 bp. Soluble chromatin was diluted 10 fold in ChIP Dilution buffer (1% Triton X-100, 2 mM EDTA pH 8.0, 150 mM NaCl) precleared with Agarose Protein A/G beads (Santa Cruz), and then incubated with antibody specific for total RNA Pol II (N-20, sc-899x, Santa Cruz) or specific for the RNA Pol II Ser5P-phoshorylated form (Abcam #ab5131). After incubation, immunocomplexes were collected with Agarose Protein A/G beads (Santa Cruz). Next, the immunocomplexes were washed sequentially with Low Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 150mM NaCl), High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 500mM NaCl), LiCl Wash Buffer (0.25M LiCl, 1% NP40, 1% deoxycholate-Na, 1mM EDTA, 10mM Tris-HCl, pH 8.1) and washed twice with TE (10 mM Tris-HCl pH7.5, 1mM EDTA). Immunocomplexes were eluted in ChIP elution buffer (1%SDS, 0.1M NaHCO3) and the crosslinking was reverted by incubation at 65 ºC for 8 hrs with 200 mM NaCl. Samples were treated with Proteinase K and RNase A, and DNA was extracted using Phenol-Chloroform. DNA precipitation was in 100% ethanol with 0.1 M NaAcetate ph5.2 and 2 uLs glycogen (Roche). The DNA pellet was washed with 70% ethanol, and resuspended in ddH2O. Purified chromatin was used for library construction.

We performed 6 biological replicates for each condition (3: serum/LIF, 2i, CDK8/19i) and for each antibody (3: total RNA Pol II, S5P-RNA Pol II, control IgG). Three

replicates were used for ChIP-qPCR validations, and the other three replicates were pooled for sequencing. We note that our RNA Pol II ChIPseq data for serum/LIF and 2i-naïve cells very closely match previously published ChIPseq involving the same comparison, that is, mouse ES cells in primed *versus* 2i-naïve states[60].
For ChIP-seq the amount of DNA used was ~5 ng from each sample (as quantitated by fluorometry). Samples were processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters as in Illumina's "TruSeq DNA Sample Preparation Guide" (part # 15005180 Rev. C). Adapter-ligated libraries were completed by limited-cycle PCR with Q5 High-Fidelity DNA Polymerase (NEB) and Illumina PE primers (15 cycles), and further purified with a double-sided SPRI size selection to obtain a size distribution in the range of 230-500bp. Libraries were applied to an Illumina flow cell for cluster generation (TruSeq cluster generation kit v5) and sequenced on the Genome Analyzer IIx with SBS TruSeq v5 reagents by following manufacturer's protocols, to 20-25 million reads per replicate, to a total of >60 million reads per condition. The complete set of reads has been deposited in GEO (GSE112208).

RNA Pol II ChIP-seq data analyses
Definition of promoter and gene body regions (See: **Figures 4G and S5H**) and the calculation of RNA Pol II total and Ser5P abundance along genes was based on methods of Young and colleagues [76] (see abundance data in: **Table S6**). ChIP-seq data analysis was performed with the RUbioSeq pipeline (v3.8) [77], as follows: Sequencing quality for ChIP-seq samples was analyzed with FastQC (Andrews, 2011). Reads were aligned with Bwa 0.7.10 [78] to the mouse reference genome (NCBIm37/mm9) using the default parameters. SAMtools 0.1.19 [46] were used to convert the output alignment SAM files to the BAM file format and sort the alignments. Picard tools v1.107 (https://broadinstitute.github.io/picard/) were used to eliminate duplicated reads. Bedtools v2.16.2 were 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 MACS2 v2.1.1.20160309 [79] using the input sample as control for each one of the ChIP samples, and the distribution of peaks was plotted with SeqMiner 1.3.3e [80] with color-scaled intensities in units of reads per 50bp window, normalized per million mapped reads. Transcription Start Site (TSS) and Transcription Termination Zone (TTZ) were identified using the Database of Transcriptional Start Sites (http://dbtss.hgc.jp). Metagenes were aligned +/- 5 Kb or +/- 2 Kb around the TSS, and visualized by SeqMiner in **Figures 4E, S5E, S5G and S6D**, where genes were listed in order of RNA Pol II abundance in the promoter region of the control serum/LIF condition (see ranked list in **Table S6, Sheet#11**). The promoter, gene body and transcription termination zone (TTZ), and the ratios between these three regions for each gene (see schematic in **Figures 4G and S5H, Table S6**, Sheet#1), were defined similar to previous descriptions [76,81,82]. Total and Ser5P RNA Pol II abundance were quantified at promoter, gene body and TTZ for 31,167 Refseq gene loci where the transcription start and stop sites are known (**Table S6**, sheets #2 and #3) in four steps, similar to previous reports [76]. (i) the number of reads per nucleotide was computed with BEDTools 'genomecov'; (ii), to extend this number to the number of reads per gene promoter or gene body, BEDTools 'map' was used; (iii), to correct for region size, the RNA Pol II abundance was calculated as: ((number of reads in region / region size)*scaling factor)*10^5. Scaling factor = (total number of reads in sample/genome length). (iv) For the analysis of Pol II abundance according to inhibitor treatment, genes were first filtered for high confidence Pol II detected at threshold of >3,000 units at the promoter, and detected in all three conditions (Serum/LIF, 2i or CDK8/19i), yielding 12,072 genes (see: **Table S6**, sheet#4 for full filter

calculations). In **Figures 4E, S5E, S5G and S6D**, genes were arranged in rank by the abundance of RNA Pol II in the promoter region in the control serum/LIF condition.

CDK8/19 ChIP-Seq and definition of enhancers, target genes, and eRNA levels

For **Figures S6D to S6G and S7A to S7C**, CDK8/19 enrichment across the genome of wild-type mixed background V6.5 (C57BL/6-129) mouse ES cells was determined using the published dataset: GSE44286, GSM1082346, as previously described [83,84], with peak calling by MACS v1.4.1 [79] using standard settings and compared to the input negative control. Note, the ChIP antibody for this ChIP-seq (Santa Cruz #sc-1521) is reported to bind to both CDK8 and CDK19 [2,85]. Peak annotation within local genomic features for **Figures 5A, S6E and S6F**, was done using HOMER and the enhancer regions previously defined as constituent regions of typical enhancers (n= 9,981) or super-enhancers (n=646) [84], and of super-enhancer extended regions (n=231) as defined [83], where enhancers were defined by co-enrichment for Oct4, Sox2, Nanog, and Med1. For peak calls, CDK8/19 abundance at called peaks, and local annotation see **Table S7**.

For **Figures S7A to S7C**, in order to identify the single-nearest target gene to each PSC super-enhancer and analyze their biological functions, GREAT analysis was performed as described (GREAT v3.0.0; [86] using the list of CDK8/19 peaks identified above (**Table S7**).

For **Figures 4G to 4M**, RNA Pol II abundance was assessed by normalizing the total number of reads between treatments, and using FeatureCounts [51] to calculate the background-subtracted Log2_RPKM of RNA Pol II abundance in the indicated regions. In **Figures 5B and S7D**, the regions were defined according to the intensity of CDK8/19 MACS peaks within regions annotated as intergenic by HOMER, as described above. In **Figure 5C**, the regions were defined as above, that is, the enhancer regions previously defined as constituent regions of typical enhancers (n= 9,981) or super-enhancers (n=646) [84]. In **Figures S7E and S7F**, the regions were defined by filtering the PREStige database of enhancers [87,88] which identifies enhancers by enrichment of H3K4 me1methylation in multiple tissues and lineages. Using the PREStige data, we identified enhancer regions with H3K4me1 enrichment >20 units, and specific only to pre-implantation naïve ES cells, or post-implantation EpiSC cells, versus all other tissue-specific enhancer regions listed in the database, by subtraction of overlapping enhancers (1bp overhang threshold) as outlined in the schematic of Figure S7E. See **Table S7** for the list identified for the set of naïve ES-specific enhancers (n=1,424), or EpiSC-specific enhancers (n=1,005).

For **Figures 5D and S7G**, enhancer RNA (eRNA) levels were quantified by qRT-PCR using primers previously described as indicative of activity in naïve ES-specific super-enhancers [1], and these primer sequences are listed in **Resources Tables**. In the plots, the Mean and SEM of three independent experiments are displayed. In the heatmap, the values of the three independent experiments are shown at each timepoint.

For **Figure 5E**, GSEA was run with a geneset of the single nearest genes to super-enhancers (as identified by GREAT analysis above; (GREAT v3.0.0; [86]), using the super-enhancers previously described in mouse ES cells [84], versus, the ranked list of differential gene expression determined by RNA-Seq for serum/LIF control compared to CDK8/19i-adapted mouse ES cells.

For **Figure S7H**, GREAT analysis (GREAT v3.0.0; [86]) was used to identify the set of single nearest genes (n=3,553 genes) to all the ES cell enhancer regions previously

identified in mouse ES cells (n=10,627), where enhancers were defined by co-enrichment for Oct4, Sox2, Nanog and Med1 [83,84]. Then the of Log_2_fold-change in RNA expression of these genes was ranked high-to-low, using the RNAseq in mouse for control vs 2i, or control vs CDK8/19i, where the control was the serum/LIF condition. The extent of hypergeometric overlap of these two ranked lists is shown as a heatmap in Figure S7H, and was performed by RRHO (Rank Rank Hypergeometric Overlap) using standard settings [58], available at: http://systems.crump.ucla.edu/rankrank/rankranksimple.php. Colour intensity of RRHO heatmap indicates the -log10 p-value after Benjamini-Yekutieli correction of the hypergeometric overlap.

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise specified quantitative data are presented as mean +/- SD and significance was assessed by the two-tailed Student's t test; *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. Rank Rank Hypergeometric Overlap (RRHO) was performed as described [58] using standard settings and after Benjamini-Yekutieli correction of the hypergeometric overlap. For differential gene expression by RNA-Seq, a threshold of FDR q-value of q<0.05, or q<0.01 was applied, as indicated in each case. In GSEA analysis, the standard threshold for significance was applied, where p <0.05 and FDR q-value < 0.25. Immunofluorescent image analysis is described in detail above in section on embryo analysis.

DATA AND SOFTWARE AVAILABILITY

Data Resources. Accession Numbers

Five datasets (four RNA-seq and one ChIP-seq experiment) are available from the GEO database: GSE112208. The mass spectrometry proteomics data are available from the ProteomeXchange Consortium/PRIDE repository with the dataset identifier PXD009200.

SUPPLEMENTAL REFERENCES

1. Respuela, P. *et al*. Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification. *Cell Stem Cell* **18**, 118–133 (2016).

2. Broude, E. V *et al*. Expression of CDK8 and CDK8-interacting Genes as Potential Biomarkers in Breast Cancer. *Curr. Cancer Drug Targets* **15**, 739–49 (2015).

3. Diéguez-Hurtado, R. *et al*. A Cre-reporter transgenic mouse expressing the far-red fluorescent protein Katushka. *genesis* **49**, 36–45 (2011).

4. Chambers, I. *et al*. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).

5. Macfarlan, T. S. *et al*. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 1–10 (2012). doi:10.1038/nature11244

6. Zalzman, M. *et al*. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).

7. Ying, Q.-L. *et al*. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).

8. Palmero, I. & Serrano, M. Induction of senescence by oncogenic Ras. *Methods Enzymol.* **333**, 247–56 (2001).

9. Wang, J. *et al*. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409 (2014).

10. Chen, H. *et al*. Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *Nat. Commun.* **6**, 7095 (2015).

11. Gafni, O. *et al*. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).

12. Porter, D. C. *et al*. Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities. *Proc. Natl. Acad. Sci.* **109**, 13799–13804 (2012).

13. Pease, S. & Saunders, T. *Advanced Protocols for Animal Transgenesis*. (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-20792-1

14. Buecker, C. *et al*. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).

15. Guo, G. *et al*. Klf4 reverts developmentally programmed restriction of ground state pluripotency. **1069**, 1063–1069 (2009).

16. Tosolini, M. & Jouneau, A. From Naive to Primed Pluripotency : In Vitro Conversion of Mouse Embryonic Stem Cells in Epiblast Stem Cells. (2015). doi:10.1007/7651

17. Savatier, P., Lapillonne, H., van Grunsven, L. A., Rudkin, B. B. & Samarut, J. Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene* **12**, 309–22 (1996).

18. Marikawa, Y., Tamashiro, D. A. A., Fujita, T. C. & Alarcón, V. B. Aggregated P19 mouse embryonal carcinoma cells as a simple in vitro model to study the molecular regulations of mesoderm formation and axial elongation morphogenesis. *Genesis* **47**, 93–106 (2009).

19. Abad, M. *et al*. Reprogramming in vivo produces teratomas and iPS cells with

totipotency features. *Nature* **502**, 340–345 (2013).

20. Montserrat, N. *et al*. Reprogramming of human fibroblasts to pluripotency with lineage specifiers. *Cell Stem Cell* **13**, 341–350 (2013).

21. Salvetti, P. *et al*. Effect of the luteinizing hormone on embryo production in superovulated rabbit does. *Theriogenology* **67**, 1185–93 (2007).

22. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).

23. Li, H. *et al*. P27Kip1 directly represses Sox2 during embryonic stem cell differentiation. *Cell Stem Cell* **11**, 845–852 (2012).

24. Li, H. *et al*. The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).

25. 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).

26. Németh, G. *et al*. Novel, selective CDK9 inhibitors for the treatment of HIV infection. *Curr. Med. Chem.* **18**, 342–58 (2011).

27. Kwiatkowski, N. *et al*. Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).

28. Ali, S. *et al*. The development of a selective cyclin-dependent kinase inhibitor which demonstrates anti-tumor activity. *Cancer Res.* **69**, 6208–6215 (2009).

29. 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).

30. 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).

31. Fernandez, A. F., Valledor, L., Vallejo, F., Cañal, M. J. & Fraga, M. F. Quantification of Global DNA Methylation Levels by Mass Spectrometry. in *Methods in molecular biology (Clifton, N.J.)* **1708**, 49–58 (2018).

32. Tajuddin, S. M. *et al*. LINE-1 methylation in leukocyte DNA, interaction with phosphatidylethanolamine N-methyltransferase variants and bladder cancer risk. *Br. J. Cancer* **110**, 2123–2130 (2014).

33. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–5 (2012).

34. Bedzhov, I., Leung, C. Y., Bialecka, M. & Zernicka-Goetz, M. In vitro culture of mouse blastocysts beyond the implantation stages. *Nat. Protoc.* **9**, 2732–9 (2014).

35. Shahbazi, M. N. *et al*. Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* (2017). doi:10.1038/nature24675

36. Bedzhov, I. & Zernicka-Goetz, M. Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* **156**, 1032–1044 (2014).

37. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

38. Cox, J. *et al*. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–805 (2011).

39. Tyanova, S. *et al*. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

40. Ritchie, M. E. *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43**, e47–e47 (2015).

41. Wieczorek, S. *et al*. DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **33**, 135–136 (2017).

42. Yuan, J. S., Reed, A., Chen, F. & Stewart, C. N. Statistical analysis of real-time PCR data. *BMC Bioinformatics* **7**, 85 (2006).

43. Ortega-molina, A., Lopez-guadamillas, E., Cabo, R. De & Serrano, M. Pharmacological Inhibition of PI3K Reduces Adiposity and Metabolic Syndrome in Obese Mice and Rhesus Monkeys Article Pharmacological Inhibition of PI3K Reduces Adiposity and Metabolic Syndrome in Obese Mice and Rhesus Monkeys. 1–13 (2015). doi:10.1016/j.cmet.2015.02.017

44. Trapnell, C. *et al*. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc*. **7**, 562–578 (2012).

45. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10**, R25 (2009).

46. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

47. Rosenbloom, K. R. *et al*. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res*. **43**, D670–D681 (2015).

48. Ye, T. *et al*. seqMINER: An integrated ChIP-seq data interpretation platform. *Nucleic Acids Res*. **39**, 1–10 (2011).

49. Robinson, J. T. *et al*. Integrative genomics viewer. *Nat. Biotechnol*. **29**, 24–26 (2011).

50. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

51. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–30 (2014).

52. Team, R. D. C. *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing, Vienna, Austria*. (2008).

53. Durinck, S. *et al*. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–40 (2005).

54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550 (2014).

55. Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293 (2014).

56. Subramanian, A. *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A*. **102**, 15545–50 (2005).

57. Reich, M. *et al*. GenePattern 2.0. *Nat. Genet*. **38**, 500–1 (2006).

58. 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).

59. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

60. Marks, H. *et al*. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).

61. Bulut-Karslioglu, A. *et al*. Inhibition of mTOR induces a paused pluripotent state. *Nature* **540**, 119–123 (2016).

62. Fidalgo, M. *et al*. Zfp281 Coordinates Opposing Functions of Tet1 and Tet2 in Pluripotent States. *Cell Stem Cell* **19**, 355–369 (2016).

63. Kolodziejczyk, A. A. *et al*. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **17**, 471–485 (2015).

64. 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).

65. Sahakyan, A. *et al*. Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* **20**, (2017).

66. Takashima, Y. *et al*. Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* **158**, 1254–1269 (2014).

67. 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).

68. Guo, G. *et al*. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports* **6**, 437–446 (2016).

69. 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).

70. Xue, Z. *et al*. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).

71. 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).

72. Petropoulos, S. *et al*. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012–1026 (2016).

73. Mootha, V. K. *et al*. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–73 (2003).

74. Boroviak, T. *et al*. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev. Cell* **35**, 366–382 (2015).

75. Nakamura, T. *et al*. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016).

76. Rahl, P. B. *et al*. C-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).

77. Rubio-Camarillo, M. *et al*. RUbioSeq+: A multiplatform application that executes parallelized pipelines to analyse next-generation sequencing data. *Comput. Methods Programs Biomed.* **138**, 73–81 (2017).

78. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

79. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).

80. Ye, T., Ravens, S., Krebs, A. R. & Tora, L. Interpreting and visualizing ChIP-seq data with the seqMINER software. *Methods Mol. Biol.* **1150**, 141–52 (2014).

81. Chen, F. X. *et al*. PAF1, a molecular regulator of promoter-proximal pausing by RNA Polymerase II. *Cell* **162**, 1003–1015 (2015).

82. Booth, G. T., Wang, I. X., Cheung, V. G. & Lis, J. T. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Res.* **26**, 799–811 (2016).

83. Whyte, W. A. *et al*. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

84. Hnisz, D. *et al*. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

85. Tsutsui, T., Fukasawa, R., Tanaka, A., Hirose, Y. & Ohkuma, Y. Identification of target genes for the CDK subunits of the Mediator complex. *Genes to Cells* **16**, 1208–1218 (2011).

86. McLean, C. Y. *et al*. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

87. 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).

88. Corradin, O. *et al*. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).