Integrated multi-omics reveal polycomb repressive complex 2 restricts human trophoblast induction

Dick W. Zijlmans1,13, Irene Talon2,13, Sigrid Verhelst3,13, Adam Bendall4,13, Karlien Van Nerum2, Alok Javali5, Andrew A. Malcolm6, Sam S. F. A. van Knippenberg2, Laura Biggins7, San Kit To2, Adrian Janiszewski2, Danielle Admiraal8, Ruth Knops9, Nikky Corthout10, Bradley P. Balaton2, Grigorios Georgolopoulos2, Amitesh Panda2, Natarajan V. Bhanu11, Amanda J. Collier4, Charlene Fabian4, Ryan N. Allsop2, Joel Chappell2, Thi Xuan Ai Pham2, Michael Oberhuemer2, Cankat Ertekin2, Lotte Vanheer2, Paraskevi Athanasouli2, Frederic Lluis2, Dieter Deforce3, Joop H. Jansen9, Benjamin A. Garcia11, Michiel Vermeulen1,8, Nicolas Rivron5, Maarten Dhaenens2,3,13, Hendrik Marks8,14,12,14,13, Peter J. Rugg-Gunn4,6,12,14,13 and Vincent Pasque2,1,2,14

Human naive pluripotent stem cells have unrestricted lineage potential. Underpinning this property, naive cells are thought to lack chromatin-based lineage barriers. However, this assumption has not been tested. Here we define the chromatin-associated proteome, histone post-translational modifications and transcriptome of human naive and primed pluripotent stem cells. Our integrated analysis reveals differences in the relative abundance and activities of distinct chromatin modules. We identify a strong enrichment of polycomb repressive complex 2 (PRC2)-associated H3K27me3 in the chromatin of naive pluripotent stem cells and H3K27me3 enrichment at promoters of lineage-determining genes, including trophoblast regulators. PRC2 activity acts as a chromatin barrier restricting the differentiation of naive cells towards the trophoblast lineage, whereas inhibition of PRC2 promotes trophoblast-fate induction and cavity formation in human blastoids. Together, our results establish that human naive pluripotent stem cells are not epigenetically unrestricted, but instead possess chromatin mechanisms that oppose the induction of alternative cell fates.

Epiblast and trophoblast cells of the human embryo display a prolonged period of developmental plasticity. Contrary to the mouse blastocyst, where the epiblast and trophoblast lineages are restricted, these lineages are not yet committed in the human blastocyst14–17. This unrestricted lineage potential of cells of early human blastocysts is retained in naive human pluripotent stem cells (hPSCs), derived from pre-implantation blastocysts, which have the potential to differentiate into both embryonic and extra-embryonic cell types including the trophoblast lineage18,19. The developmental plasticity of naive hPSCs also endows them with the capacity to form blastoids, which are generated from naive hPSCs that self-organize into structures resembling blastocysts15–19. In contrast, primed hPSCs share properties with postimplantation epiblast cells and differentiate into trophoblast cells less efficiently. Hence, they are not suitable to generate blastoids13,17–21.

Trophoblast cells rarely arise spontaneously in robust naive hPSC cultures but they can be converted from this state using trophoblast stem cell culture conditions6,11–14. This suggests that the trophoblast fate is actively suppressed in naive hPSCs and is activated in response to appropriate cues and in a regulated manner. Considering the important role of chromatin-based processes in regulating cell identity, this raises the possibility that epigenetic barriers could exist to regulate the transition from naive pluripotency towards the trophoblast lineage. Defining these barriers would shed light on developmental mechanisms regulating developmental transitions and lead to better control of trophoblast specification and differentiation.

---

1Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences (RIMLS), Oncode Institute, Radboud University Nijmegen, Nijmegen, The Netherlands. 2Department of Development and Regeneration, Leuven Stem Cell Institute, Leuven Institute for Single Cell Omics (LISCO), KU Leuven, Leuven, Belgium. 3ProGenToMics, Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium. 4Epigenetics Programme, Babraham Institute, Cambridge, UK. 5Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna, Austria. 6The Welcome–MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. 7Bioinformatics Group, Babraham Institute, Cambridge, UK. 8Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences (RIMLS), Radboud University Nijmegen, Nijmegen, The Netherlands. 9Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre (RadboudUMC), Nijmegen, the Netherlands. 10VIB Center for Brain and Disease Research, KU Leuven, VIB Bioimaging Core, Leuven, Belgium. 11Epigenetics Institute, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. 12The Centre for Trophoblast Research, University of Cambridge, Cambridge, UK. 13These authors contributed equally: Dick W. Zijlmans, Irene Talon, Sigrid Verhelst, Adam Bendall. 14These authors jointly supervised this work: Maarten Dhaenens, Hendrik Marks, Peter J. Rugg-Gunn, Vincent Pasque.

E-mail: Maarten.Dhaenens@UGent.be; Hendrik.Marks@ru.nl; peter.rugg-gunn@babraham.ac.uk; vincent.pasque@kuleuven.be

858
Chromatin and epigenetic-based processes are key regulators of cell identity, fate specification and developmental gene expression programmes\(^{22-24}\). Striking differences in the transcriptome, DNA methylome and genome organization have been uncovered between naive and primed hPSC states, which correspond to their distinct developmental identities\(^{21-29}\). A limited number of other chromatin-based epigenetic properties have also been examined in naive and primed hPSCs, including histone H3 lysine 27 trimethylation (H3K27me3), which is a histone modification catalysed by polycomb repressive complex 2 (PRC2) and is associated with transcriptional repression\(^{23}\). H3K27me3 levels differ between human pluripotent states, although it remains unclear whether global levels of H3K27me3 are higher in primed hPSCs compared with naive hPSCs\(^{31}\), or the opposite\(^{42}\). Genome mapping by chromatin immunoprecipitation with sequencing showed that a greater number of gene promoters are marked by H3K27me3 in primed hPSCs compared with naive hPSCs\(^{32,33}\). It therefore remains enigmatic which chromatin-associated proteins and histone post-translational modifications (hPTMs) characterize and regulate the unrestricted lineage potential of naive hPSCs.

Naive hPSCs can be maintained in the absence of epigenetic repressors, including PRC2, DNMT1 and METTL3, whereas these factors are required for stable self-renewal and maintaining the pluripotent status of primed hPSCs\(^{21,24-26}\). Based on these observations, naive hPSCs are considered ‘epigenetically unrestricted’. However, because the role of chromatin-based mechanisms in controlling the transcriptome, epigenome and differentiation potential of naive hPSCs has not been examined, whether these mechanisms establish a lineage barrier in human cell pluripotency and control fate specification remains an important unresolved question.

Here we apply an integrated multi-omics approach to comprehensively map the chromatin-associated proteome, hPTMs and transcriptome of naive and primed hPSCs. We unexpectedly discovered that PRC2 activity opposes the induction of trophoblast in naive hPSCs and blastoids, thereby establishing that naive pluripotent cells are not epigenetically unrestricted but that instead, chromatin barriers limit their ability to differentiate into trophoblast.

Results

Comprehensive chromatin profiling in hPSCs. To define the chromatin landscapes of hPSCs, we performed an integrated multi-omics analysis of naive (cultured in PXGL medium) and primed (cultured in E8 medium) H9 human embryonic stem cells (hESCs; Fig. 1a). This analysis incorporated chromatin proteomes, DNA methylation levels, hPTMs and transcriptomes, thus including chromatin regulatory factors as well as their modifications and transcriptional outcomes. For convenient access to these data, we created a searchable tool to explore the data (https://www.bioinformatics.babraham.ac.uk/shiny/shiny_omics/Shiny_omics). Transcriptional analyses validated the anticipated expression of pluripotent-state markers (Extended Data Fig. 1a and Supplementary Tables 1,2).

To identify chromatin regulators associated with both pluripotent states, we analysed chromatin-bound proteins using chromatin enrichment for proteomics (ChEP), followed by mass spectrometry\(^{37,38}\). We identified 4,576 proteins, of which 1,819 changed significantly between the naive and primed states (\(P<0.05\), fold change (FC) > 2; Fig. 1b and Supplementary Table 3). Gene ontology analysis of the chromatin-bound proteins that were more abundant in primed hPSCs showed an association with development and neuronal differentiation (Fig. 1c), in agreement with the more advanced developmental stage of primed hPSCs\(^{39}\). Gene ontology terms associated with proteins that were more abundant in naive hPSCs included transcriptional regulation and RNA processing (Fig. 1c).

We next analysed prominent proteins involved in pluripotency, DNA methylation and chromatin remodelling (Fig. 1d and Extended Data Fig. 1b). In naive hPSCs, we identified an increase in the chromatin-associated levels of known naive factors (KLFL4, KLFL5 KLFL17, TFCP2L1, PRDM14 and TFPAP2C) in addition to unanticipated factors (UTF1, DPPA2, LIN28B and MYCN)\(^{35,36}\). In primed hPSCs, transcription factors including ZIC2, ZIC5, LIN28A and LITD1 were more abundant compared with naive hPSCs\(^{37}\). Shared proteins included core pluripotency factors (POU5F1, SALL4 and SOX2) and chromatin remodelers (BRD3, BRD4 and SMARCC2; Fig. 1d and Extended Data Fig. 1b,c).

We confirmed that naive hPSCs were globally DNA hypomethylated compared with primed hPSCs (Extended Data Fig. 1d), corroborating previous findings\(^{26,42}\). Despite this difference in DNA methylation, our ChEP analysis showed that there was little to no change in the chromatin-bound levels of the DNA methyltransferases DNMT3A and DNMT3B when naive and primed hPSCs were compared (Fig. 1d). However, we detected a decrease in DNMT1 and its known interactor UHRF1 (ref. 48) as well as an increase in TET1 in naive hPSCs, which are differences that could potentially reinforce the hypomethylated state of naive hPSCs.

We detected changes between members of several chromatin regulatory complexes between naive and primed hPSCs, including PRC1, PRC2, NuRD and histone deacetylase complexes (Fig. 1e and Extended Data Fig. 1c). For PRC2, we noticed a modest increase in core components in naive cells as well as increased MTF2 and decreased JARID2, which suggests a shift in PRC2 subcomplexes from PRC2.2 to PRC2.1. Finally, we found changes in ATP-dependent chromatin remodelling complexes (Fig. 1d and Extended Data Fig. 1c). Notably, we detected higher levels of SMARCA2 (also known as BRM) in the naive state and SMARCA1 (also known as SNF2L) in the primed state, in line with the OCT4-specific association of these factors to regulate chromatin accessibility\(^{49,50}\).

Together, this analysis identified a compendium of chromatin-associated proteins in naive and primed pluripotent states, including widespread differences in DNA-binding factors as well as in the writers, readers and erasers of hPTMs, highlighting the distinct chromatin landscapes of human pluripotent states.

hPTMs of pluripotent cells. Although hPTMs are pivotal mediators of chromatin structure and function, they have mainly been studied in hPSCs using targeted sequencing-based approaches\(^{50,51,52,53,54}\). We performed a bottom-up mass spectrometry analysis following acid extraction to assay the global abundance of hPTMs and histone variants. This approach quantified 43 individual hPTMs on histones H3 and H4 (Fig. 2a,b and Supplementary Table 4), of which 23 were significantly different between the two cell states (\(P<0.05\)). There was a strong increase in PRC2-mediated H3K27me2 and H3K27me3, and DOT1L-mediated H3K79me2 in naive cells compared with primed cells. Modications that were lower in naive hPSCs included H3K27 acetylation and H3K36me2, which is consistent with the antagonism of these modifications with H3K27me2 and H3K27me3 (refs. 40-48), as well as a global decrease in H4-tail acetylation. Histone variants also affect chromatin states (Extended Data Fig. 2a). Particularly in naive hPSCs, the histone H1 and H2A repertoire shifts significantly, with a prominent increase in H1.1 and H2A1B/H1, and a decrease in the H2A variants H2AW and H2AY (macroH2A). Extending these findings, we performed hPTM profiling in the same H9 hPSC line but using alternative naive cell culture conditions (t2iLg0 medium)\(^{22}\) and mass spectrometry protocols, and furthermore compared the results with previously profiled H9 hPSCs cultured in ENHSM medium\(^{32}\) (Fig. 2c). Overall, the hPTM patterns were similar for the three culture conditions, further validating the differences between naive and primed states.

Histone profiling additionally identifies alkaline proteins that are co-purified, referred to as the acid extractome, which contains many nucleic-acid binders (Supplementary Table 5),\(^{32}\). There was a good correlation between the abundance of proteins detected in both the chromatin proteomes and acid extractomes (Extended...
Naive and primed hPSCs contain distinct chromatin proteomes. a, Schematic of the workflow used for multi-omics profiling of naive and primed hPSCs. Major classes of chromatin regulators and protein complexes, indicated in the top-left corner, are highlighted. Complex members and regulators with significantly changed levels of expression (two-sided Student’s t-test, \(P < 0.05\), FC > 2) are labelled; \(n\) = 3 biologically independent samples for each cell type. Horizontal dashed lines represent FC > 2 (two-sided Student’s t-test).

The acid extractome adds insights by identifying proteins that were not detected in the chromatin proteome; for example, the WNT signalling regulator APC2 is increased in naive hPSCs (Extended Data Fig. 2c). In addition, the acid extractome showed higher levels of ribosomal and nucleolar proteins in naive compared with primed hPSCs (Extended Data Fig. 2d), in line with the enrichment of the gene ontology term ‘RNA processing’ observed in the transcriptomics data (Fig. 1c). Related to this, MMP-2 activates ribosome biogenesis by enzymatic clipping of the histone H3 amino (N)-terminal tail following binding of the ribosomal–RNA gene promoter\(^*_2\), which can initiate ribosome synthesis in preparation of large cellular transitions such as between naive and primed...
Fig. 2 | Profiling of hPTMs reveals decoupling of chromatin-modifier activity and abundance when comparing naive and primed pluripotency. a, b. Levels of H3 (a) and H4 (b) hPTMs in naive and primed hPSCs. The relative abundance of each hPTM as a percentage of the total for the histone residue (for example, the relative abundances of H3K79me1, H3K79me2 and H3K79 unmodified all add up to 100%) is provided (left). Unmodified histones are only shown for residues with >1 modification. Data are presented as the mean ± s.d. Change in hPTMs between naive and primed hPSCs as log2-transformed values (right). The red bars indicate significantly changed hPTMs (two-sided Student’s t-test with Benjamini–Hochberg correction, \( P < 0.05 \)).

c. Comparison of hPTMs in naive and primed H9 hPSCs; the naive hPSCs were (naive hPSCs) and 5 (primed hPSCs) biologically independent samples.

d. Integration of the chromatin proteome and hPTM measurements for naive and primed hPSCs, separated by histone H3 and H4 modules. Nodes represent chromatin modifiers and hPTMs, and are coloured according to the log2-transformed abundance FC. Edges indicate known functional connections (write or erase) between the nodes. Highlighted hubs indicate major hPTM groups. Chromatin modifiers in grey nodes were not detected.

e. Comparison of the human and mouse chromatin proteomes of naive relative to primed pluripotent states. Only proteins identified in both human and mouse datasets were retained. Proteins with \( P < 0.05 \) were deemed as significantly changed (two-sided Student’s t-test). Red dashed lines indicate FC = 2.

f. Comparison of the human and mouse hPTMs in naive relative to primed pluripotent states. Only hPTMs identified in both human and mouse datasets were retained. The blue line indicates the best-fit linear regression; the shaded grey area indicates the 95% confidence interval. Mouse hPTM data were obtained from98. Source data are provided.
DNMT3A and DNMT3B are strongly enriched on chromatin in opposite trend in mouse pluripotent cells (Fig. 2e). In addition, which were strongly enriched in human naive cells but showed the strongly enriched on the chromatin of naive human cells compared proteins showed an opposite trend between mouse and human. JARDI2. However, despite these similarities, several prominent and PRC2 core and sub-complex members, such as MTF2 and TFCP2L1, that occur at higher levels in the naive state (Fig. 2e) and mouse. This includes transcription factors, such as KLF4 and that many naive and primed factors are similar between human GATA3 genes.

Fig. 3] H3K27me3 localization, as determined by cCUT&RUN in naive and primed hPSCs. a. Kernel density estimate of H3K27me3 cCUT&RUN reads in naive and primed hPSCs after normalization to the Dro sophila spike-in. The genome was divided into 1-kb bins, the number of H3K27me3 reads in each bin was quantitated and the log2-transformed value of the counts was calculated; n = 2 biologically independent experiments for all samples (primed and naive H3K27me3 and IgG cCUT&RUN) except naive IgG, which is from n = 1 experiment. b. Normalized H3K27me3 reads mapped at repetitive element classes in the human genome as a percentage of the total sequenced reads for naive and primed hPSCs. SINE and LINE, short and long interspersed nuclear elements, respectively; LTR, long terminal repeat. c. Heatmap of normalized H3K27me3 (left) and IgG (right) cCUT&RUN read counts within a 10-kb peak-centred window in naive and primed hPSCs. Regions were subsetted into primed-enriched (n = 5,086 regions; top), common (n = 7,851 regions; middle) and naive-enriched (n = 6,308 regions; bottom) sites. d. Metaplots showing average profiles of normalized H3K27me3 counts across peaks, with relative abundance and distribution within 25 kb either side of the peak centre for primed-specific (middle), shared (right) and naive-specific (left) peaks. e. Percentage of normalized H3K27me3 reads within defined peaks for naive and primed hPSCs. f. Normalized H3K27me3 (top) and IgG (bottom) cCUT&RUN genome browser tracks over naive-specific (DUSP6 and SFRP2; left) and primed-specific (KLFL4 and TFCP2L1; right) H3K27me3-marked genes. g. Normalized H3K27me3 (top) and IgG (bottom) cCUT&RUN genome browser tracks for exemplar trophoblast (CDX2, GATA3, GATA2, KRT8 and KRT18; top), primitive endoderm (GATA4, GATA6, PDGFRα and FOXA2; middle) and additional alternative lineage marker genes (HAND1, PAX6 and SOX17; bottom) in naive and primed hPSCs. Regions with P < 0.05 after Benjamini–Hochberg multiple-testing correction were identified as differentially enriched. Source data are provided.

In conclusion, naive and primed hPSCs have distinct chromatin landscapes with specific transcription factors as well as their own and shared chromatin complexes. Each state has its own unique hPTM signature, with naive hPSCs containing more H3K27me3 overall compared with primed cells. Surprisingly, the hPTM signature of each pluripotent state does not always directly correlate with the protein abundances of their writers and erasers on chromatin.

Conserved and species-specific chromatin features. To compare the chromatin-based properties of mouse and human pluripotent states, we integrated our dataset with a previous study of mouse PSCs38. Global analysis of chromatin-bound proteomes revealed that many naive and primed factors are similar between human and mouse. This includes transcription factors, such as KLF4 and TFCP2L1, that occur at higher levels in the naive state (Fig. 2e) and PRC2 core and sub-complex members, such as MTF2 and JARDI2. However, despite these similarities, several prominent proteins showed an opposite trend between mouse and human. Notable exceptions include KLF4, TFAP2C and DPPA2, which were strongly enriched on the chromatin of naive human cells compared with primed cells but not in mouse cells (Fig. 2e), which for TFAP2C is consistent with its human-specific role in early development46. LIN28B is mainly present in the chromatin of naive PSCs in humans, whereas in mouse it is associated with primed pluripotency8. Other striking differences included UTF1 and DNMT3L, which were strongly enriched in human naive cells but showed the opposite trend in mouse pluripotent cells (Fig. 2e). In addition, DNMT3A and DNMT3B are strongly enriched on chromatin in primed mouse cells but were detected on chromatin at similar levels in naive and primed human cells (Fig. 2e). We also identified proteins that were detected uniquely in the ChEP proteomes of hPSCs but not in mouse ChEP proteomes, which might therefore have human-specific roles in pluripotent cells (Extended Data Fig. 2f).

Finally, hPTM patterns are largely conserved between human and mouse pluripotent states in naive and primed cells (Fig. 2f), as is H3K27 clipping (Extended Data Fig. 2e)36,50.

H3K27me3 marks lineage-determining genes in the naive state. As H3K27me3 is associated with the control of gene regulation and cell identity and showed the largest difference between human pluripotent states, we investigated the genome-wide distribution of this chromatin mark in naive and primed hPSCs. We adapted the cleavage under targets and release using nuclease (CUT&RUN) method47,48 by incorporating calibrated spike-in normalization (cCUT&RUN) to enable quantitative comparisons (Extended Data Fig. 3a,b and Supplementary Table 6). Consistent with our mass spectrometry results, cCUT&RUN confirmed there was a higher global level of H3K27me3 in naive compared with primed hPSCs (Fig. 3a). Several repetitive element classes also had higher levels of H3K27me3 in naive hPSCs (Fig. 3b), potentially also contributing to the pluripotent state-specific differences in H3K27me3 levels.

Contrary to the global trend, however, peak-based analysis revealed stronger and more focused regions of H3K27me3 enrichment in primed hPSCs compared with naive hPSCs (Fig. 3c,d and Extended Data Fig. 3c,d). Furthermore, a threefold-higher proportion of cCUT&RUN reads were within peaks in primed cells (Fig. 3c). We detected elevated levels of H3K27me3 in the regions surrounding peaks, providing further evidence that H3K27me3 coats the genome of naive hPSCs (Fig. 3d). These results show that although primed hPSCs have lower global H3K27me3 signal, the cCUT&RUN reads are more concentrated within defined and narrower peak regions.

As expected, a large proportion of the peaks in either cell type were near promoters and this proportion was higher in primed hPSCs (36%) compared with naive hPSCs (21%; Extended Data Fig. 3e). The reduced number of promoters marked by H3K27me3 in naive hPSCs is consistent with observations in naive hPSCs cultured in other media conditions49,50 and similar to observations in mice46. However, based on our quantitative profiling, the number of H3K27me3-marked promoters in naive hPSCs is substantially higher than previously reported26,27. Differential analysis categorized peaks into regions enriched for H3K27me3 in either naive or primed hPSCs. We found that a subset of primed-enriched peaks marked naive-specific genes, including KLF4 and TFCP2L1 (Fig. 3f and Extended Data Fig. 3f). In addition, many primed-specific peaks were marked in both cell types but accumulated more H3K27me3 in the primed state (Fig. 3c), suggesting that regions marked by H3K27me3 in primed hPSCs are often already established in the naive state. Conversely, the naive-enriched regions were largely...
devoid of H3K27me3 in primed cells (Fig. 3c,f and Extended Data Fig. 3f). The gain and loss of H3K27me3 correspond to transcriptional differences between pluripotent states (Extended Data Fig. 3g) and, overall, the presence of H3K27me3 at naive-specific genes was associated with reduced expression levels compared with primed cells (Extended Data Fig. 3h).
Many of the genes marked by H3K27me3 were shared between naive and primed cells (n = 2,384 genes; Extended Data Fig. 3f and Supplementary Table 7). Importantly, this category contained genes associated with embryonic- and extra-embryonic-lineage specification, which were unexpectedly marked by H3K27me3 in naive hPSCs as well as primed hPSCs (Fig. 3g). This gene set included germ-layer determinants including PA6X; primitive endoderm factors, such as PDXG1, GATA4 and GATA6; and trophoblast regulators such as CDX2, GATA3, GATA2, KRT7 and KRT18 (Fig. 3g). The unexpected presence of H3K27me3 at the promoters of key lineage regulators in naive hPSCs raises the possibility that PRC2-mediated H3K27me3 might oppose cell-fate specification in naive hPSCs. Several of the trophoblast factors marked by H3K27me3 in naive hPSCs are expressed at high levels in trophoderm cells of human blastocysts and their enforced expression induces the trophoblast cell fate\textsuperscript{34,35}. Consequently, because naive hPSCs have the capacity to produce trophoblasts in vitro\textsuperscript{11,14}, we sought to use trophoblast differentiation as a cell model to investigate a potential role for H3K27me3 in controlling lineage induction in human naive pluripotency.

**PRC2 activity opposes the induction of trophoblast fate.** To test the hypothesis that PRC2 activity in naive hPSCs restricts the induction of the trophoblast lineage, we established conditions that could rapidly deplete PRC2-mediated H3K27me3 in naive hPSCs, thereby limiting secondary effects or cell culture adaptations. Application of the PRC2 inhibitor UNC1999 (ref. \textsuperscript{36}) in PXGL culture medium for 4 d robustly and reversibly depleted global H3K27me3 levels (Fig. 4a and Extended Data Fig. 4a–d) and removed H3K27me3 from gene promoters (Fig. 4b,c and Extended Data Fig. 4e). We observed minimal disruption to the chromatin-bound proteome (Fig. 4d and Extended Data Fig. 4a–d) excepting naive IgG, which is from PRC2i without PRC2i. Regions were subsetted into primed-enriched (n = 6,308) and naive-enriched (n = 5,806 regions) sites; n = 2 biologically independent experiments for all samples (primed and naive H3K27me3 and IgG cCUT&RUNs, both with and without PRC2i) excepting naive IgG, which is from n = 1 experiment. Samples without PRC2i treatment are reproduced from Fig. 3. e, Genome browser tracks show normalized H3K27me3 and IgG cCUT&RUN reads for trophoblast genes (CDX2, GATA2, GATA3, and KRT7 and KRT18) in naive and primed hPSCs with and without PRC2i. Regions were subsetted into primed-enriched (n = 8,096 regions), common (n = 7,985 regions) and naive-enriched (n = 6,308 regions) sites; n = 2 biologically independent experiments for all samples (primed and naive H3K27me3 and IgG cCUT&RUNs, both with and without PRC2i) excepting naive IgG, which is from n = 1 experiment. Samples without PRC2i treatment are reproduced from Fig. 3. e, Genome browser tracks show normalized H3K27me3 and IgG cCUT&RUN reads for trophoblast genes (CDX2, GATA2, GATA3, and KRT7 and KRT18) in naive and primed hPSCs with and without PRC2i. Principal component (PC) analysis of the chromatin proteome (left), hPMT landscape (middle) and transcriptome (right) of naive and primed hPSCs with and without PRC2i (n = 3 biologically independent samples for chromatin proteome and transcriptome); f, Gene expression levels, determined through RNA-seq analysis, of trophoblast-associated genes (GIF2, SOCS3, SATB1, SATB2 and SOX21) in naive hPSCs with and without PRC2i (n = 3 biologically independent samples). Data are presented as the mean ± s.d. \textsuperscript{f.g.} Differential gene expression in naive (f) and primed (g) hPSCs with and without PRC2i (n = 3 biologically independent samples). Genes enriched in the untreated condition are highlighted in red and those enriched after PRC2i are highlighted in blue; the number of differentially expressed genes in both conditions are indicated. Dashed lines indicate P < 0.05 and log\textsubscript{2}(FC) > 1 (two-sided Student’s t-test). Source data are provided.

![Fig. 4 | Histone, chromatin and transcriptional responses following short-term acute PRC2 inhibition. a, Levels of H3 hPMTs in naive and primed hPSCs with and without PRC2 activity inhibition for 4 d (PRC2i). Data are presented as the log\textsubscript{2}-transformed FC between the two conditions indicated above each panel. Data are ordered according to the left panel. The red bars indicate significantly changed hPMTs (two-sided Student’s t-test with Benjamini–Hochberg correction, P < 0.05; n = 7 naive hPSCs, 5 primed hPSCs, 6 naive hPSCs + PRC2i) and 8 (primed hPSCs + PRC2i) biologically independent samples. b, Heatmaps of normalized H3K27me3 cCUT&RUN read counts within a 10-kb peak-centred window in naive and primed hPSCs with and without PRC2i. Regions were subsetted into primed-enriched (n = 5,086 regions), common (n = 7,851 regions) and naive-enriched (n = 6,308 regions) sites; n = 2 biologically independent experiments for all samples (primed and naive H3K27me3 and IgG cCUT&RUNs, both with and without PRC2i) excepting naive IgG, which is from n = 1 experiment. Samples without PRC2i treatment are reproduced from Fig. 3. c, Genome browser tracks show normalized H3K27me3 and IgG cCUT&RUN reads for trophoblast genes (CDX2, GATA2, GATA3, and KRT7 and KRT18) in naive and primed hPSCs with and without PRC2i. d, Principal component (PC) analysis of the chromatin proteome (left), hPMT landscape (middle) and transcriptome (right) of naive and primed hPSCs with and without PRC2i (n = 3 biologically independent samples for chromatin proteome and transcriptome); e, Gene expression levels, determined through RNA-seq analysis, of trophoblast-associated genes (GIF2, SOCS3, SATB1, SATB2 and SOX21) in naive hPSCs with and without PRC2i (n = 3 biologically independent samples). Data are presented as the mean ± s.d. \textsuperscript{f.g.} Differential gene expression in naive (f) and primed (g) hPSCs with and without PRC2i (n = 3 biologically independent samples). Genes enriched in the untreated condition are highlighted in red and those enriched after PRC2i are highlighted in blue; the number of differentially expressed genes in both conditions are indicated. Dashed lines indicate P < 0.05 and log\textsubscript{2}(FC) > 1 (two-sided Student’s t-test). Source data are provided.](naturecellbiology)
Fig. 6a–c). Cells in the C1 ‘intermediate’ cell cluster had reduced expression levels of pluripotency genes and increased expression of KRT18, TFP2C (Fig. 5g), ARID3A and EPCAM (Extended Data Fig. 6b,c). Most of the day 4 conversion samples, both with and without PRC2-inhibitor treatment (74 and 88% of cells, respectively), contributed to the intermediate cluster (C1; Extended
Data Fig. 6d and Supplementary Table 9). The C2 cells showed a strong decrease in expression of pluripotency genes and the activation of multiple trophoblast markers^{5,6} (Fig. 5g,h and Extended Data Fig. 6b,c). We termed cells in the C2 cluster as trophoblast cells because they aligned to human embryo trophoderm and early trophoblast (Fig. 6a–d and Extended Data Fig. 6e,f). Importantly, PRC2 inhibition promoted the acquisition of trophoblast fate, as the proportion of cells in the C2 trophoblast cluster was more than doubled following PRC2 inhibition (26% for the PRC2-inhibited cells versus 11% for the control cells; Fig. 6e and Supplementary Table 9). Together, these results show that PRC2 inhibition promotes naive pluripotency exit and increases trophoblast-fate induction.

**PRC2 inhibition enhances trophoblast formation in blastoids.** To further investigate a role for PRC2 in human trophoblast specification and morphogenesis, we used human blastoids as a three-dimensional blastocyst-like model^{7} (Fig. 7a). We tested whether PRC2 inhibition affects trophoblast specification and epithelial morphogenesis, processes that are necessary to form a blastocoeel-like cavity. We inhibited PRC2 in naive hPSCs for 4 d before blastoid formation and assessed the effect of PRC2 inhibition on trophoblast-fate induction by measuring the proportion of trophoblast cells in human blastoids using the pan-trophoblast markers TROP2 and GATA3. PRC2 inhibition increased the proportion of TROP2⁺ and GATA3⁺ cells in blastoids at 36 h and 60 h (Fig. 7b,c and Extended Data Fig. 7a–d). The increase in trophoblast induction was accompanied by a decrease in the ratio of NANOG⁺ epiblast-like cells (Fig. 7c and Extended Data Fig. 7a–d). Although there were very few primitive endoderm-like cells (FOXA2⁺) at 60 h, there was a trend towards increased primitive endoderm induction following PRC2 inhibition (Extended Data Fig. 7c).

During blastocyst development the trophoblast forms an epithelium that pumps water directionally to generate the blastocoeel cavity. PRC2 inhibition accelerated cavity formation by about 24 h (Fig. 7d–f and Supplementary Table 10). Multiple cavities seemed to progressively coalesce, possibly through the action of aquaporin 3, the water transporter most highly expressed in human blastocysts^{8}. This effect of the PRC2 inhibitor was no longer observed at 60 h, as blastoids had already reached their maximum size. During human blastocyst and blastoid development aquaporin 3 is initially expressed in all cells and then becomes restricted to trophoblast cells^{9,10}. This restriction seemed to occur more rapidly with PRC2 inhibition (Extended Data Fig. 7e). These observations are consistent with an acceleration of trophoblast-fate induction, as observed in the earlier monolayer experiments (Fig. 5). The results show that PRC2 inhibition promotes the formation of functional trophoblast with some trophoderm-like morphogenetic properties. Together, we conclude that PRC2 functions as a barrier to trophoblast formation in naive hPSCs (Fig. 7g).

**Discussion**

Human naive PSCs and epiblast cells have the potential to generate trophoblast with high efficiency in response to inductive cues^{6,11–14}. The molecular properties that enable this highly regulated developmental plasticity, however, have not been comprehensively defined. Here we have demonstrated that repressive chromatin pathways oppose trophoblast induction in naive hPSCs. We showed that PRC2-mediated H3K27me3 marks trophoblast regulators in naive hPSCs, including genes that are expressed in trophoderm cells of human blastocysts and can promote trophoblast fate^{13,15}. By establishing that PRC2 is a lineage gatekeeper stabilizing the undifferentiated naive state, these findings overturn the current assumption that naive hPSCs are epigenetically unrestricted. Protection of trophoblast genes against low-level or inappropriate transcriptional activation signals is anticipated to support robust growth of undifferentiated naive hPSCs while maintaining their broad developmental potential. Sustained exposure to strong trophoblast-inductive signals overcomes these repressive mechanisms and efficient trophoblast differentiation is initiated. By uncovering a role for this pathway in opposing trophoblast induction and finding that in naive hPSCs H3K27me3 also marks key regulators of additional cell types, such as primed pluripotency and primitive endoderm, our work lays the foundation for future studies to determine whether PRC2 could also control the specification of alternative lineages from naive pluripotency.

We have also shown that human naive and primed pluripotent cells have striking differences in the relative abundance and activities of chromatin-associated proteins. Integrating these data sets enabled a systems-level view of the chromatin proteomes and revealed that state-specific differences in the abundance of chromatin modifiers and their associated histone marks are not always concordant. These findings highlight the importance of regulating protein activities in addition to protein abundance in changing chromatin states in pluripotency and thereby raise caution in using methods like transcriptional or proteomic profiling to predict differences in chromatin states between cell types. Our work also resolved discrepancies in the literature. Different methodological approaches have previously reported conflicting results on the global level of H3K27me3 in naive hPSCs^{16,17}. Our findings establish that global H3K27me3 levels are increased in naive compared with primed hPSCs and in multiple culture conditions, which is consistent with a previous report in hPSCs and with mouse pluripotent states^{18,19}. These findings are in line with a possible need for high polycomb-protein activity in cells, such as naive hPSCs, to retain low levels of global DNA methylation^{20}. More generally, because the changes in the relative abundance of most histone modifications were similar when human and mouse pluripotent states were compared, our results also suggest a general conservation of the histone code between human and mouse in these cell types. This raises the prospect of applying histone profiling to define mammalian PSC states. However, despite these general similarities, we also uncovered
species-specific differences, particularly at the level of chromatin mediators. For example, the DNA methylation machinery seems to operate differently in human and mouse naive PSCs, with major state-specific differences in chromatin association of DNMT3A, DNMT3B and DNMT3L, which show opposite trends when human and mouse cells are compared. Curiously, the catalytically
inactive protein DNMT3L is strongly upregulated in naive hPSCs and its role in global hypomethylation is not intuitive as it might be expected to boost de novo methyltransferase activity by stimulating DNMT3A and DNMT3B. However, knockdown of DNMT3L during primed-to-naive hPSC resetting does not affect the levels of DNA methylation, and it is possible that DNMT3L might have roles in human naive pluripotency that are methylation-independent. It is of particular interest to establish whether DNMT3L might recruit chromatin-modifying repressor proteins to silence transposable elements and other target regions, as has been recently reported in mouse PSCs and fibroblasts.

The state-specific global differences in H3K27me3 prompted us to examine this modification in further detail. Using a quantitative CUT&RUN assay, we found that the levels of H3K27me3 in the genome of naive hPSCs were substantially higher than previously shown, corroborating the global H3K27me3 quantification of our hPTM profiling. Importantly, the promoters of developmental regulators of multiple lineages are marked by H3K27me3 in naive hPSCs, thereby uncovering a more prominent role for H3K27me3 in these cells than recognized thus far. This finding builds on our recent study showing that H3K27me3 tends to co-occur in naive hPSCs with the active histone marks H3K4me3 and H3K4me1, which is a signature of bivalent chromatin. Whether these regions in naive hPSCs have other molecular hallmarks of bivalent chromatin is important to establish. In primed human and mouse PSCs, regions containing bivalent chromatin are connected through long-range chromatin interactions that are thought to constrain and coordinate transcriptional regulation. In contrast, naive...
human and mouse PSCs lack long-range connections between bivalent chromatin sites, suggesting that although developmental genes are marked by H3K27me3 in naive cells, the mode of regulation might differ25,70.

Following the unexpected discovery of H3K27me3 at trophoblast-associated genes in naive hPSCs, we hypothesized that this repressive modification might functionally oppose the induction of trophoblast cell identity. We tested this prediction using

**Fig. 7 | PRC2 inhibition accelerates trophoblast development and cavity formation in human blastoids.**

| a | Schematic of the experimental set-up for studying the role of PRC2 in trophoblast formation in human blastoids. Blastoids are formed by aggregating naive hPSCs in microwells19. Created with BioRender.com. |
| b | Proportion of TROP2+ trophoblast cells in human blastoids at 36 h and 60 h with and without PRC2i; n = 3 biologically independent samples. c, NANOG, GATA3 and FOXA2 expression in 36 h (left) and 60 h (right) blastoids with or without PRC2i, quantified from immunofluorescence images (Extended Data Fig. 7a–d). The boxplots show the interquartile range (box limits showing the 25th and 75th percentile) and median (centre line) of the ratio of cells belonging to individual lineages, represented as percentages of the total number of cells per blastoid. Whiskers indicate 1.5x the interquartile range; n = 21 (36 h without PRC2i (−PRC2i)), 23 (36 h + PRC2i), 27 (60 h − PRC2i blastoids) and 17 (60 h + PRC2i) blastoids were quantified from a single experiment. Two-sided Wilcoxon rank-sum test with Bonferroni correction; 36 h, **P = 3.7 × 10−3 and ****P = 1.1 × 10−7; 60 h, *P = 1.1 × 10−2 and ****P = 2.5 × 10−7. 

| d | Bright-field images showing accelerated cavity formation during human blastoid formation (0–96 h) following PRC2i treatment (d) and at 36 h of human blastoid formation following PRC2i treatment (e). f, Fold change in cavitated human blastoids after 36 h with PRC2i (left) and 60 h with or without PRC2i (right). Data are shown as the FC normalized to 36 h without PRC2i; n = 3 biologically independent samples. g, Model showing that PRC2 restricts the induction of trophoblast fate from naive hPSCs. For color bars, darker colors indicate higher levels, except for the chromatin proteome, where pink represents naive chromatin proteome and grey represents primed chromatin proteome. Our findings establish that PRC2 acts as a barrier to lineage specification in naive hPSCs, opposing the formation of trophoblast cells in the presence of differentiation cues. In addition, our results uncover a potential role for PRC2 to safeguard the naive epigenome against adopting features of primed pluripotency, similar to observations in mice38. PRC2 activity establishes a higher global level of H3K27me3 in naive hPSCs compared with primed hPSCs, whereas the number of defined H3K27me3 peaks shows the opposite pattern. Our study also defined distinct chromatin proteomes that differ between naive and primed pluripotent states. Source data are provided.

---

**Fig. 7** | PRC2 inhibition accelerates trophoblast development and cavity formation in human blastoids. a, Schematic of the experimental set-up for studying the role of PRC2 in trophoblast formation in human blastoids. Blastoids are formed by aggregating naive hPSCs in microwells19. Created with BioRender.com. b, Proportion of TROP2+ trophoblast cells in human blastoids at 36 h and 60 h with and without PRC2i; n = 3 biologically independent samples. c, NANOG, GATA3 and FOXA2 expression in 36 h (left) and 60 h (right) blastoids with or without PRC2i, quantified from immunofluorescence images (Extended Data Fig. 7a–d). The boxplots show the interquartile range (box limits showing the 25th and 75th percentile) and median (centre line) of the ratio of cells belonging to individual lineages, represented as percentages of the total number of cells per blastoid. Whiskers indicate 1.5x the interquartile range; n = 21 (36 h without PRC2i (−PRC2i)), 23 (36 h + PRC2i), 27 (60 h − PRC2i blastoids) and 17 (60 h + PRC2i) blastoids were quantified from a single experiment. Two-sided Wilcoxon rank-sum test with Bonferroni correction; 36 h, **P = 3.7 × 10−3 and ****P = 1.1 × 10−7; 60 h, *P = 1.1 × 10−2 and ****P = 2.5 × 10−7. d, Bright-field images showing accelerated cavity formation during human blastoid formation (0–96 h) following PRC2i treatment (d) and at 36 h of human blastoid formation following PRC2i treatment (e). f, Fold change in cavitated human blastoids after 36 h with PRC2i (left) and 60 h with or without PRC2i (right). Data are shown as the FC normalized to 36 h without PRC2i; n = 3 biologically independent samples. g, Model showing that PRC2 restricts the induction of trophoblast fate from naive hPSCs. For color bars, darker colors indicate higher levels, except for the chromatin proteome, where pink represents naive chromatin proteome and grey represents primed chromatin proteome. Our findings establish that PRC2 acts as a barrier to lineage specification in naive hPSCs, opposing the formation of trophoblast cells in the presence of differentiation cues. In addition, our results uncover a potential role for PRC2 to safeguard the naive epigenome against adopting features of primed pluripotency, similar to observations in mice38. PRC2 activity establishes a higher global level of H3K27me3 in naive hPSCs compared with primed hPSCs, whereas the number of defined H3K27me3 peaks shows the opposite pattern. Our study also defined distinct chromatin proteomes that differ between naive and primed pluripotent states. Source data are provided.
two different cellular models of naive-to-trophoblast specification and found that the acute inhibition of PRC2 activity promoted trophoblast-fate induction. Curiously, a recent study reported that PRC2-deficient primed hPSCs upregulate GATA3 and KRT7 when transferred into trophoblast stem cell medium\(^1\). However, the low efficiency and prolonged timing of these events suggest that the mechanisms and developmental relevance when starting from a primed state are distinct from the naive-to-trophoblast transition that we uncover here\(^1\). Our experiments showed that PRC2 inhibition was indeed not sufficient to increase the efficiency of trophoblast-fate induction in primed hPSCs to levels that are comparable to naive hPSCs.

Many of our conclusions are in line with another study published in this issue\(^3\). One of the few differences between the two studies relates to whether PRC2 inhibition causes naive hPSCs to differentiate in self-renewing conditions. We found there is no miscellaneous differentiation of naive hPSCs following PRC2 inhibition with UNC1999 in PXGL medium, whereas Kumar et al.\(^5\) report significant levels of differentiation of naive hPSCs following PRC2 inhibition with EPZ-6438 in t2iLGö media. We tested whether this difference could be due to the different inhibitors used. Naive hPSCs treated with EPZ-6438 in PXGL medium also showed no change in naive hPSC differentiation (Extended Data Fig. 7g). We believe this difference can instead be attributed to the different naive hPSC media used, which alter the permissiveness of naive hPSCs to induce differentiation. In particular, PXGL medium contains a WNT antagonist (XAV939) whereas t2iLGö medium contains a WNT activator (CHIR99021). Shielding from WNT stimulation protects naive hPSCs against the induction of differentiation-associated genes\(^9\).

We also examined the role of PRC2 in a second model of human trophoblast development. Our results in blastoids showed that PRC2 inhibition accelerates trophoblast induction and the appearance of a blastocoel-like cavity during blastoid formation. These findings raise the possibility that the controlled inhibition of PRC2 could be a way to improve the timing and efficiency of blastoid formation. Furthermore, because extended developmental plasticity is also a property of epiblast cells in human pre-implantation embryos\(^9\), our results raise the possibility that PRC2 might also fulfill a similar role in human development. Whether PRC2 opposes trophoblast specification in human embryos is an exciting line of future research with important implications for understanding the causes of infertility and developmental disorders.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-022-00932-w.

Received: 2 September 2021; Accepted: 5 May 2022; Published online: 13 June 2022

References
1. Rossant, J. Investigation of the determinative state of the mouse inner cell mass. II. The fate of isolated inner cell masses transferred to the oviduct. J. Embryol. Exp. Morphol. 33, 991–1001 (1975).
2. De Paep, C. et al. Human trophoderm cells are not yet committed. Hum. Reprod. 28, 740–749 (2012).
3. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human pre-implantation embryos. Cell 165, 1012–1026 (2016).
4. Gerri, C. et al. Initiation of a conserved trophoderm program in human, cow and mouse embryos. Nature 587, 443–447 (2020).
5. Meistermann, D. et al. Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. Cell Stem Cell 28, 1625–1640 (2021).
6. Guo, G. et al. Human naive epiblast cells possess unrestricted lineage potential. Cell Stem Cell 28, 1046–1056 (2021).
7. Mølë, M. A. et al. A single cell characterisation of human embryogenesis identifies pluripotency transitions and putative anterior hypoblast centre. Nat. Commun. 12, 3679 (2021).
8. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell 158, 1254–1269 (2014).
9. Rostovskaya, M., Sturparo, G. G. & Smith, A. Capacitation of human naive pluripotent stem cells for multi-lineage differentiation. Development 146, dev172916 (2019).
10. Linneberg Agerholm, M. et al. Naive human pluripotent stem cells respond to Wnt, Nodal and LIF signalling to produce expandable naive extra-embryonic endoderm. Development 146, dev180620 (2019).
11. Dong, C. et al. Derivation of trophoblast stem cells from naive human pluripotent stem cells. Cell Stem Cell 9, e52504 (2020).
12. Cinkorppunin, J. K. et al. Naive human embryonic stem cells can give rise to cells with a trophoblast-like transcriptome and methylome. Stem Cell Rep. 15, 198–213 (2020).
13. Castel, G. et al. Induction of human trophoblast stem cells from somatic cells and pluripotent stem cells. Cell Rep. 33, 108419 (2020).
14. Io, S. et al. Capturing human trophoblast development with naive pluripotent stem cells in vitro. Cell Stem Cell 28, 1023–1039 (2021).
15. Yu, L. et al. Blastocyst-like structures generated from human pluripotent stem cells. Nature 591, 620–626 (2021).
16. Liu, X. et al. Modelling human blastocysts by reprogramming fibroblasts into iPS cells. Nature 591, 627–632 (2021).
17. Yanagida, A. et al. Naive stem cell blastocyst model captures human embryo lineage segregation. Cell Stem Cell 28, 1016–1022 (2021).
18. Sozen, B. et al. Reconstructing aspects of human embryogenesis with pluripotent stem cells. Nat. Commun. 12, 5550 (2021).
19. Kagawa, H. et al. Human blastoid model blastocyst development and implantation. Nature https://doi.org/10.1038/s41586-021-04267-8 (2021).
20. Dong, C., Fischer, L. A. & Theunissen, T. W. Recent insights into the naive state of human pluripotency and its applications. Exp. Cell Res. 385, 111645 (2019).
21. Bayerl, J. et al. Principles of signaling pathway modulation for enhancing human naive pluripotency induction. Cell Stem Cell 28, 1549–1565 (2021).
22. He, L. & Crabtree, G. R. Chromatin remodelling during development. Nature 463, 474–484 (2010).
23. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. 12, 36–47 (2011).
24. Cavalli, G. & Heard, E. Advances in epigenetics link genetics to the environment and disease. Nature 571, 489–499 (2019).
25. Chowancz, P. et al. Widespread reorganisation of pluripotent factor binding and gene regulatory interactions between human pluripotent states. Nat. Commun. 12, 2098 (2021).
26. Killens, S. et al. Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. Nat. Commun. 9, 360 (2018).
27. Liu, X. et al. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. Nat. Methods 14, 1055–1062 (2017).
28. Theunissen, T. W. et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15, 524–526 (2014).
29. 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).
30. van Mierlo, G., Veenstra, G. J. C., Vermeulen, M. & Marks, H. The complexity of PRC2 subcomplexes. Trends Cell Biol. 29, 660–671 (2019).
31. Sperber, H. et al. The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. Nat. Cell Biol. 17, 1523–1535 (2015).
32. De Clerck, L. et al. Untargeted histone profiling during naive conversion uncovers conserved modification markers between mouse and human. Sci. Rep. 9, 17240 (2019).
33. Batlle, S. L. et al. Enhancer chromatin and 3D genome architecture changes from naive to primed human embryonic stem cell states. Stem Cell Rep. 12, 1129–1144 (2019).
34. Collinson, A. et al. Deletion of the polycomb-group protein EZH2 leads to compromised self-renewal and differentiation defects in human embryonic stem cells. Cell Rep. 17, 2700–2714 (2016).
35. Moody, J. D. et al. First critical repressive H3K27me3 marks in embryonic stem cells identified using designed protein inhibitor. Proc. Natl Acad. Sci. USA 114, 10125–10130 (2017).
36. Shao, Y. et al. PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naive ESCs. Nat. Commun. 8, 672 (2017).
37. Kustatscher, G., Wills, K. L. H., Furlan, C. & Rappsilber, J. Chromatin enrichment for proteomics. Nat. Protoc. 9, 2090–2099 (2014).
38. van Mierlo, G. et al. Integrative proteomic profiling reveals PRC2-dependent epigenetic crosstalk maintains ground-state pluripotency. Cell Stem Cell 24, 123–137 (2019).
39. Hackett, J. A. & Surani, M. A. Regulatory principles of pluripotency: from the ground state up. Cell Stem Cell 15, 416–430 (2014).
40. Pastor, W. A. et al. TEAP2C regulates transcription in human naive pluripotency by opening enhancers. Nat. Cell Biol. 20, 553–564 (2018).
41. Zhang, J. et al. LIN28 regulates stem cell metabolism and conversion to primed pluripotency. Cell Stem Cell 19, 66–80 (2016).
42. 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).
43. Liu, X. et al. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat. Commun. 4, 1563 (2013).
44. Huang, X. et al. OCT4 cooperates with distinct ATP-dependent chromatin remodelers in naive and primed pluripotent states in human. Nat. Commun. 12, 5123 (2021).
45. Pontis, J. et al. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. Cell Stem Cell 24, 724–735 (2019).
46. Alabert, C. et al. Domain model explains propagation dynamics and stability of histone H3K27 and H3K36 methylation landscapes. Cell Rep. 30, 1223–1234 (2020).
47. Streubel, G. et al. The H3K36me2 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3 domains in embryonic stem cells. Mol. Cell 70, 371–379 (2018).
48. Yuan, W. et al. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. 286, 7983–7989 (2011).
49. Ali, M. A. M. et al. Matrix metalloproteinase-2 mediates ribosomal RNA transcription by cleaving nuclear histones. FEBS J. 288, 6736–6753 (2021).
50. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. Cell 149, 590–604 (2012).
51. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. eLife 8, e46314 (2019).
52. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife 6, 62186 (2017).
53. Blakeley, P. et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development 142, 3613 (2015).
54. Fogarty, N. M. E. et al. Direct reprogramming of human embryonic to trophoblast stem cells. Preprint at bioRxiv https://doi.org/10.1101/2021.08.18.456785 (2021).
55. Niakan, K. K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. Dev. Biol. 375, 54–64 (2013).
56. Konz, K. D. et al. An orally bioavailable chemical probe of the lysine methyltransferases EZH2 and EZH1. ACS Chem. Biol. 8, 1324–1334 (2013).
57. Constância, M. et al. Placental-specific IGF-II is a major modulator of placental and fetal growth. Nature 417, 945–948 (2002).
58. Li, T. et al. CTCF regulates allelic expression of Igf2 by orchestrating a promoter–polycomb repressive complex 2 intrachromosomal loop. Mol. Cell. Biol. 28, 6473–6482 (2008).
59. Assanoma, K. et al. SATB homeobox proteins regulate trophoblast stem cell renewal and differentiation. J. Biol. Chem. 287, 2257–2268 (2012).
60. Moretto Zita, M., Soncin, F., Natale, D., Pizzo, D. & Parast, M. Gene expression profiling reveals a novel regulatory role for Sox21 protein in mouse trophoblast stem cell differentiation. J. Biol. Chem. 290, 30152–30162 (2015).
61. Takahashi, Y., Dominici, M., Swift, J., Nagy, C. & Ilhé, J. N. Trophoblast stem cells rescue placental defect in SOCS3-deficient mice. J. Biol. Chem. 281, 11444–11445 (2006).
62. Okae, H. et al. Derivation of human trophoblast stem cells. Cell Stem Cell 22, 50–63 (2018).
63. Soncin, F. et al. Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. Development 145, dev156273 (2018).
64. Patani, H. et al. Transition to naive human pluripotency mirrors pan-cancer DNA hypermethylation. Nat. Commun. 11, 3671 (2020).
65. Kao, T.-H. et al. Ectopic DNMT3L triggers assembly of a repressive complex for retroviral silencing in somatic cells. J. Virol. 88, 10680–10695 (2014).
66. Yu, Y. C.-Y. et al. Transient DNMT3L expression reinforces chromatin surveillance to halt senescence progression in mouse embryonic fibroblast. Front Cell Dev. Biol. 8, 103 (2020).
67. Azuara, V. et al. Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. 8, 532–538 (2006).
68. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326 (2006).
69. Schoenfelder, S. et al. Polycomb repressive complex 1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. 47, 1179–1186 (2015).
70. Joshi, O. et al. Dynamic reorganization of extremely long-range promoter–promoter interactions between two states of pluripotency. Cell Stem Cell 17, 748–757 (2015).
71. Denholtz, M. et al. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell 13, 602–616 (2013).
72. Wei, Y. et al. Efficient derivation of human trophoblast stem cells from primed pluripotent stem cells. Sci. Adv. 7, eabd4416 (2021).
73. Kumar, B. et al. Polycomb repressive complex 2 shields naive human pluripotent cells from trophoderm-differentiation. Nat. Cell Biol. https://doi.org/10.1038/s41556-022-00916-w (2022).
74. Zhou, F. et al. Reconstituting the transcriptome and DNA methylome landscapes of human implantation. Nature 572, 660–664 (2019).
Methods

Our research complies with all relevant ethical regulations and guidelines. Experiments with hPSCs were approved by the UZ/KU Leuven Ethics Committee (S52426, S66185 and S66492), the Flemish Government (SBB 219 2020/0435) and the Steering Committee of the UK Stem Cell Bank (SCSC11-58). The WiCell line H9 (WA/09) was used under the agreements 20-WO-341, 12-WO-202 and 18-WO-026. Blastoid generation was approved by the Commission for Science Ethics of the Austrian Academy of Sciences and the KU/IZU Leuven Ethics Committee (S66185 and S66492). This work did not exceed a developmental stage normally associated with 14 consecutive days in culture after fertilization. The animal work carried out in this study was covered by project licences (ECD_P003-2016 and ECD_P170_2019 to VP, and to F.L., respectively) approved by the KU Leuven Animal Ethics Committee.

Human cell lines. Experiments were carried out using the following cell lines: H9 hESCs (obtained from WiCell) and ICSIG-1 1PSC0028 hiPSCs (obtained from Sigma-Aldrich). The H9 line chemically reset to the naive state was provided by A. Smith75 (with permission from WiCell) and was used for all experiments in Figs. 1–4 and Extended Data Figs. 1–3,4a–j. Other naive hPSC lines (H9 and Human cell lines.) were cultured in PXGL medium9,76, consisting of a 1:1 mixture of DMEM/F12 and Neurobasal media supplemented with 0.5% N2 supplement, 0.5% B27 supplement, 2 mM l-glutamine, 50 U ml −1 penicillin–streptomycin (Gibco, 25300-081), 0.5% N2 supplement (Gibco, 17502-048), 1% B27 supplement (Gibco, 17504-044), 1% penicillin–streptomycin (Gibco, 15104022) and 0.1 mM β-mercaptoethanol (Gibco, 31500-010)) supplemented with 1 µM PD325901 (Axon Medchem, 1408), 10 ng ml −1 recombinant human LIF (PeproTech, 305-05) and 1 mM valproic acid (Sigma-Aldrich, V0033900). After 3 d in CR1 medium, the cells were switched to CR2 medium—N2B27 medium supplemented with 1 µM PD325901, 1 µM CHIR99021 (Axon Medchem, 1386), 2 µM Gö6983 and 10 mM l-histidine. These cells were passed as single cells every 4–5 d through a 5-min incubation in accutase (Sigma-Aldrich, A6964-100ML) at 37 °C. Naive ICSIG-1 hPSCs were switched at passage 10 into PXGL medium for maintenance and expansion.

To convert primed hPSCs (used in Extended Data Fig. 5c) to naive hESCs, primed hPSCs were trypsinized and seeded onto gelatin-coated plates with MEFs in human KSR primed medium along with 10 µM Y-27632 (Tocris, 1254) in humidified normoxia conditions (5% CO2) for 2 d using a previously described protocol77. On day 3, after a PBS wash, the medium was changed to S1A medium—composed of 50% DMEM/F12 medium, 50% Neurobasal medium, 1% N2 supplement, 2% B27 supplement, 20 ng ml −1 recombinant human LIF; 2 mM t-glutamine, 1× non-essential amino acids, 0.1 µM β-mercaptoethanol, 1×penicillin–streptomycin and 1×µg ml −1 BSA (Sigma-Aldrich, A3059) supplemented with five inhibitors, that is, PD025901 (Stemgent, 1 µM), IM-12 (Enzo, 1 µM), SB508858 (R&D Systems, 0.5 µM), WH-4-023 (A Chemtek, 1 µM), Y-27632 (Tocris, 10 µM) and activin A (Peprotech, 20 ng ml −1) in hypoxia conditions (5% CO2 and 5% O2) at 37 °C. Dome-shaped naïve colonies were observed after 10–14 d. Naïve cells were passaged as single cells every 4–5 d using accutase with an incubation of 5 min at 37 °C. The cells were switched into PXGL medium at passage 12.

Due to the limited availability of ICSIG-1 hPSCs, we used hESCs derived from the same line in a comparison experiment of Extended Data Fig. 7g, where EPZ-6438 (10 µM) was used instead of the PRC2 inhibitor as well as on days 0 and 4 of the PRC2-inhibitor treatment. Due to the limited availability of ICSIG-1 hPSCs, we used hESCs derived from the same line in a comparison experiment of Extended Data Fig. 7g, where EPZ-6438 (10 µM) was used instead of the PRC2 inhibitor as well as on days 0 and 4 of the PRC2-inhibitor treatment.
Cell counting and viability. Cells were counted and viability was assessed using a LUNA-FI dual fluorescence cell counter (Logos Biosystems) on days 0, 1, 2, 3 and 4 of naive hPSC-to-trophoblast conversion. Cells were collected for cell count and viability measurements by collecting the supernatant and dissociating the attached cells to single cells using TrypLE. The cells were centrifuged for 5 min at 400 g and the pellet was resuspended in 100 μl culture medium. The sample was prepared by adding 2 μl acridine orange–propidium iodide stain solution (Logos Biosystems, F25001) to 18 μl of sample (pellet diluted in culture medium). The sample preparation (10 μl) was loaded into a chamber of a PhotonSlade cell counter (Logos Biosystem) to count the total number of cells and measure cell viability.

Human blastoids. Naïve hPSCs cultured on MEFs in PXGL medium were pre-treated with PRC2 inhibitor (1 μMUNC1999) for 4 d before blastoid induction. Blastoids were induced 72 h after the last PRC2 inhibitor addition. Blastoids were induced 72 h after the last PRC2 inhibitor addition. The cells were resuspended in PXGL medium supplemented with 10 μM Y-27632 (MedChemExpress), seeded onto gelatin-coated plates and incubated at 37 °C for 70 min to deplete the MEFs. The unattached cells were collected, pelleted through centrifugation and resuspended in N2B27 medium containing 10 μM Y-27632 with or without PRC2 inhibitor (aggregation medium), after which 30,000 cells were seeded onto an array of 200-μm microwells inserted into a well of a 96-well plate. Note that microwell arrays comprising microwells were imprinted into 96-well plates 77,78. After 24 h, the aggregation medium was replaced with N2B27 medium supplemented with 1 μM PD0325901, 1 μM A83-01 (MedChemExpress, HY-10496, 500 μl) and 0.2% fish-skin gelatin. Following overnight incubation with primary antibodies in blocking solution at 4 °C, the cells were washed three times for 5 min with PBS, fixed in 10% formaldehyde (PFA; Life Technologies, 28908) for 10 min × antibodies in blocking solution at 4 °C, the cells were washed three times for 5 min and centrifuged again under the same conditions. The cell pellets were resuspended in 50 μl FACS buffer and incubated for 50 min at 4 °C. The cells were centrifuged, washed twice with flow buffer and incubated with antibody secondary mouse antibody conjugated to Alexa Fluor 488 (Invitrogen, A-21201) for 30 min at 4 °C. The cells were washed and centrifuged twice with flow buffer and resuspended in fresh flow buffer for flow cytometry analysis. Antibody information is provided in Supplementary Table 11.

Western blotting for histone proteins. Cells were washed once with PBS and dissociated using accutase followed by N2B27 medium supplemented with 10 μM Y-27632 and incubated with complete EDTA-free protease inhibitor cocktail (Roche) for 10 min at 4 °C and centrifuged at 300 g for 5 min at 4 °C. The cell pellet was resuspended in 0.2 M sucrose, incubated for 30 min at 4 °C and centrifuged for 12,000 g at 4 °C. The supernatant was removed, and the pellets were washed with acetone and incubated for 10 min at 4 °C. After centrifuging at 1,200 g for 10 min at 4 °C, a second acetone wash and centrifugation step was performed. Histone proteins were dissolved overnight in 100 mM Tris–HCl pH 9.0 containing protease inhibitors at 4 °C. The samples were centrifuged at 12,000 g for 10 min at 4 °C and the supernatant was retained.

Histone proteins were quantified using the Bradford assay and denatured by heating at 95 °C in 5 μl Protein loading dye (4% SDS, 0.25 M Tris HCl, 6.8 μl bromophenol blue, 0.5 M dithiothreitol and 30% glycerol) for 5 min. The histones were separated by electrophoresis on a 15% SDS–PAGE gel alongside a pre-stained protein standard (Bio-Rad) to assess the protein molecular weights. The histones were transferred onto nitrocellulose membranes using an iBlot transfer system at 25 V for 10 min. The membranes were blocked for 3 h at room temperature in TBS-T (1× Tween-buffered saline and 0.05% Tween 20) containing 5% dried skimmed milk and hybridized overnight at 4 °C with primary antibodies diluted in TBS-T containing 5% milk. The membranes were washed three times with TBS-T for 10 min before incubation with fluorescent-conjugated secondary antibodies diluted in TBS-T containing 5% milk at 4 °C for 1 h at room temperature and protected from light. The membranes were washed three times with TBS-T for 10 min and then once with 1× TBS before detection using an Odyssey imaging system (LI-COR Biosciences) or Clarity western ECL reagent (Bio-Rad). Antibody information is provided in Supplementary Table 11.

RNA extraction. RNA extraction was performed using one of two methods—either TRizol reagent or a RNeasy micro kit (Qiagen, 74004). RNA extraction with TRizol reagent (TRizol, 15596-018) was performed according to the TRizol reagent user guide. Briefly, cells were washed once with 1× PBS and dissociated with 400 μl TRizol reagent for 15 min at room temperature. After collection, the samples were stored at −80 °C until further use. Chloroform (80 μl) was added to the sample, mixed, incubated for 2–3 min and centrifuged at 12,000 g for 15 min at 4 °C in the presence of 1 μl glycogen. The aqueous phase containing the RNA was transferred to a new tube. Isopropanol (200 μl) was added to the sample and incubated for 10 min. Total RNA was precipitated by centrifugation at 12,000 g for 10 min at 4 °C. The pellet was resuspended in 200 μl of 75% ethanol. The sample was briefly vortexed at 20°C. After discarding the supernatant, the RNA pellet was air dried for 10 min and resuspended in 20 μl MilliQ water. RNA extraction using the RNeasy micro kit was performed according to the RNeasy micro handbook.

Reverse transcription. Two reverse transcription protocols were used: (1) reverse transcription with homemade reverse transcriptase (RT) and (2) first-strand complementary DNA synthesis using SuperScript II RT (Thermo Fisher, 18064-022). The homemade RT was used for the samples that were collected with the TRizol reagent and the SuperScript II RT for the samples collected using the RNeasy micro kit. RNA from the TRizol-isolated samples (50 μg) was added to a mixture of 4 μg random primers (Sigma, M6041), 4 μg dNTPs (Thermo Fisher, 18148-012), 50 μg μl random hexamers (Thermo Fisher, S0142), 0.25 μM of each gene-specific primer (reverse), 0.2 μM of each dNTP (Thermo Fisher, 10297-018), 1× first-strand buffer (homemade), 2 mM dithiothreitol (Sigma, G717-1318-01), 2 units μl−1 RNaseOUT.
Bulk RNA-seq library preparation. RNA extraction from 0.5 × 10^6 cells per sample was performed using an RNeasy micro kit (Qiagen, 74004) following the manufacturer’s protocol. Messenger RNA-seq libraries were prepared starting from 500 ng enriched total RNA using a KAPA stranded mRNA-Seq kit v2 (KAPA Biosystems; Cat. No. 07962193001) with KAPA single-indexed adapter kit sets A and B (Illumina, KR1317) following the manufacturer’s protocol. Libraries were pooled with a final library concentration of 7 nM. The quality of the input RNA, cDNA and individual libraries was assessed using an Agilent 2100 Bioanalyzer system at the KU Leuven Nucleomics core. Sequencing was performed at the KU Leuven Genomics Core on an Hiseq1500 (illumina) in single-end mode (50 bp), yielding an average of 29 × 10^6 reads per sample (Supplementary Table 1).

Bulk RNA-seq analysis. Quality assessment of the bulk RNA-seq data was performed using FastQC (v0.11.8; Babraham Bioinformatics). Samples were mapped to the human GRCh38.p12 reference genome with the corresponding GENCODE v31/Ensembl 97 using STAR (v2.7.1a)86. The count table was generated using featureCounts (v2.0.1-2) with default parameters. Downstream analyses were performed using the R package DESeq2 (v1.26.0)82. Samples were filtered to keep genes with average count > 10. Naive and primed hPSCs and Drosophila S2 cells were merged and cell counts were normalized using the Seurat global-scaling normalization method ‘LogNormalize’, which normalizes the feature expression measurements for each cell by the total expression, multiplies this by a 10,000-scale factor and log-transforms the result. Differential expression testing was performed using the FindMarkers function in Seurat based on the non-parametric Wilcoxon rank-sum test applying the log-transformed FC threshold of averaged log(FC) > 0.25. A graph-based cell-clustering approach was used to cluster cells with the FindClusters function in Seurat. Loom files were generated in R using build_loom and add_col_attr from SCopeLoom (version 0.3.1).

Integration of scRNA-seq data. Datasets used for gene expression integration can be found in ArrayExpress under the accession number E-MTAB-3929 (ref.7) and in the Gene Expression Omnibus (GEO) database under the accession number GSE109555 (ref.61). Integration of published scRNA-seq embryonic datasets3,20 with the day 4 trophoblast conversion + PRG2 scRNA-seq dataset generated in this study was performed using the scRNA-seq integration tool. Anchors for integration were found using the FindIntegrationAnchors function with default arguments. Parameters and data were integrated across all features. An integration-based UMAP was constructed using the runUMAP function with dims: 150. Published scRNA-seq embryonic datasets3,20 were annotated following19 for UMAP (Fig. 6c). scRNA-seq reads from of3,20 were simplified as follows: embryo trophectoderm (comprising early, medium and late trophectoderm), embryo trophoblast (comprising early, medium, late and apoptosis trophoblast); embryo extravillous trophoblast (comprising pre-extravillous trophoblast and extravillous trophoblast); embryo syncytiotrophoblast (comprising pre-synctyiotrophoblast and syncytiotrophoblast) and (pre)blastocyst (comprising morula and B1/B2 blastocysts).

Calibrated CUT&RUN. Concavanal A-conjugated paramagnetic beads (Epichery, 21-1401) were resuspended and washed twice on a magnetic rack in bead activation buffer composed of 20 mM HEPES pH 7.9, 10 mM KCl, 1 mM CaCl2 and 1 mM MnCl2. Naive or primed hPSCs and Drosophila S2 cells were dissociated and counted. Cell pellets were washed twice with a wash buffer composed of 20 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM spermidine supplemented with EDTA-free protease inhibitor. The cells, 50,000 human cells and 20,000 Drosophila cells in wash buffer, were added to concavanal A beads and incubated for 10 minutes at room temperature with rotation to immobilize the cells. The concavanal A beads were collected with a magnetic rack, the supernatant was discarded and the beads were resuspended in antibody buffer (wash buffer supplemented with 0.08% digitonin (Millipore) and 2 mM EDTA). Antibody (1 μg) was added to the beads, which were incubated overnight at 4 °C with rotation. Antibody information is provided in Supplementary Table 11. The following day, the beads were washed twice with digitonin buffer (wash buffer supplemented with 0.08% digitonin). CUTANA pA/G MNase (2.5 μl; Epichery, 15-1016) was added to 50 μl digitonin buffer and incubated for 10 minutes at room temperature. The beads were washed twice with digitonin buffer, and the MNase was activated with 2 mM CaCl2 and incubated for 2 hours at 37 °C. MNase activity was terminated by adding 40 μM CaCl2 and stop buffer (10 mM EDTA, 50 μg ml−1 RNase A and 50 μg ml−1 glycogen). Cleaved DNA fragments were released from nuclei by incubation for 10 minutes at 37 °C, centrifugation for 5 minutes at 16,000 g at 4 °C and collection of the supernatant from the beads on a magnetic rack. The DNA was purified by incubation with 1 μl SDS (20%) and 1.5 μl proteinase K (20 mg ml−1) at 70 °C for 10 minutes, followed by a 1:8X AMPure XP bead (Beckman Coulter) clean-up into DNA to bind-tubes (Eppendorf) and elution in 50 μl of 0.1X TE. Libraries were prepared using a NEBNext ultra II DNA library preparation kit for Illumina (NEB) using the manufacturer’s protocol, with libraries indexed using NEBNext multiplex oligos for Illumina (index primers sets 1 and 2; NEB). Following library preparation, the library fragment size and concentration were determined using a Qubit fluorometer double stranded DNA high sensitivity assay kit with an Agilent Bioanalyzer 2100 and using a KAPA library quantification kit (KAPA Biosystems). The samples were sequenced on an Illumina NexSeq500 instrument as HighOutput 75-base paired-end reads at the Babraham Institute Next Generation Sequencing Facility (highest read count = 53,117,572, lowest read count = 20,861,009, and average read count = 32,213,678).

Calibrated CUT&RUN analysis. Raw FastQ data were trimmed using TrimmGalore (v0.6.6, Babraham Bioinformatics) and aligned to the GRCh38 genome or the Drosophila BDGP6 genome using Bowtie2 (v2.3.2) with the following parameters –very-sensitive -I 10 -X700. High-quality reads with a mapping quality value of >20 were retained by filtering the reads (tools view v1.0). For each sample, the following features were retained as the ratio of the sample with the lowest number of unique mapped Drosophila spike-in tags over the number of unique mapped Drosophila spike-in tags per sample. These calibration factors were then used to scale the human genome-mapped reads by random downscaling. Calibrated browser extendible data (BED) files were produced using BEDTools genome cov

Articles Nature Cell Biology

Quantitative PCR with reverse transcription. Quantitative PCR was performed using a Platinum SYBR Green qPCR SuperMix-UDG kit (Invitrogen, 17373046) on an ABI Viia7 real-time PCR system (Applied Biosystems) following the manufacturer’s protocol. Each quantitative PCR well contained 1.25 ng cDNA together with the Platinum SYBR Green qPCR SuperMix-UDG, ROX reference dye (ABI, 7500) and a primer mix of the forward and reverse primers (0.25 μM each). Primer sequences are provided in Supplementary Table 12.

Single-cell preparation and scRNA-seq. Cells were washed with 1×PBS before dissociation from the culture dish via incubation with TrypLE express enzyme (Thermo Fisher, 12605-036) for 8 min at 37 °C. The single-cell suspensions were filtered through a Falcon 40-μm cell strainer and centrifuged at 200 g for 5 min. After resuspension in 1×PBS with 0.04% RSA, the concentration of cells in the single-cell suspension was determined using a Luminex FL-automated Fluorescence Cell Counter (Luminex). Cells were loaded onto the 10X Chromium single-cell platform (10X Genomics) at a concentration of 1,000 cells μl−1. The cells were washed with wash buffer, were added to concavanal A beads and incubated for 10 minutes at room temperature. The beads were washed twice with digitonin buffer, and the MNase was activated with 2 mM CaCl2 and incubated for 2 hours at 37 °C. MNase activity was terminated by adding 40 μM CaCl2 and stop buffer (10 mM EDTA, 50 μg ml−1 RNase A and 50 μg ml−1 glycogen). Cleaved DNA fragments were released from nuclei by incubation for 10 minutes at 37 °C, centrifugation for 5 min at 16,000 g at 4 °C and collection of the supernatant from the beads on a magnetic rack. The DNA was purified by incubation with 1 μl SDS (20%) and 1.5 μl proteinase K (20 mg ml−1) at 70 °C for 10 minutes, followed by a 1:8X AMPure XP bead (Beckman Coulter) clean-up into DNA to bind-tubes (Eppendorf) and elution in 50 μl of 0.1X TE. Libraries were prepared using a NEBNext ultra II DNA library preparation kit for Illumina (NEB) using the manufacturer’s protocol, with libraries indexed using NEBNext multiplex oligos for Illumina (index primers sets 1 and 2; NEB). Following library preparation, the library fragment size and concentration were determined using a Qubit fluorometer double stranded DNA high sensitivity assay kit with an Agilent Bioanalyzer 2100 and using a KAPA library quantification kit (KAPA Biosystems). The samples were sequenced on an Illumina NexSeq500 instrument as HighOutput 75-base paired-end reads at the Babraham Institute Next Generation Sequencing Facility (highest read count = 53,117,572, lowest read count = 20,861,009, and average read count = 32,213,678).

Calibrated CUT&RUN analysis. Raw FastQ data were trimmed using TrimmGalore (v0.6.6, Babraham Bioinformatics) and aligned to the GRCh38 genome or the Drosophila BDGP6 genome using Bowtie2 (v2.3.2) with the following parameters –very-sensitive -I 10 -X700. High-quality reads with a mapping quality value of >20 were retained by filtering the reads (tools view v1.0). For each sample, the following features were retained as the ratio of the sample with the lowest number of unique mapped Drosophila spike-in tags over the number of unique mapped Drosophila spike-in tags per sample. These calibration factors were then used to scale the human genome-mapped reads by random downscaling. Calibrated browser extendible data (BED) files were produced using BEDTools genome cov
scouting by these calibration factors and were used for peak calling (v.2.29.2)\textsuperscript{15}. Peak calling was performed using the CUT&RUN optimized Sparse Enrichment Analysis for CUT&RUN (SEACR) algorithm and the top 1% peaks were retained (v.1.3)\textsuperscript{15}. Peaks closer than 200 bp were merged, and peaks common to both replicates were determined by BEDTools intersect to generate final peak sets for naive and primed. Peaks called in naive and primed hPSC datasets were concatenated into a combined peak list and de-duplicated. Peaks that were differentially enriched between naive and primed hPSCs were then determined from this concatenated list using a DSEseq implementation in SeqMonk (v.1.47.2; Babraham Bioinformatics) to identify differential regions enriched with P < 0.05 after Benjamin–Hochberg multiple-testing correction. Common peaks were classified as peaks in the concatenated list that were not statistically enriched in either condition. These peaks were then filtered against the ENCODE GRCh38 exclusion list to remove coverage outliers. The fraction of reads in peak scores were calculated using deepTools bamCoverage with scaling settings --missingDataAZero and plotted with plotHeatmap or plotProfile.

For analysis of 1-kb windows of the genome, the coverage of 1-kb bins of sorted and indexed binary alignment map (BAM) files were processed by deepTools multiBamSummary scaling by the calculation normalization factors previously calculated. Replicate reproducibility was assessed by Pearson's correlation of signal across these pronounced peaks using deepTools with a correlation parameter --log1p for plotting. Principal component analysis plots were produced using signals at combined naive and primed peak sets on downsampled BAM files in SeqMonk (v.1.47.2; Babraham Bioinformatics). Scatter, violin, box, bar and density plots were produced using the R package ggplot2.

Data processing and visualization. Figures were produced in R using the R package ggplot2 (v.3.3.3), Microsoft Excel, BioRender and Adobe Illustrator. The online application for viewing the multi-omics dataset was developed using the Shiny package for R.

Histone extraction, propionylation and digestion. Histone preparation of naive (cultured in PXGL medium) and primed (cultured in E8 medium) hPSCs was performed starting from frozen cell pellets by isolating the nuclei through resuspension in hypotonic lysis buffer (10 mM Tris–HCl pH 8.0, 1 mM KCl and 1.5 mM MgCl\textsubscript{2}) complemented with 1 mM dithiothreitol and complete protease inhibitors (Roche) at 4 °C. Histones were collected by centrifugation at 3,400 g for 5 min at 4 °C, washed twice with 0.1% HCl in acetone and once with 100% acetone. The washed histones were collected by centrifugation at 20,000 g for 5 min at 4 °C, washed once more with acetone and then left to air dry. The dried pellets were resuspended in mass spectrometry-grade water and adjusted to pH 8.0 with 100 mM NH\textsubscript{4}HCO\textsubscript{3}.

The histone concentration was estimated using the Bradford assay. About 20 μg of purified histones in 20 μl of 100 mM NH\textsubscript{4}HCO\textsubscript{3} (pH 8.0) were derivatized in 5 μl propionylation reagent (mixture of propionic anhydride with acetonitrile at 1:1 ratio) at 1 °C for 1 h. After derivatization, this reaction was processed to ensure complete propionylation of unmodified lysines. Following this, the histones were digested overnight with 1 μg trypsin at room temperature. After digestion, the N termini of peptides were derivatized by two more rounds of propionylation. For injection into the mass spectrometer, the samples were desalted using C18 stage tips.

Liquid chromatography with MS/MS (hPTMs). The propionylated naive (cultured in PXGL medium) and primed (cultured in E8 medium) samples, complemented with a β-galactosidase (Sciex) and MPDS (Waters) internal standards, were digested overnight with 1 μl of 0.5 M NH\textsubscript{2}OH and 15 μl 100% formic acid. After digestion, the biologically most interesting hPTMs (acetylations and methylations), were identified in the cRAP database (https://www.thegpm.org/crap/). Subsequently, a FASTA file (version 5.197) was used as a standard, were resuspended in 0.1% formic acid resulting in 1.5 μl of each sample. Data-dependent acquisition was performed on a TripleTOF 6600+ system (AB Sciex) operating in positive mode coupled to an Eksigent NanoLC 425 HPLC system operating in capillary flow mode (5 μl/min)\textsuperscript{15}. Trapping and separation of the peptides was carried out on a reverse-phase column (5 μm, 150 mm × 0.2 mm, C18, 180 A\textsubscript{o}) and a Phenomenex Luna Omega Polar C18 column (50 μm × 0.3 mm, particle size 3 μm), respectively, using a low pH reverse-phase gradient. Buffers A and B of the mobile phase consisted of 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. A 60-min gradient going from 3% to 45% Buffer B, with a total run time of 75 min per sample, was applied. The samples were run in a randomized fashion and a quality control injection was incorporated every five samples. For each cycle, one full MS1 scan (m/z 400–1,250) of 250 ms was followed by an MS2 (m/z 65–2,000, high-sensitivity mode) of 200 ms. A maximum of ten precursors (charge state +2 to +5) exceeding 300 c.p.s. were monitored, followed by an exclusion for 10 s per cycle. A rolling collision energy with a spread of 15 V and a cycle time of 2.3 s was applied.

The propionylated naive (cultured in 2iLG0 medium) and primed (cultured in E8 medium) samples were analyzed using an EASY-nLC nanoHPLC (Thermo Scientific) fitted with a nano–column packed with inner diameter of 75 μm × 17 cm Reprosil-Pur C18-AQ (3 μm; Dr. Maisch GmbH). Online mixing of solvents was as follows: 2–28% solvent B (solvent A, 0.1% formic acid; solvent B, 95% acetonitrile and 0.1% formic acid) over 45 min, followed by 28–80% solvent B in 5 min and 80% solvent B for 10 min at a flow rate of 300 nL/min\textsuperscript{15}, which allowed separation of the analyte components and sprays into the Q-Exactive mass spectrometer (Thermo Scientific). A data-independent acquisition method, consisting of a full-scan mass spectrometry spectrum (m/z 300–2,000) at a resolution of 70,000 and tandem mass spectrometry (MS/MS) of windows of m/z 50 to a resolution of 15,000, was used. MS/MS data were acquired in centroid mode. The data-independent acquisition data were searched using EpiProfile\textsuperscript{15}.

Analysis of hPTMs. Analysis of the mass spectrometry data was performed as previously described\textsuperscript{15}. For all runs, raw data were imported in Progenesis QIP 4.2. (Nonlinear Dynamics, Waters), followed by alignment, feature detection and normalization. Next, a mascot generic format (MGF) file was created based on the twenty MS/MS spectra closest to the elution apex and exported for searches using Mascot (Matrix Science). First, a standard search was performed on the extracted MGF file to identify the propionylated standard peptides (β-galactosidase and MPDS) and to verify underpropionylation. Second, to identify the proteins present in the sample and to detect unexpected hPTMs, an error-tolerant search without biological modifications was carried out against a complete Human Swiss–Prot database (downloaded from UniProt and supplemented with contaminants from the cRAP database (https://www.thegpm.org/crap)). Subsequently, a FASTA database was created based on the reference to the error-tolerant search. In addition, the highest-ranking hPTMs that emerged from this search, complemented with the biologically most interesting hPTMs (acetylations and methylations), were selected to establish a set of nine hPTMs for further analysis. Next, the three MS/MS spectra closest to the elution apex per feature were merged into a single MGF file and exported for a MaxSearch search including the following parameters: (1) a mass error tolerance of 10 ppm and 50 ppm for the precursor and fragment ions, respectively; (2) Arg-C enzyme specificity, allowing for up to one missed cleavage site; (3) variable modifications included acetylation, butyrylation, crotonylation as well as trimethylation on lysine, methylthionation on arginine, dimethylation on both lysine and arginine, deamidation on asparagine, glutamine and arginine.
Log transformation and quantile normalization of the data were performed. Each run belongs (that is, naive, primed and primed + inhibitor, primed versus primed + inhibitor, primed versus primed + inhibitor, naive versus primed + inhibitor) to which the name of the runs included in the experiment as well as the condition to which ‘Analysis of hPTMs’ was used). MSqRob requires an annotation file that contains the name of each peak in the data. The individual nucleosides were quantified using area-based linear regression analysis per sample: flow through as well as pH 8 and pH 2 elutions. Peptides were grouped to calculate differential proteins. Data were filtered for three valid triplicates were calculated using a Student’s t-test (P < 0.05) and FC > 2. Generation of volcano plots and downstream analysis of proteomics data were performed using R. For data integration of the hPTM and the chromatin proteome, writers and erasers were selected from the Histome database and literature. The quantified epigenetic modifiers and their target modification were integrated using Cytoscape v.3.7.2 (ref. 13).

Acid extractome analysis after histone extraction. Next to histones, other alkaline proteins remain in the HCl during acid extraction. For this purpose, we used the MSqRob software, which was developed for relative protein quantification by implementing the peptide-level robust ridge-regression method. First, the deconvoluted peptide ion data of all histones were exported from Progenesis QIP 4.2 for further analysis.

The hPTM mass spectrometry proteomics datasets have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository (https://www.proteomexchange.org) under the dataset identifier PXD028162 and PXD032792. The project with the identifier PXD028162 (consultable through ProteomeXchange) was licensed on a single-run basis and is fully accessible and editable by the researchers after free download of the Progenesis QIP 4.2 software (https://www.proteomexchange.org). The ChEP mass spectrometry proteomics datasets have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository (https://www.ebi.ac.uk/pride) under the dataset identifier PXD028111. Datasets were downloaded as provided by Petropoulos et al. (ArrayExpress: E-MTAB-39293), Zhou et al. (GEO: GSE109558) and van Merlo et al. (GEO: GSE101675). Public databases used in this manuscript include HumanProteinAtlas database (https://www.proteinatlas.org/), cRAP database (https://www.thegpm.org/crap/), Histome database (https://www.histome.org/), UniProtKB (https://www.uniprot.org/). Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Code availability

Codes pertaining to analyses in this study are available from https://github.com/AndrewAMalcolm/Zijlmans-et-al.-2022 (ref. 112). Codes pertaining to analyses in this study are available from https://github.com/laurabiggins/Shiny_omics (ref. 111). Codes pertaining to analyses in this study are available from https://github.com/AndrewAMalcolm/Pasque-et-al-2022 (ref. 114).

References

75. Guo, G. et al. Epigenetic resetting of human pluripotency. Development 144, 2748–2763 (2017).
76. Bredenkamp, N. et al. Wnt inhibition facilitates RNA-mediated reprogramming of human immune cells to naive pluripotency. Stem Cell Rep. 13, 1083–1098 (2019).
77. Rivron, N. C. et al. Tissue deformation spatially modulates VEGF signaling and angiogenesis. Proc. Natl Acad. Sci. USA 109, 6888–6891 (2012).
78. Vijay, E. J. et al. 3D high throughput screening and profiling of embryoid bodies in thermoformed microwell plates. Lab Chip 16, 734–742 (2016).
79. Pasque, V. et al. X chromosome reactivation dynamics reveal stages of histones’ interactions and degradation. Stem Cell Rep. 13, 1083–1098 (2019).
80. Birney, E. et al. The genome sequence of Drosophila melanogaster: a test case for integrative genomics. Nature 407, 1–20 (2000).
81. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).
82. 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).

Nature Cell Biology | www.nature.com/naturecellbiology
83. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191 (2009).
84. Satija, R., Farahbod, A. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502 (2015).
85. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 377–379 (2012).
86. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
87. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
88. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383 (2015).
89. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).
90. Ramírez, F. et al. deepTools2: a next generation web server for deep sequencing analysis. Nucleic Acids Res. 44, W160–W165 (2016).
91. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and compeBioN: a review of this work. Peer reviewer reports are available.
92. Perino, M. et al. Two functional axes of feedback-enforced PRC2 maintenance, and human disease. Mol. Cell. Biol. 39, 2498–2504 (2019).
93. Meert, P. et al. Tackling aspecific side reactions during histone propionylation: the promise of reversing overpropionylitation. Proteomics 16, 1970–1974 (2016).
94. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 25, 2078–2079 (2009).
95. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
96. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
97. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383 (2015).
98. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).
99. Ramírez, F. et al. deepTools2: a next generation web server for deep sequencing analysis. Nucleic Acids Res. 44, W160–W165 (2016).
100. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and compeBioN: a review of this work. Peer reviewer reports are available.
101. Wiśniewski, J. R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J. Proteome Res. 8, 5674–5678 (2009).
102. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896–1906 (2007).
103. 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).
104. Khare, S. P. et al. Histone-a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res. 40, D337–D343 (2012).
105. Huang, F., Abmayr, S. M. & Workman, J. L. Regulation of KAT6 complex 2 restricts human trophoblast induction. Nat. Biotechnol. 30, W158–W165 (2012).
106. Wu, S., Wang, H., Hsu, J.-L. & Tsai, H.-J. Identification of KAT6a-modified peptides bioinformatic analysis. Nucleic Acids Res. 44, D442–D450 (2016).
107. Van Aller, G. S. et al. Smyd3 regulates cancer cell phenotypes and catalyzes the use of Tapestry and 10X Chromium Controller and in particular G. Bervel for helpful feedback. We thank A. S. Puyvelde for his expertise and support in operating the liquid chromatography with mass spectrometry system. Research in the VP laboratory is supported by The Research Foundation–Flanders (FWO); Olydecx Studentship (grant nos. 1S72719N and 1S29419N to V.P.); the Pandora project (grant nos. 1S72719N, 1S72722N, 1S29419N, 1S75720N and T.X.A.P. (grant nos. 1N1322N. We thank P. J.-R.G. and S. V. for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383 (2015).
108. Malcolm, A. Integrated multi-omics reveal polycomb repressive complex 2 restricts human trophoblast induction. Zenodo https://doi.org/10.5281/ZENODO.6399297 (2022).
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Global analysis of chromatin proteome and transcriptome in naive and primed hPSCs. a. Volcano plot showing differential gene expression as detected by RNA-seq between naive and primed hPSCs (n = 3 biologically independent samples). Core pluripotency factors (black) as well as factors specific to each pluripotent state (blue for primed; green for naive) are highlighted. Dashed lines indicate p-value < 0.05 and log₂fold change > 2.5 (two-sided student’s t test). P-values can be found in Supplementary Table 8. b. Chromatin occupancy of all proteins with a potential role in stem cell maintenance in naive and primed hPSCs (n = 3 biologically independent samples). Protein names were collected from AmiGO (http://amigo.geneontology.org/; ‘stem cell maintenance’ GO:0019827)\(^{38,50}\). The following pluripotency-associated proteins were not detected in our dataset: ASCL2, BMP7, BMPRIA, DAZL, DLL1, ERAS, ESRRB, FANCC, FGFI10, FGFI4, FGFR1, FOXO3, FZD7, HES1, HES5, HESX1, ID1, ID2, ID3, JAG1, KIT, KLF10, KLF2, LBH, LDB2, LIF, LOXL2, LRP5, MCPH1, MED21, MED27, MMP24, MYC, NANO2, NANO52, NOAL, NOG, NOTCH2, NR0B1, NR2E1, PADI4, PAX2, PAX8, PELO, PHF19, PIWIL2, PRDM16, PROX1, PRRX1, PTN, RAFI, SETD6, SFRP1, SFRP1, SIX2, SMO, SOX9, SPI1, TAL1, TBX3, TCF15, TCL1, TERT, TP63, TUT4, WNT7A, WNT9B, ZFP36L2, ZNF322, ZNF358, ZNF706. c. ChIP analysis of differential abundance of chromatin-associated complexes in naive and primed hPSCs (n = 3 biologically independent samples), supplementing Fig. 1d,e. Data are presented as mean values +/− SEM. Dashed lines represent 2-fold change. Asterisks indicate p-value < 0.05 and fold change > 2 (two-sided student’s t test). Low-change proteins (log₂ FC < 0.5) involved in ATP-dependent chromatin remodelling (Fig. 1e) are shown. d. Mass spectrometry analyses of global levels of DNA methylation (green) and DNA hydroxymethylation (purple) in naive and primed hPSCs (n = 3 biologically independent samples) Data are presented as mean values +/− SD. Underlying source data is provided in Source Data Extended Data Fig. 1.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Global analysis of acid extractome and hPTM clipping in naive and primed hPSCs. a. Overview of histone variants identified in acid extractomes. Data are visualized as log2 transformed normalized expression of primed over naive hPSCs. Naive hPSCs, n = 7 biologically independent samples; primed hPSCs, n = 5 biologically independent samples. b. Comparison of proteins identified in the chromatin proteome (n = 4,576 proteins) and acid extractome (n = 894 proteins) in naive and primed hPSCs. Only proteins identified in both conditions were retained (n = 355 proteins). Proteins significantly changing (two-sided student’s t test for chromatin proteome, moderated t test with Benjamini–Hochberg correction for acid extractome, p-value < 0.05, > 2-fold change) in both datasets are indicated with red dots, while proteins significantly changing in only one dataset are highlighted in green dots. Strongest changing proteins (p-value < 0.05 & > 4-fold change in one dataset) are labelled by name. Blue line indicates best fit linear regression, while the shaded grey area indicates 95% confidence interval. c. Quantification of proteins uniquely identified in the acid extractome of naive and primed hPSCs (n = 539 proteins). Significantly changing proteins (moderated t test with Benjamini–Hochberg correction, p-value < 0.05 & >2-fold change) are indicated with red dots. The 15 most strongly changing proteins for each pluripotent state are labelled by name. d. Quantification of ribosomal and nucleolar proteins identified in the acid extractome of naive and primed hPSCs (n = 128 proteins). Significantly changing proteins (moderated t test with Benjamini–Hochberg correction, p-value < 0.05 & >2-fold change) are indicated with red dots and labelled by name. e. Log2 transformed abundance of H3 tail clipping events (normalized against all histone peptidoforms) as identified by mass spectrometry. The boxplots show the interquartile range (box limits) and the median (centre line) of the abundance of clipping events. Naive hPSCs, n = 7 biologically independent samples; primed hPSCs, n = 5 biologically independent samples. f. Comparison of chromatin-associated proteins uniquely identified in human between naive and primed pluripotent states. The listed proteins have no known mouse ortholog or homologue. Underlying source data is provided in Source Data Extended Data Fig. 2.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Chromatin profiling of H3K27me3 in naive and primed hPSCs. a. Schematic of the cCUT&RUN method and normalization strategy. b. Scatterplots comparing the log2 transformed H3K27me3 normalized read count across 1 kb windows for naive (left) and primed (right) hPSCs. Correlation r is determined by Pearson correlation prior to transformation. 2 biologically independent experiments for primed and naive H3K27me3 and IgG calibrated CUT&RUNs, with the exception of naive IgG cCUT&RUN which was performed once. c. Principal Component Analysis of normalized cCUT&RUN data. d. Violin plots showing the H3K27me3 peak width of normalized cCUT&RUN data in naive and primed hPSCs. The difference in width between primed and naive hPSCs is statistically significant, indicated by * (two-sided student’s t test, p < 2.2×10^-16; n = 10,187 peaks for naive PSCs and n = 17,626 peaks for primed PSCs). e. Venn diagram showing the extent of overlap of nearest genes (within 10 kb of promoter) to H3K27me3 peaks in naive and primed hPSCs, with example genes added for each category. f. Stacked bar plot describing genomic annotation of H3K27me3 peaks in untreated naive and primed hPSCs, compared to a background sample of 10,000 randomly generated peaks. g. Gene expression values of primed-specific (DUSP6 and SFRP2) and naive-specific (KLF4 and TFCP2L1) transcripts from bulk RNA-sequencing data. Individual data points from n = 3 biologically independent samples are shown, and the bar indicates the mean expression value. h. Violin plots show the fold change in gene expression between naive and primed hPSCs for naive-specific and primed-specific H3K27me3-marked genes from bulk RNA-sequencing data. The plots show the 25th and 75th quartiles (white dotted lines) and the median (black dashed lines). Underlying source data is provided in Source Data Extended Data Fig. 3.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | Four-day PRC2 inhibition by UNC1999 efficiently removes H3K27me3, with minor alterations in chromatin proteome and hPTM landscape. a. Western blot validation of H3K27me3 removal in naive and primed hPSCs upon treatment with UNC1999 (Data shown represent 3 biologically independent samples [AU: modified as “n = X” statement is retained for cases where statistics are derived]). b. Histone PTM quantification for H4 in naive and primed hPSCs, with and without PRC2i. Data are visualized as log2 fold changes between two conditions, which are listed on top of each panel. Data are ordered according to the first panel. Red bars indicate significantly changing hPTMs (two-sided student’s t test with Benjamini–Hochberg correction, p-value < 0.05). Naive hPSCs, n = 7 biologically independent samples; primed hPSCs, n = 5 biologically independent samples; naive hPSCs + inhibitor n = 6 biologically independent samples; primed hPSCs + inhibitor, n = 8 biologically independent samples. c–d. Integration of the chromatin proteome and hPTM measurements for naive (c) and primed (d) hPSCs with and without PRC2i. Nodes represent chromatin modifiers and hPTMs, and are coloured by log2 fold change in abundance. Edges indicate functional connection (write or erase) between the nodes. Chromatin modifiers in grey nodes were not detected. e. Metaplots showing average profiles of normalized H3K27me3 reads across peaks, with relative abundance and distribution within 25 kb either side of the peak centre for primed-enriched, common and naive-enriched peaks in naive and primed hPSCs, with or without PRC2i. Two biologically independent experiments were used for primed and naive H3K27me3 and IgG cCUT&RUN experiments, both with and without PRC2i, with the exception of naive IgG cCUT&RUN which was performed once. Non-inhibitor treated samples are replicated from Extended Data Fig. 3. f. Specific changes in chromatin-associated proteins induced by PRC2i in naive hPSCs. Volcano plot of chromatin-associated proteins (n = 3,784 proteins) quantified in n = 3 biologically independent samples for naive hPSCs with or without PRC2i. Significantly changing (two-sided student’s t test, p-value < 0.05, >2-fold change) proteins are indicated with red dots and labelled by name. g. Numbers of significantly changing chromatin proteins (two-sided student’s t test, p-value < 0.05 & >2-fold change) between naive and primed hPSCs and after PRC2i. N = 3 biologically independent samples. h. Global DNA methylation (green) and DNA hydroxymethylation (purple) levels in naive hPSCs, with and without PRC2i (n = 3 biologically independent samples). Data are presented as mean values +/- SD. i. Flow cytometry analysis of cell viability (left) and state-specific pluripotency markers (right) in naive and primed hPSCs treated with and without UNC1999 for four days. Cell viability was assessed using a live-dead dye, and the values shown represent the percentage of live cells in the total cell population. For the protein marker analysis, naive hPSCs were assayed for the expression of cell-surface markers SUSD2 and CD75, and primed hPSCs for the cell-surface markers SSEA4 and CD24. The values shown are the percentage of double-positive cells (SUSD2 and CD75, or SSEA4 and CD24) out of the total population of live cells. N = 3 biologically independent samples. j. Immunofluorescence analysis for KLF17 (magenta), NANOG (green) and DAPI (blue) in naive hiPSCs after 8 days with or without PRC2i in PXGL medium. Right panel: quantification of KLF17 and/or NANOG positive colonies. Images are representative of 2 experiments. Scale bar = 100 mm. Underlying source data is provided in Source Data Extended Data Fig. 4.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | PRC2 inhibition promotes naive to trophoblast induction. a. Immunofluorescence analysis of trophoblast and pluripotency markers with or without PRC2 inhibition. Naive hiPSCs were stained for GATA3 (magenta), NANOG (green) and DAPI (blue) at day 4 of naive to trophoblast conversion with PRC2 inhibition (PRC2i) during the 4 days preceding trophoblast conversion only (Naive + PRC2i → Trophoblast), or during trophoblast conversion only (Naive → Trophoblast + PRC2i), or throughout the experiment (Naive + PRC2i → Trophoblast + PRC2i). Representative images from 3 experiments. Scale bar = 100 mm. b. Proportion of naive hiPSCs with GATA3 expression using different thresholds of scRNA-seq data counts to deem a cell GATA3 positive after 4 days of PRC2i in PXGL or in DMSO control conditions. c. Immunofluorescence analysis of naive hPSC with and without PRC2i. Cells were stained for GATA3 and DAPI at day 4 (H9 hPSCs; n = 1,925 cells for day 4 Naive; n = 2,025 cells for day 4 Naive + PRC2i, in two experiments) and day 8 (ICSIG-1 hiPSCs; n = 496 cells for day 8 Naive; n = 881 cells for day 8 Naive + PRC2i, in one experiment). Middle panel shows representative images for day 4 (H9 hPSCs). Left and right panels show quantification of GATA3+ nuclei for day 4 and day 8, respectively. Scale bar = 200 mm. d. Flow cytometry analysis of TROP2 and SUSD2 in naive hiPSCs at 4 and 8 days with or without PRC2i treatment. This experiment was performed once. e. Quantification of the proportion of colonies expressing or lacking NANOG protein expression in 2 experiments as evaluated by immunofluorescence. The number of colonies counted is shown on top of the panels. f. Phase contrast images of representative cells at day -4 (d -4), 0 (d0) and 4 (d4) of naive hiPSC to trophoblast conversion. Representative images of 3 experiments. g. Cell viability (left) and cell number (right) during naive hiPSC to trophoblast conversion with and without PRC2i. Cells have been pretreated for 4 days with PRC2i. N = 3 biologically independent samples. h. RT-qPCR assay for trophoblast marker VGLL1. Expression was normalized to GAPDH. Squares, triangles and circles represent n = 3 biologically independent samples, each with averaged biological triplicates or duplicates in each experiment. A.U. = arbitrary unit. Two-sided t test with Bonferroni adjustment, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. i. Flow cytometry analysis of TROP2 and SUSD2 markers during naive hiPSC to trophoblast conversion with and without PRC2i treatment or pre-treatment at day 4. Day 12 trophoblast cells were included as a positive control. This experiment was performed once. j. Quantification of immunofluorescence analysis of H9 naive hPSC to trophoblast conversion with and without PRC2i from 2 experiments. Cells were stained for GATA3 at day 0 and day 4. k. RT-qPCR assay for pluripotency-associated genes, DPPA5, SOX2, KLF17 and NANOG (normalized to GAPDH) at day 0 and 4 of naive hiPSC to trophoblast conversion with and without PRC2i pre-treatment, treatment during trophoblast conversion alone, or PRC2i throughout the experiment. Squares, triangles and circles represent independent experiments (n = 3 experiments). For each experiment, the average of 3 or 2 biologically independent samples is shown. A.U. = arbitrary unit. Two-sided t test with Bonferroni adjustment, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. l. Immunofluorescence analysis of primed hPSC to trophoblast conversion with and without PRC2i. Cells were stained for GATA3, NANOG and DAPI at day 0, day 4 and day 10. Right panel shows quantification of GATA3+ and NANOG+ nuclei. Representative images of 2 experiments. Underlying source data is provided in Source Data Extended Data Fig. 5.
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Integration with human embryo data.  

**a.** Violin plots with single cell expression distributions combined with boxplots in naive hiPSCs with and without PRC2i of pluripotency-associated genes. Data are visualized as log(nUMI). The boxplots show the interquartile range (box limits) and median (centre line) of gene expression levels. Number of single cells measured: n = 2,903 cells for the naive sample, and n = 3,338 cells for the naive + PRC2i sample.  

**b.** scRNA-seq analysis of pluripotency-associated and trophoblast-associated genes (UMAPs). Data are visualized as log normalized counts. Darker red intensity represents higher levels of gene expression, while lower red represents lower gene expression levels.  

**c.** scRNA-seq analysis showing the 20 most differentially expressed genes between the naive, intermediate and trophoblast cell clusters. Point size represents the proportion of cells in the cluster with the indicated gene expression enrichment. Data are visualized as average expression scale. Darker red intensity represents higher levels of gene expression, while softer red represents lower gene expression levels.  

**d.** Proportion of day 4 converted cells with intermediate cell identity. Purple indicates the day 4 trophoblast and red indicates the day 4 + PRC2i samples.  

**e.** Single-cell UMAP representation comparing in vitro day 4 trophoblast and day 4 trophoblast + PRC2i with human pre-implantation\(^1\) and post-implantation\(^7,4\) by data integration. Annotations from\(^13\).  

**f.** Single-cell UMAP representation of pluripotency, trophoblast and primitive endoderm marker genes from data in **e.** Data are visualized as log normalized counts. Underlying source data is provided in Source Data Extended Data Fig. 6.
Extended Data Fig. 7 | See next page for caption.
Extended Data Fig. 7 | Human blastoids. a–b. Boxplots with the immunofluorescence quantification at 36 h (a) and 60 h (b) blastoids with or without PRC2i. Blastoids were stained for NANOG, GATA3 and FOXA2. The boxplots show the interquartile range (box limits) and median (centre line) of the total number of positive cells per blastoid (left panel). The right panel indicates the total number of cells per blastoid. N = 21 blastoids for 36 h -PRC2i blastoids; n = 23 blastoids for 36 h + PRC2i; n = 27 blastoids for 60 h -PRC2i; and n = 17 blastoids for 60 h + PRC2i, quantified from n = 1 experiment. A two-sided Wilcoxon rank-sum test with Bonferroni correction was used for significance testing. 36 h: ****P = 1 × 10⁻⁶; 60 h: ****P = 3.9 × 10⁻⁵, **P = 1.6 ×10⁻³. c. Close-up for average number of FOXA2 + cells as found in (b). N = 27 blastoids for -PRC2i, and n = 17 blastoids for +PRC2i, quantified from n = 1 experiment. PrEnd = Primitive Endoderm. d. Representative immunofluorescence images for Fig. 7c and Extended Data Fig. 7a–b. NANOG is shown in green, GATA3 in magenta and FOXA2 in yellow. Scale bar: 200 μm. The experiment was performed once. e. Immunofluorescence analysis of human blastoids with and without PRC2i. Cells were stained for AQP3 (white) and NANOG (green) after 24 h. Representative image from 1 experiment. Scale bar: 200 μm. f. Quantification of cavitated human blastoids after 36 and 60 h with and without PRC2i. N = 3 biologically independent samples. g. Quantification of immunofluorescence in naïve human pluripotent stem cells cultured in PXGL treated with PRC2i EPZ-6438 or UNC1999 for 7 days. Cells were stained for GATA3 and GATA4. The experiment was performed once. Underlying source data is provided in Source Data Extended Data Fig. 7.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a  Confirmed

☐  The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐  A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐  The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐  A description of all covariates tested

☐  A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐  A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) and variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐  For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

☐  For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐  For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐  Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on Statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection  CUT&RUN (SEACR) algorithm (v1.3)

Data analysis  Flowllo V10.1 software

Nikon NIS-Elements AR with DeNoise AI and 3D Richardson-Lucy algorithm (v5.30.01)
Axio Vision software (v4.9.1.0)
ImageJ (v1.37a)

Microsoft Excel (Microsoft Office Professional Plus 2016)
R (v4.0.3)

ggplot2 (v3.3.3)
FastQC (v0.11.8; Babraham Bioinformatics)
STAR (v2.7.1a)

featureCounts (v2.0.1)
DESeq2 (v1.26.0)
10X Genomics Cell Ranger (v4.0.0)
biomart (v2.64.3)
Seurat (v4.0.1)

TTrimGalore (v0.6.6; Babraham Bioinformatics)
Bowtie2 (v2.3.2)
DESEQ2 implementation within SeqMonk (v1.47.2; Babraham Bioinformatics)
ChIPseeker (v1.30.3)
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data generated during this study has been deposited in public repositories as follows:

The multi-omic data can be explored using the online resource: https://www.bioinformatics.babraham.ac.uk/shiny/shiny_omics/Shiny_omics

Figures 1, 2, 3, 4 and Extended Data Figures 1, 2, 3, 4.

Raw and processed sequencing data for RNA-seq, cUTR&RUN and scRNA-seq (including scRNA-seq lom files) to be visualised on the SCope platform: https://scope.aerslab.org/#/HumanPluripotencyPRC2/*/welcome have been submitted to the NCBI GEO [http://www.ncbi.nlm.nih.gov/geo/] under accession number GSE176175.

Figures 3, 4, 5, 6 and Extended Data Figures 3, 4, 5, 6.

Raw and processed sequencing data generated in this study (CheP mass spectrometry proteomics data):

PRIDE partner repository, https://www.ebi.ac.uk/pride/

[dataset identifier PXD028111]

Figures 1, 2, 4 and Extended Data Figures 1, 2, 4.

Raw and processed sequencing data generated in this study (hPTM mass spectrometry proteomics):

PRIDE partner repository, https://www.ebi.ac.uk/pride/

[dataset identifier PXD028162, 10.6019/PXD028162 and PXD032792]

Figures 2, 4 and Extended Data Figures 2, 4.

Public databases/datasets used in this manuscript:

Human Swissprot database [https://www.uniprot.org/]
cRAP database [https://www.theogm.org/crap/]
Histome database [http://www3.iser.pune.ac.in/~cgee/histome/]
UniProtKB human proteome [https://www.uniprot.org/]

ArrayExpress under accession number E-MTAB-39293 (Petropoulos et al., 2016)

GEO database under accession number GSE10955589 (Zhou et al., 2019)

van Mierlo et al., 2019; PRIDE partner repository, https://www.ebi.ac.uk/pride/ [dataset identifier PXD007154].

DATA AVAILABILITY

The multi-omic data can be explored using the online resource: https://www.bioinformatics.babraham.ac.uk/shiny/shiny_omics/Shiny_omics. The scRNA-seq lom files can be visualised on the SCope platform: https://scope.aerslab.org/#/HumanPluripotencyPRC2/*/welcome.

RNA sequencing, cUTR&RUN, and single-cell RNA sequencing datasets have been deposited in the Gene Expression Omnibus (GEO) under the accession code of GSE176175.

hPTM mass spectrometry proteomics datasets have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository [https://www.ebi.ac.uk/pride/] under the dataset identifiers PXD028162, 10.6019/PXD028162 and PXD032792. The project with identifier PXD028162 (consultable via ProteomeXchange) was licensed on a single run basis and is fully accessible and editable by the readership after free download of the Progenesis QI 4.2 software [https://www.nonlinear.com/progenesis/qi-for-proteomics/]. CheP mass spectrometry proteomics datasets have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository [https://www.ebi.ac.uk/pride/] under the dataset identifier PXD028111. Furthermore, datasets were downloadable as provided by Petropoulos et al. (ArrayExpress: E-MTAB-39293), Zhou et al. (Zhou et al., 2019) (GEO: GSE10955580) and van Mierlo et al. (GEO: GSE101675). Public databases used in this manuscript include Human Swissprot database [https://www.uniprot.org/], cRAP database [https://www.theogm.org/crap/], Histome database [http://www3.iser.pune.ac.in/~cgee/histome/], UniProtKB human proteome [https://www.uniprot.org/]. Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

Life sciences study design

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

Sample size
- All sample size are indicated on the figure legends or in the manuscript.
- In stem cells, qPCR, flow cytometry, western blotting and immunostaining, experiments were performed in two or three independent replicates unless stated in figure legends.
- In blastoids experiments, flow cytometry and cavity formation experiments were performed in three independent experiments.
- Immunofluorescence in blastoids were performed in several blastoids from one independent experiment.
- In bulk RNAseq, ChEP, and CUT&RUN experiments were performed in two or three biological replicates.
- hPTM mass spectrometry was performed for at least five biological replicates per condition.
- No statistical method was used to determine sample size. Sample sizes were chosen based on previous experience from similar studies.

Data exclusions
- For ChEP, proteins flagged as ‘reverse’, ‘potential contaminant’ or ’only identified by site’ were filtered from the final protein list. Biological triplicates were grouped to calculate differential proteins. Data were filtered for 3 valid values in at least 1 group. Missing values were imputed using default settings in Perseus, based on the assumption that they were not detected because they were under or close to detection limit.
- For bulk RNAseq analysis samples were filtered to keep genes that had more than 1 count in at least 2 conditions.
- For single-cell RNAseq analysis human cells were retained and mouse cells (MEFs) were filtered out by adjusting the number of counts per cell (nCount_RNA) and the number of mapped genes per cell (nFeature_RNA) to only keep cells that were mostly mapped to the human GRCh38 (hg38) genome (for naive cells: nCount_RNA < 40000, nCount_RNA > 3000, nFeature_RNA < 8000 and nFeature_RNA > 1500; for day 4 of naive to trophoblast conversions: nCount_RNA < 300000, nCount_RNA > 10000, nFeature_RNA < 12000 and nFeature_RNA > 3000). Naive cells with more than 25% of mitochondrial counts were filtered out. Day 4 trophoblast converted cells with more than 30% of mitochondrial counts were filtered out.
- For the hPTM analysis, outliers were removed first on the basis of the normalisation factor in Progenesis QI 4.2: a normalisation factor less than 0.5 or more than 2 were filtered out. Secondly, outliers were removed based on the principal component analysis (PCA).
- For calibrated CUT&RUN analysis, low quality reads with a MAPQ value > 20 were removed by filtering with samtools view.
- All exclusion criteria used in this manuscript were pre-established.

Replication
- All of the main experiments listed on this manuscript have been successfully replicated at least twice as independent experiments. When this is not the case, this has been indicated in figure legends.

Randomization
- For scRNA-seq we used canonical correspondence analysis to remove potential technical/sample-to-sample effects and variation.
- For the mass spectrometry experiments, all samples were run in a randomized fashion. Sample allocation was random in all other experiments.

Blinding
- ChEP, Acid Extractome, RNA-seq, Western blot, scRNA-seq, imaging
- Samples from naive versus prime hPSCs, untreated or treated with UNC1999 inhibitor, were used, hence the analyses could not be performed with blinding. No blinding was performed because none of the analyses reported involved procedures that could be influenced by investigator bias.

Imaging
- Quantification of immunostaining experiments were performed by different researchers to confirm scoring.

For blastoids lineage analysis, counting of nuclei was performed blindly using automated nuclei scoring based on fluorescence using NIS-Elements AR 5.30.01 via a GA3 script.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.
### Antibodies

**Antibodies used**

| Antibodies                                                                                       | Quantity | Catalogue Number | Reference |
|--------------------------------------------------------------------------------------------------|----------|------------------|-----------|
| Anti-NANOG Mouse, 1 in 50, BD Pharmingen, 560482 (clone N31:355)                                |          |                  |           |
| Anti-mouse-Alexa Fluor 488; 555; 647 Donkey, 1 in 500, Invitrogen, A21202; A31570; A31571         |          |                  |           |
| Anti-klf17 Rabbit, 1 in 333, Abcam, HPA0246329; Anti-rabbit-Alexa Fluor 647 Donkey, 1 in 500, Invitrogen, A31573 |          |                  |           |
| Anti-GATA3 Rabbit, 1 in 250, Abcam, ab199428 (clone EPR16651); Anti-rabbit Alexa Fluor 647 Donkey, 1 in 500, Invitrogen, A31573 |          |                  |           |
| Anti-GATA3 Rat, 1 in 100, Thermo Fisher, 14-9966-82 (clone TWAJ); Anti-rat-Alexa Fluor 488 Donkey, 1 in 500, Invitrogen, A21208 |          |                  |           |
| Anti-AQP3 Rabbit, 1 in 200, Antibodies-online.com, ABIN863208; Anti-rabbit-Alexa Fluor 488 Donkey, 1 in 500, Invitrogen, A21208 |          |                  |           |
| Anti-fox2a Goat, 1 in 40, R&D Systems biotechne, AF2400; Anti-goat-Alexa 546 Donkey, 1 in 500, Invitrogen, A11056 |          |                  |           |
| Anti-4a4 Rat, 1 in 400, ebioscience, 14-9980-82 (clone eBioEv); Anti-rat-Alexa Fluor 488 Donkey, 1 in 500, Invitrogen, A21208 |          |                  |           |

**Flow Cytometry:**

- Anti-TROP2 Mouse, 1 in 100, R&D Systems, MA6500 (clone 77220); Anti-mouse Alexa Fluor 488 Donkey, 1 in 500, Invitrogen, A21202
- Anti-TROP2-488 Mouse, 1 in 25, R&D systems, FAB650X (clone 77220)
- Anti-SUSD2-APC Mouse, 1 in 50, Miltenyi Biotec, 130-121-134 (clone W5CS)
- Anti-SUSD2-PE Mouse, 1 in 100, Miltenyi Biotec, 130-111-641 (clone REA795)
- Anti-CD75-eF660 Mouse, 1 in 20, ebioscience, 15519896 (clone LN-1)
- Anti-4a4-APC Mouse, 1 in 40, R&D Systems, FAB1435A (clone MC-813-70)
- Anti-CD24-BUV395 Mouse, 1 in 40, BD Bioscience, 563818 (clone ML5)

**Calibrated CUT&RUN:**

- Tri-Methyl-Histone H3 Lysine 27 Rabbit, 1 in 50, Cell Signaling Technology, 9733 (clone C36811)
- Anti-IgG Rabbit, 1 in 50, Invitrogen, 31188
- Anti-Histone H2Av spike-in antibody, 1 in 50, Active Motif, 61686
- Western Blotting:
  - Anti-Tri-Methyl-Histone H3 lysine 27 Rabbit, 1 in 10000, Cell Signaling Technology, 9733 (clone C36811); Anti-Rabbit IgG (H+L)-DyLight-800 Donkey, 1 in 10000, Invitrogen, S5-1044
  - Anti-Rabbit IgG (H+L)-HRP Goat, 1 in 10000, Bio-Rad, 170651

**Validation**

- Anti-NANOG Mouse, 1 in 400, BD Pharmingen, 560482 (clone N31:355)
- Anti-mouse-Alexa Fluor 488; 555; 647 Donkey, 1 in 500, Invitrogen, A21202; A31570; A31571
- Anti-klf17 Rabbit, 1 in 200-1 in 500, Abcam, HPA0246329
- Anti-rabbit-Alexa Fluor 647 Donkey, 1 in 200-1 in 2000, Invitrogen, A31573
- Anti-GATA3 Rabbit, 1 in 250, Abcam, ab199428 (clone EPR16651)
- Anti-GATA3 Rat, 1 in 100, Thermo Fisher, 14-9966-82 (clone TWAJ)
- Anti-rat-Alexa Fluor 488 Donkey, 1 in 2000, Invitrogen, A21208
- Anti-AQP3 Rabbit, 1 in 200, Antibodies-online.com, ABIN863208
- Anti-rabbit-Alexa Fluor 488 Donkey, 1 in 1000, Invitrogen, A21206
- Anti-fox2a Goat, 1 in 13-3:1 in 40, R&D Systems biotechne, AF2400
- Anti-goat-Alexa 546 Donkey, 1 in 200-1 in 2000, Invitrogen, A11056
- Anti-4a4 Rat, 1 in 200-1 in 400, ebioscience, 14-9980-82 (clone eBioEv)

**Flow Cytometry:**

- Anti-TROP2 Mouse, 0.25 μg/10^6 cells, R&D Systems, MA6500 (clone 77220)
- Anti-mouse-Alexa Fluor 488 Donkey, no validation for Flow Cytometry, Invitrogen, A21202
- Anti-TROP2-488 Mouse, 10 μL/10^6 cells, R&D systems, FAB650X (clone 77220)
- Anti-SUSD2-APC Mouse, 1 in 50, Miltenyi Biotec, 130-121-134 (clone W5CS)
- Anti-SUSD2-PE Mouse, 1 in 50, Miltenyi Biotec, 130-111-641 (clone REA795)
- Anti-CD75-eF660 Mouse, 5μl/Test, ebioscience, 15519896 (clone LN-1)
- Anti-4a4-APC Mouse, 10 μL/10^6 cells, R&D Systems, FAB1435A (clone MC-813-70)
- Anti-CD24-BUV395 Mouse, 5μl/Test, BD Bioscience, 563818 (clone ML5)

**Calibrated CUT&RUN:**

- Tri-Methyl-Histone H3 Lysine 27 Rabbit, 1 in 50, Cell Signaling Technology, 9733 (clone C36811)
Eukaryotic cell lines

Policy information about cell lines

**Cell line source(s)**
- H9 hESCs [iCell]
- ICSC: 1 iPSC0028 hiPSCs (Sigma)
- Wild-type iPSC0028 hiPSCs (Sigma)
- Male mouse embryonic fibroblasts (MEFs) isolated from wild-type mouse B6 embryos
- Drosophila S2 cells (ThermoFisher Scientific)

**Authentication**
All cell lines in this study were authenticated via gene expression analysis.

**Mycoplasma contamination**
Periodic mycoplasma contamination testing was carried out and confirmed the absence of mycoplasma contamination.

**Commonly misidentified lines**
No commonly identified cell lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

**Laboratory animals**
Wild-type male mouse embryos C57Bl/6J X 14.5. Animals were kept with a maximum of 5 animals per cage and separated by sex. Illumination was controlled on a 14h light, 10h dark light cycle from 7h to 21h. Temperature was checked daily and should be 22±2°C. Humidity in mouse rooms was checked daily and should be between 45-70% but can vary with weather conditions, especially in winter.

**Wild animals**
No wild animals were used in this study.

**Field-collected samples**
No field-collected samples were used in this study.

**Ethics oversight**
Animal work carried out in this study was covered by project licences [ECD_P003-2016 and ECD_P170/2019 to V.P and to F.L., respectively] approved by the KU Leuven Animal Ethics Committee.

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

ChIP-seq

**Data deposition**
- 
- Confirm that both raw and final processed data have been deposited in a public database such as GEO.
- 
- Confirm that you have deposited or provided access to group files (e.g. BED files) for the called peaks.

**Data access links**
May remain private before publication.

Calibrated CUT&RUN data provided in GEO (http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE19675.

http://ftp1.babraham.ac.uk/ftp/asr/79/Files in database submission

| GSM5569915 | Primed_H3K27me3_Replicate_1 |
|-------------|----------------------------|
| GSM5569916 | Primed_H3K27me3_Replicate_2 |
| GSM5569917 | Primed_UNC1999_H3K27me3_Replicate_1 |
| GSM5569918 | Primed_UNC1999_H3K27me3_Replicate_2 |
| GSM5569919 | Primed_lgG_Replicate_1 |
| GSM5569920 | Primed_lgG_Replicate_2 |
| GSM5569921 | Primed_UNC1999_lgG_Replicate_1 |
| GSM5569922 | Primed_UNC1999_lgG_Replicate_2 |
| GSM5569923 | Naive_H3K27me3_Replicate_1 |
| GSM5569924 | Naive_H3K27me3_Replicate_2 |
| GSM5569925 | Naive_UNC1999_H3K27me3_Replicate_1 |
| GSM5569926 | Naive_UNC1999_H3K27me3_Replicate_2 |
| GSM5569927 | Naive_lgG_Replicate_1 |
| GSM5569928 | Naive_UNC1999_lgG_Replicate_1 |
| GSM5569929 | Naive_UNC1999_lgG_Replicate_2 |
Methodology

Replicates
Two independent biological replicates for all samples except for naïve IgG which has one biological replicate. Calibrated CUT&RUN performed for H3K27me3 for experimental samples and IgG, with an anti-Drosophila H2Av spike-in antibody, as a control used for peak calling. Strong replicate agreement between samples of the same condition - Pearson correlation coefficient of 0.76 between two naïve H3K27me3 samples and 0.98 for two primed H3K27me3 samples.

Sequencing depth
Total number of sequencing reads: 483,205,175. Highest read count = 53,117,572, Lowest read count = 70,861,009, Average read count = 32,213,678. Uniquely mapping human reads = 275,974,550. Highest uniquely mapping read count = 35,268,030, Lowest uniquely mapping read count = 6,450,381. Average uniquely mapping read count = 18,398,303. Samples were sequenced as 150 bp paired-end sequencing.

Antibodies
Tri-Methyl-Histone H3 Lysine 27 - Cell Signaling Technology, 9733 - (C36811) Rabbit mAb #9733
Rabbit anti-Mouse IgG (H+L) Secondary Antibody - Invitrogen, 31188 - # 31188
Anti-Histone H2Av spike-in antibody - Active Motif, 61686 - #61686

Peak calling parameters
Peak calling was performed using the CUT&RUN optimised Sparse Enrichment Analysis for CUT&RUN (SEACR) algorithm, using H3K27me3 samples as the sample and their corresponding IgG as a control, against which peaks were called. The top 1% of peaks were retained (v1.3). Peaks closer than 3000bp were merged using bedtools merge and peaks common to both replicates were determined by bedtools intersect to generate final peak sets for naïve and primed. Peaks called in naïve and primed hPSCs data sets were concatenated into a combined peak list and deduplicated. Differentially enriched peaks between naïve and primed hPSCs were then determined from this concatenated list using a DESEQ2 implementation within SeqMonk (v1.47.2; Babraham Bioinformatics) to identify differential regions with a p-value < 0.05 after Benjamini-Hochberg Multiple-Testing Correction. Common peaks were classified as peaks in the concatenated list that were not statistically-enriched in either condition. These peaks were then filtered against the ENCODE GRCh38 exclusion list to remove coverage outliers.

Data quality
Raw FastQ data were trimmed with Trim Galore and aligned to GRCh38 human genome or Drosophila BG6-P6 genome using Bowtie2 with the following parameters - very-sensitive -I 10 -X700. High quality reads with a MAPQ value > 20 were retained by filtering with samtools view. Peak calling was performed using the Sparse Enrichment Analysis for CUT&RUN (SEACR) algorithm and the top 1% peaks were retained. Differentially enriched peaks were determined from a concatenated peak list using a DESEQ2 implementation within SeqMonk to identify differential regions with a p-value < 0.05 after Benjamini-Hochberg Multiple Testing Correction to ensure a false discovery rate of less than 5%.

Software
Trim Galore (v0.6.6: Babraham Bioinformatics)
Bowtie2 (v2.3.2)
samtools (v1.11)
bedtools (v2.29.2)
Sparse Enrichment Analysis for CUT&RUN (SEACR) algorithm (v1.3)
DESEQ2 package within SeqMonk (v1.47.2; Babraham Bioinformatics)
bedtools API suite
Python (v3.7.3)
Chipseeker package within R (v1.30.3:93)
HOMER
deeptools bamCoverage (v3.43)
UCSC-tools bigWigMerge
UCSC-tools bedGraphToBigWig
Washu epigenome browser (v5)
deeptools computeMatrix
deeptools MultiBamSummary
deeptools plotCorrelation
ggplot2 package within R

Flow Cytometry

Plots

Confirm that:
☒ The axis labels state the marker and fluorochrome used (e.g, CD4-FITC).
☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
☒ All plots are contour plots with outliers or pseudocolor plots.
☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
For Figure 7, blastoids were harvested from the microwell arrays, sequentially treated with 300 U/ml collagenase type IV and 10x Trypsin-EDTA (Thermofisher) at 37C on a shaker. The blastoids were dissociated into single cells by pipetting. Cells were washed 3 times with flow buffer (1% FBS in PBS) and incubated with primary antibody diluted in flow buffer and incubated...
for 30 minutes at 4°C. The cells were centrifuged, washed with flow buffer twice, and combined with secondary antibody and incubated for 30 minutes at 4°C. The cells were centrifuged, washed twice with flow buffer, and resuspended in fresh flow buffer for flow cytometry analysis.

For Figure 4, naive and primed hPSCs were washed once with PBS and dissociated using Accutase (BioLegend) by incubation for 5 min at 37°C. Accutase was quenched 1:1 with medium and cells were passed through a 50 μm cell strainer (VWR) and centrifuged at 300 x g for 3 minutes. Cell pellets were washed once with PBS containing 2% FBS (flow buffer) and counted. Fluorophore-conjugated antibodies and eFlour 780 fixable viability dye (eBioscience, 65-0865-14) were mixed with 50 μl Brilliant stain buffer (BD Biosciences) and applied to 500,000 cells in 50 μl flow buffer. Labelling occurred for 30 min at 4 °C in the dark. Cells were washed twice with flow buffer and analysed.

For Figure 5, cells were dissociated using Accutase (5 min incubation at 37°C) and centrifuged at 1000 rpm for 5 min. Supernatant was removed and cell pellets were resuspended in 300 ul of FACS buffer supplemented with 0.25-0.5% BSA buffer per sample and centrifuged again in same conditions. Cell pellets were resuspended in 50 ul of FACS buffer and antibody incubations were carried out at 4°C for 30 min. Afterwards, cells were washed twice with 300 ul of FACS buffer and centrifuged (1000 rpm for 5 min). Supernatant was removed and pellet was resuspended in 300 ul of PBS with 4% PFA. Samples were analysed using a BD FACSCanto II Flow Cytometer.

| Instrument          | BD FACSCanto II Flow Cytometer  |
|---------------------|---------------------------------|
| Software            | FlowJo software                 |
| Cell population abundance | No cell populations were sorted in this work. |
| Gating strategy     | Single-stained controls were used for compensation calculations and unstained cells were used in cytometer and gating setup. |

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