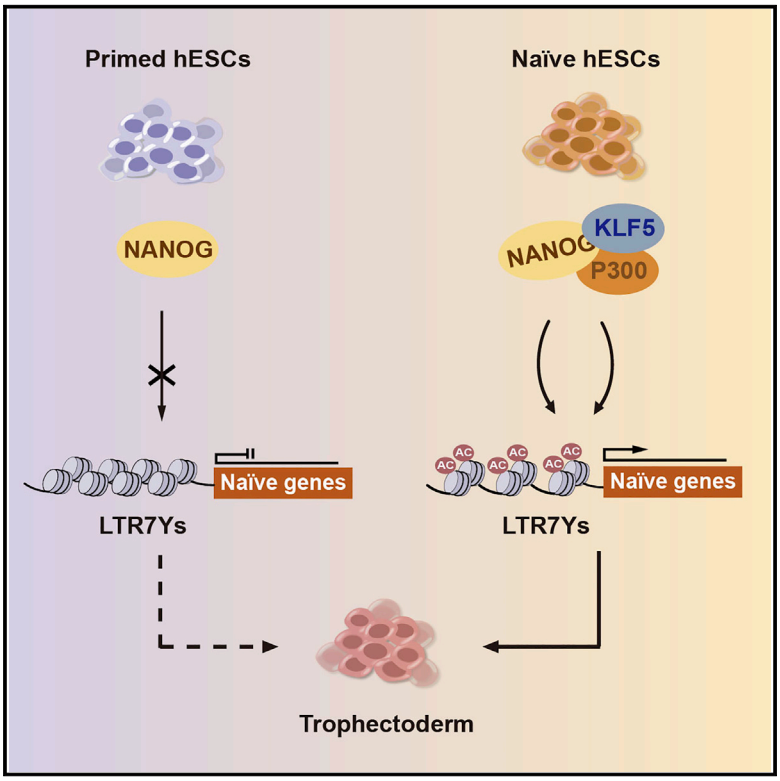
Cell Reports

Article

Krüppel-like factor 5 rewires NANOG regulatory network to activate human naive pluripotency specific LTR7Ys and promote naive pluripotency

Graphical abstract



Highlights

- Human naive pluripotency enriched LTR7Ys are activated by KLF4 and KLF5
- KLF4 and KLF5 have both overlapping and distinct gene and TE targets
- KLF5 rewires NANOG binding affinity from LTR7s to LTR7Ys
- KLF5-NANOG activates LTR7Ys to boost naive gene expression and expands cell potency

Authors

Zhipeng Ai, Xinyu Xiang, Yangquan Xiang, ..., Yun-Shen Chan, Wanlu Liu, Hongqing Liang

Correspondence

zhangdan@zju.edu.cn (D.Z.), chan_yunshen@gzlab.ac.cn (Y.-S.C.), wanluliu@intl.zju.edu.cn (W.L.), lianghongqing@zju.edu.cn (H.L.)

In brief

Endogenous retrovirus (ERV) elements are highly activated during early human embryonic development, but their regulation and function during embryogenesis are largely unclear. Ai et al. find two closely related ERVs that can be specifically activated by the naive and primed pluripotency transcription machineries, respectively, and that contribute to the cell fate plasticity.

Ai et al., 2022, Cell Reports *40*, 111240
August 23, 2022
https://doi.org/10.1016/j.celrep.2022.111240

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Krüppel-like factor 5 rewires NANOG regulatory network to activate human naive pluripotency specific LTR7Ys and promote naive pluripotency

Zhipeng Ai,[1,10] Xinyu Xiang,[2,10] Yangquan Xiang,[1,10] Iwona Szczerbinska,[3,10] Yuli Qian,[4,10] Xiao Xu,[1] Chenyang Ma,[1] Yaqi Su,[2] Bing Gao,[2] Hao Shen,[1] Muhammad Nadzim Bin Ramli,[3] Di Chen,[2] Yue Liu,[5] Jia-jie Hao,[5] Huck Hui Ng,[3,6,7] Dan Zhang,[4,*] Yun-Shen Chan,[3,5,*] Wanlu Liu,[2,8,9,*] and Hongqing Liang[1,11,*]

[1]Division of Human Reproduction and Developmental Genetics, Women's Hospital, and Institute of Genetics, Zhejiang University School of Medicine, Hangzhou 310006, China

[2]Zhejiang University-University of Edinburgh Institute (ZJU-UoE Institute), Zhejiang University School of Medicine, International Campus, Zhejiang University, 718 East Haizhou Road, Haining 314400, China

[3]Stem Cell and Regenerative Biology, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore

[4]Key Laboratory of Reproductive Genetics (Ministry of Education) and Department of Reproductive Endocrinology, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou 310006, China

[5]Guangzhou Laboratory, No. 9 Xing Dao Huan Bei Road, Guangzhou International Bio Island, Guangzhou 510005, China

[6]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117597, Singapore

[7]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 639798, Singapore

[8]Department of Orthopedic Surgery of the Second Affiliated Hospital of Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310029, China

[9]Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cell and Regenerative Medicine, Zhejiang University, Hangzhou 310058, China

[10]These authors contributed equally

[11]Lead contact

*Correspondence: zhangdan@zju.edu.cn (D.Z.), chan_yunshen@gzlab.ac.cn (Y.-S.C.), wanluliu@intl.zju.edu.cn (W.L.), lianghongqing@zju.edu.cn (H.L.)

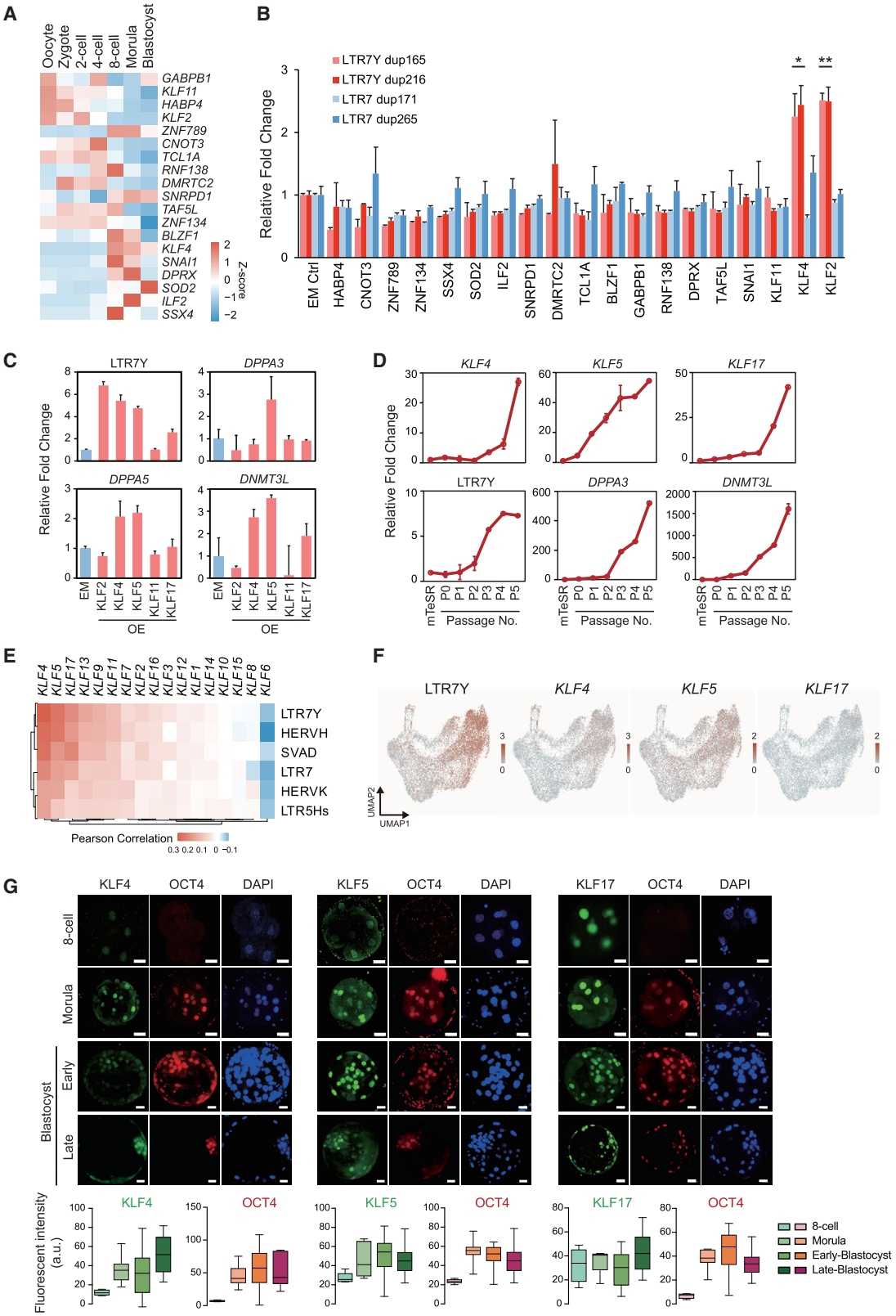
https://doi.org/10.1016/j.celrep.2022.111240

SUMMARY

Endogenous retroviruses (ERVs) have been reported to participate in pre-implantation development of mammalian embryos. In early human embryogenesis, different ERV sub-families are activated in a highly stage-specific manner. How the specificity of ERV activation is achieved remains largely unknown. Here, we demonstrate the mechanism of how LTR7Ys, the human morula-blastocyst-specific HERVH long terminal repeats, are activated by the naive pluripotency transcription network. We find that KLF5 interacts with and rewires NANOG to bind and regulate LTR7Ys; in contrast, the primed-specific LTR7s are preferentially bound by NANOG in the absence of KLF5. The specific activation of LTR7Ys by KLF5 and NANOG in pluripotent stem cells contributes to human-specific naive pluripotency regulation. KLF5-LTR7Y axis also promotes the expression of trophectoderm genes and contributes to the expanded cell potential toward extra-embryonic lineage. Our study suggests that HERVs are activated by cell-state-specific transcription machinery and promote stage-specific transcription network and cell potency.

INTRODUCTION

Transposable elements (TEs) or transposons are repetitive sequences that once had or still retain the capacity to move from one locus to another in the genome. They account for almost half of the human genome size (Bourque et al., 2018; Fedoroff, 2012; Kazazian, 2004; Klein and O'Neill, 2018; Kunarso et al., 2010). Endogenous retrovirus (ERV) is a class of retro-transposons with flanking long terminal repeats (LTRs) derived from retroviruses (de Souza et al., 2013; Feschotte and Gilbert, 2012; Goke and Ng, 2016; Kazazian, 2004). Based on sequence features and conservation, they are classified into different families and sub-families (Bourque et al., 2008; Chuong et al., 2016; Goke and Ng, 2016). Compared with genes, ERVs are much less conserved across species and are predicted to play important roles in species-specific evolution (Kunarso et al., 2010; Pontis et al., 2019; Senft and Macfarlan, 2021). In mammals, ERVs are highly active during early embryogenesis and germline development (Goke et al., 2015; Grow et al., 2015; Wang et al., 2014). They are involved in the pre-implantation transcription network, although their exact roles in early embryonic development are still under-studied (Fu et al., 2019; Grow et al., 2015; Modzelewski et al., 2021; Yang et al., 2020). In somatic cells, ERVs are transcriptionally silenced, while their dysregulation has been linked to pathological conditions like neurodegeneration and cancers (Geis and Goff, 2020; Jönsson et al., 2020).



(legend on next page)

Given the potential importance of ERVs in development and diseases, understanding how these genomic elements are regulated is critical for unraveling their functions. The studies on ERV transcriptional regulation have mainly focused on their silencing mechanism (Carter et al., 2020; Imbeault et al., 2017; Rowe et al., 2010; Sun et al., 2019). ERV activation used to be recognized as the outcome of chromatin relaxation (Kunarso et al., 2010; Wu et al., 2020); however, emerging studies revealed that specific transcription factors (TFs) are involved in ERV activation. The mouse two-cell (2C) embryo-specific ERV sub-family, MERVL, can be activated by 2C specific TFs, like Dux and Zscan4 (Guo et al., 2019; Hendrickson et al., 2017; Macfarlan et al., 2012; Whiddon et al., 2017; Yang et al., 2020). In early human embryonic development, different sub-families of ERVs exhibit stage-specific transcription from oocyte to blastocyst (Goke et al., 2015). How different sub-families of human ERVs (HERVs) are specifically activated at distinct embryonic stages is not clearly understood. HERVH is a hominoid-specific ERV sub-family, which was reported to be involved in both naive and primed pluripotency networks (Wang et al., 2014; Lu et al., 2014; Grow et al., 2015; Pontis et al., 2019), but the functional and regulatory distinctions of HERVH under these two pluripotency states is not clear. Interestingly, for different LTRs of HERVH, the LTR7 sub-family is upregulated in the primed state (Lu et al., 2014; Wang et al., 2014), while large copies of LTR7Y sub-family exhibit naive specificity and are highly enriched in morula-blastocyst stage of human embryo (Goke et al., 2015). LTR7Ys represent the most recently expanded LTRs in HERVH, the burst of which took place around 10–14 million years ago (Carter et al., 2022). The association of different LTRs with naive and primed pluripotency states implies cell-state-specific regulation for the closely related LTRs of HERVHs. Uncovering the underlying regulatory mechanism will provide insights into how ERVs contribute to the stage-specific transcription network and function to regulate cell fate decisions.

In this study, we demonstrated that the human morula-blastocyst and naive-specific LTR7Y sub-family was activated by the naive-specific TFs, KLF4, and KLF5. KLF4 and KLF5 had overlapping and distinct gene and TE targets. We further showed that KLF5 interacted with and rewired NANOG toward a naive-specific transcriptional state. This resulted in the re-direction of NANOG binding toward the naive-specific LTR7Y sites from the primed-enriched LTR7 sites. Correspondingly, the loss of KLF5 in naive state decommissioned NANOG from LTR7Y sites. The coordinated activation of LTR7Ys by KLF5 and NANOG facilitated the activation of adjacent naive and trophectoderm-related genes, contributing to the maintenance of human naive pluripotent stem cell (PSC) state and the expanded developmental potential.

RESULTS

Identifying transcriptional activators for human naive pluripotency and morula-blastocyst-specific LTR7Y element

The LTR7Y sub-family of HERVH is enriched in the morula-blastocyst stage of human embryos (Figure S1A) and is among the top enriched TEs in naive human embryonic stem cells (hESCs) (Figure S1B). In contrast, the LTR7 sub-family of HERVH is highly enriched in the primed hESCs (Figure S1B) but not enriched in the morula-blastocyst stage (Figure S1A) (Goke et al., 2015). This suggests that specific regulatory mechanisms are in place to enable the timely expression of the closely related LTRs of HERVH. To look for family-specific ERV activators, we cloned and tested a subset of nuclear factors enriched at different stages of early embryos, for their potency in activating selected copies of LTR7s and LTR7Ys (Figure 1A). Among these factors, over-expression of KLF2 and KLF4 significantly increased the expression of naive-specific LTR7Ys (dup165, dup216) but not the primed-specific LTR7s (dup171, dup265) (Figure 1B). KLF family members were reported to exhibit functional redundancies in mice (Jiang et al., 2008); thus to identify KLFs regulating LTR7Ys in human embryos, we focus on those KLFs (KLF2, KLF4, KLF5, KLF11, and KLF17) that are highly expressed in early human embryos (Figure S1C) and are also enriched in naive PSCs (Figures S1C and S1D) (Xue et al., 2013; Yan et al., 2013). Among them, ectopic expression of KLF4 and KLF5 in primed hESCs led to significant induction of endogenous LTR7Y as well as naive-specific genes (Figure 1C).

We proceeded further to investigate the expression dynamics of KLF4, KLF5, and LTR7Y in naive hESCs and early human embryos. We found that during the reversion from primed to naive state, KLF5 was significantly upregulated, preceding the increment of LTR7Y, other KLFs, and naive genes (Figures 1D and

Figure 1. Identification of human naive pluripotency and morula-blastocyst-specific LTR7Y activator

(A) Expression heatmap of 19 selected nuclear factors from early human embryos.

(B) 19 factors were over-expressed in primed hESCs, and the relative expression of LTR7Y (dup165 and dup216) and LTR7 (dup171 and dup265) was measured by qRT-PCR and normalized to empty vector control (EM Ctrl) (mean ± SEM, n = 3, technical replicates; *p < 0.05, **p < 0.01; unpaired t test).

(C) The relative expression changes of LTR7Y and naive-specific genes, upon stable over-expression of KLF2, KLF4, KLF5, KLF11, or KLF17 in primed hESCs (mTeSR) (mean ± SEM, n = 3, biological replicates).

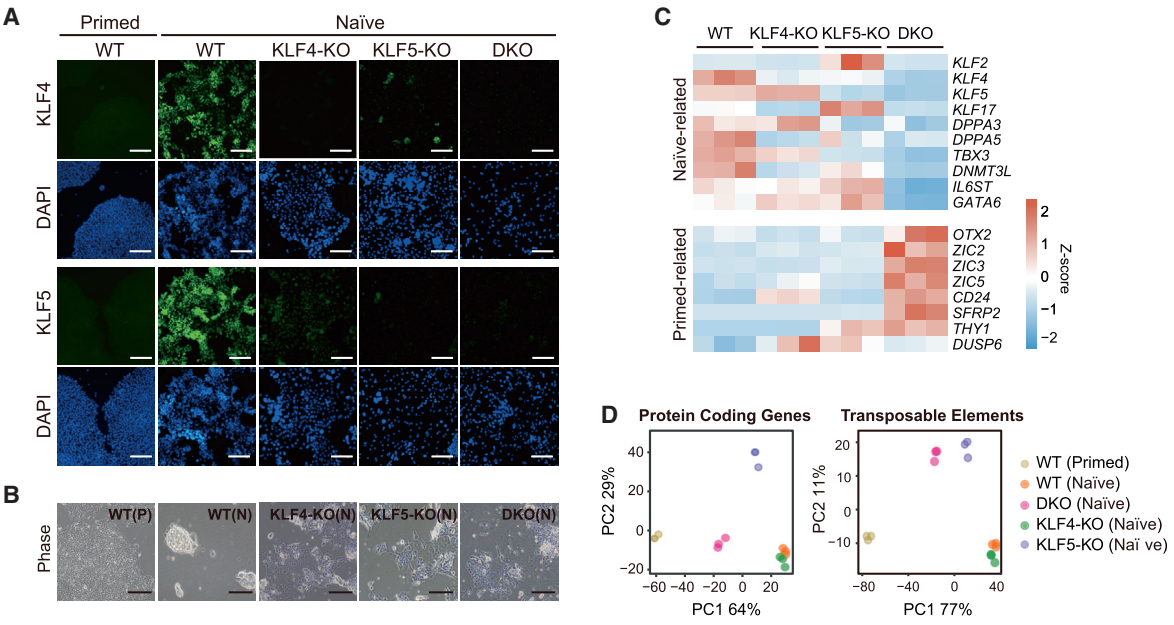
(D) The kinetics of relative expression changes for *KLF4*, *KLF5*, *KLF17*, LTR7Y, *DPPA3*, and *DNMT3L* from passage 0 to passage 5 during primed (mTeSR) to naive (FINE) state reversion process (mean ± SD, n = 2, biological replicates).

(E) The heatmap for expression correlation of KLFs and pluripotency-related TEs in single cell transcriptome of naive hESCs (cultured in FINE condition, n = 13,154).

(F) Uniform manifold approximation and project (UMAP) for single-cell transcriptome of naive hESCs showing the expression for LTR7Y, *KLF4*, *KLF5*, and *KLF17* in naive hESCs cultured in FINE condition. The color code on the right of the UMAP indicates the relative expression level of different genes.

(G) Immunofluorescence staining of KLF4, KLF5, KLF17 (green), and OCT4 (red) in early human embryos. Eight-cell, morula, early blastocyst (B1, B2 blastocysts), and late blastocyst (B3, B4 blastocysts) stages were stained respectively. Scale bar: 50 μm. Boxplots showing the quantification of KLF4, KLF5, KLF17, and OCT4 single-nucleus staining intensity (mean ± SD) at different embryonic stages; images for quantification are shown in this figure and Figure S1G.

See also Figure S1.



(legend on next page)

S1E), suggesting a pioneering role of KLF5 in LTR7Y activation and naive state reversion. In contrast, LTR7 expression was gradually reduced during the reversion process (Figure S1E). We also performed single-cell RNA-seq using naive hESCs cultured in the feeder-free condition (Szczerbinska et al., 2019) to identify the expression correlations between different KLFs and ERVs at single-cell level (Figures 1E, 1F, and S1F) (https://labw.org/hESCTE). Interestingly, the expression of KLF4/KLF5 and LTR7Ys were among the top-correlated KLF-TE pairs at single cell level in naive hESCs (Figures 1E and 1F). Comparatively, the correlation was weaker between KLF4/KLF5 and LTR7s (Figures 1E, 1F, and S1F). We further investigated the temporal dynamics of KLF4, KLF5, and KLF17 protein expression in early human embryos (Figures 1G and S1G). KLF4 and KLF5 protein levels were low at eight-cell stage and became more pronounced from morula to blastocyst stages (Figures 1G and S1G, left and middle panels). This dynamic correlated well with that of the LTR7Y expression in early embryos (Figure S1A). In contrast, KLF17 protein expression was not specific to morula and blastocyst stages but appeared pronounced throughout eight-cell to blastocyst stage (Figures 1G and S1G, right panel). Collectively, the results from gain of function screen and expression profiling in naive hESCs and early human embryos suggest that KLF4 and KLF5 are potential activators of LTR7Ys in human embryos.

KLF4 and KLF5 activate LTR7Ys and regulate naive pluripotency

To further understand how KLF4 and KLF5 functionally regulate LTR7Y and naive pluripotency, *KLF4* single-knockout (KLF4-KO), *KLF5* single-knockout (KLF5-KO), and *KLF4*/*KLF5* double-knockout (DKO) hESC lines were generated by CRISPR-CAS9 targeting (Figures 2A, S2A, and S2B). The depletion of KLF4 or/and KLF5 resulted in enlarged and flattened colony morphology (Figure 2B) with the corresponding loss of naive marker expression, and the effect was most prominent in the DKO line (Figures 2A and S2C). Surprisingly, the expression of pluripotency-associated genes and alkaline phosphatase staining were retained in the knockout lines (Figures S2D and S2E). Transcriptome profiling for the single- and double-knockout lines further supported these observations (Figure 2C; Table S1). Compared with single-knockout cells, DKO cells exhibited more pronounced alteration in global gene expression (Figures 2C and S2F; Table S1) and greater shift toward primed hESCs in principal component 1 (PC1) of the principal component analysis (PCA) (Figure 2D). This suggests redundancy of KLF4 and KLF5 function in the naive transcription network, although distinctions in gene expression were also observed between KLF4-KO and KLF5-KO lines (Figures 2C and S2F; Table S1).

Further analysis of differentially expressed TEs (DE-TEs) also revealed distinctive regulatory pattern between KLF4 and KLF5 (Figures 2D and 2E; Table S2). KLF4 depletion downregulated naive-enriched SVA, HERVK, and LTR5Hs sub-families, consistent with the previous finding on KLF4 in regulating these TEs (Pontis et al., 2019). In contrast, KLF5 depletion led to upregulation of the SVAD sub-families (Figure 2E). Despite the differences, LTR7Y sub-family was the top downregulated TE upon KLF4 or KLF5 single and double depletion (Figure 2E). In contrast, the LTR7 sub-family was instead upregulated in KLF5-KO and DKO lines, potentially due to the dissolution of naive pluripotency state (Figure 2E). The fact that KLF4 and KLF5 positively regulate LTR7Ys, but not LTR7s, further supports that LTR7Ys and LTR7s have distinct regulation patterns.

To probe the functional implications of LTR7Y activation in naive pluripotency, LTR7Y loci were directly activated by CRISPR activation (CRISPRa) system with sgRNAs targeting the LTR7Y consensus sequences (Figure 2F). The increased expression of LTR7Ys, but not other related HERVH sub-families, supported the relatively specific targeting efficiency toward LTR7Ys (Figure S2G; Table S3). Activation on LTR7Y loci by CRISPRa led to increased expression for some of the naive marker genes and decreased expression for primed marker genes (Figure 2F). To determine whether the effect of LTR7Ys in promoting naive state was through the transcribed RNAs from LTR7Y loci or the cis-regulatory function of LTR7Ys, we first knocked down the LTR7Y transcripts with siRNA targeting the consensus sequence of LTR7Ys (Figure S2H). However, reduction of LTR7Y RNA did not lead to prominent alteration of naive and primed marker genes expression (Figure S2I), suggesting LTR7Ys may function through *cis*-regulatory effect in the naive transcription network.

Figure 2. The transcription targets of KLF4 and KLF5 in naive hESCs

(A) KLF4 and KLF5 were individually or jointly knocked out in primed hESCs and converted to naive state. Immunofluorescence staining of KLF4 and KLF5 in wild-type (WT) hESC cultured in primed (mTeSR) and naive (FINE) conditions, as well as KLF4-KO, KLF5-KO, and DKO cultured in naive conditions. Scale bar: 50 μm.

(B) Morphology (phase) of WT hESCs in primed (P) and naive (N) culture conditions, as well as KLF4-KO, KLF5-KO, and DKO in naive (N) culture conditions. Scale bar: 50 μm.

(C–E) RNA-seq data for WT, KLF4-KO, KLF5-KO, and DKO hESCs (biological triplicates) in naive culture conditions. (C) Heatmap for naive pluripotency-related genes (top) or primed pluripotency-related genes (bottom) in WT, KLF4-KO, KLF5-KO, and DKO lines. (D) Principal component analysis (PCA) based on the expression of protein coding genes or transposable elements (TEs) in WT, KLF4-KO, KLF5-KO, and DKO hESCs in primed or naive culture conditions. (E) Top 10 differentially expressed TE (DE-TE) sub-families in KLF4-KO, KLF5-KO, and DKO compared with WT hESCs.

(F) LTR7Ys were activated by CRISPRa with sgRNAs targeting LTR7Y conserved sequence: the schematic was illustrated on the left; RNA-seq was performed upon the expression of CRISPRa-VP160 targeting LTR7Ys sequences in primed hESCs, and the heatmaps for naive pluripotency-related genes and primed pluripotency-related genes are shown on the right. RNA-seq was performed in biological duplicates.

(G) Luciferase assay to detect the cooperative effect of NANOG and KLFs in activating naive-specific LTR7Y (dup216): KLF4, KLF5, or KLF17 were expressed in HEK293T cells with or without the co-expression of NANOG, and luciferase activity driven by LTR7Y was measured (mean ± SEM, n = 3, technical replicates; NS, not significant, *$p < 0.05$, ***$p < 0.001$, unpaired t test).

(H) Western blots showing the co-immunoprecipitation of KLF4 or KLF5 and NANOG in HEK293T cells over-expressing NANOG and HA tagged KLF4 or KLF5.

See also Figure S2 and Tables S1–S3.

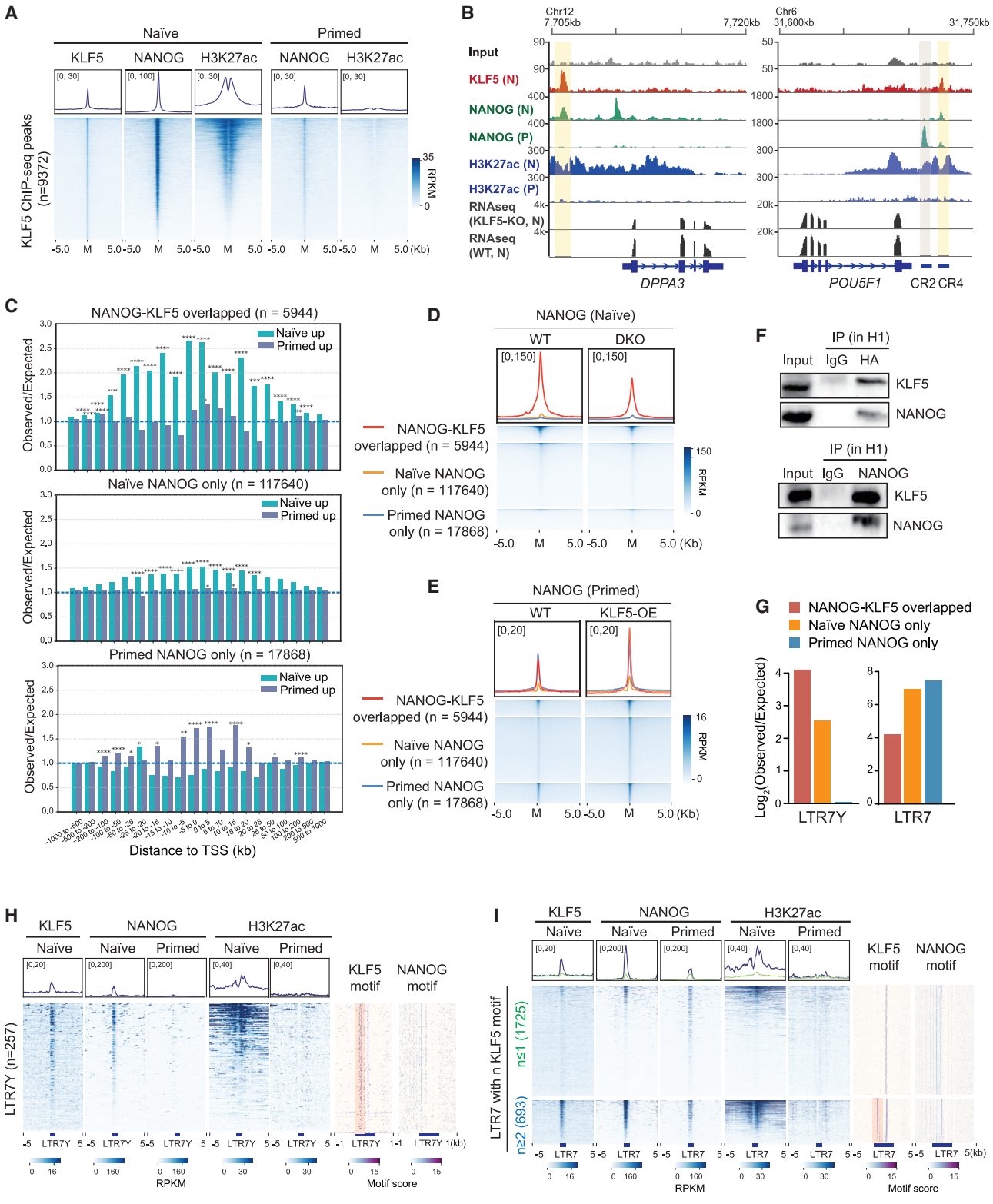


Figure 3. KLF5 coordinates with NANOG to regulate the naive transcription

(A) Heatmap of KLF5, NANOG, and H3K27ac ChIP-seq signal over KLF5 binding peaks (n = 9,372) in naive and primed hESCs. M denotes midpoint of peak.

(B) Screenshot of KLF5 (red), NANOG (green), H3K27ac (blue) ChIP-seq signals in naive (N) and primed (P) hESCs in genomic regions near *DPPA3* and *POU5F1*. Tracks of RNA-seq signal for WT and KLF5-KO in naive hESCs are shown in black.

(legend continued on next page)

To further demonstrate the *cis*-regulatory function of LTR7Ys, an LTR7Y genomic locus (dup216) highly active in the naive pluripotency state was cloned into the Renilla luciferase reporter construct. With the co-expression of KLF4 or KLF5, downstream Renilla expression was activated (Figure 2G), suggesting that KLF4 or KLF5 can directly activate gene expression through the LTR7Y DNA sequence. NANOG was previously shown to synergize with KLF in establishing the naive transcription network (Takashima et al., 2014). We therefore investigated the co-regulatory effect of NANOG and KLFs on LTR7Y. Despite being a direct activator of LTR7 as previously reported (Goke et al., 2015), NANOG exhibited minimal activation toward LTR7Y reporter (Figure 2G). Instead, NANOG can augment the activation of LTR7Y luciferase reporter by KLF5. This synergistic induction activity with NANOG was less significant for KLF4 and not observed for KLF17 (Figure 2G). To identify the molecular basis for this differential coordination, KLF4 and KLF5 were individually over-expressed with NANOG in HEK293T cells (Figure S2J), and the reciprocal pull-down efficiency between KLFs and NANOG was detected. NANOG co-precipitated more with KLF5 compared with KLF4; and conversely, NANOG also pulled down more KLF5 than KLF4 (Figure 2H), suggesting the more robust binding between KLF5 and NANOG that underlay their stronger synergistic activity on LTR7Y. These results thus far demonstrate that NANOG can synergize with KLF5 to activate the *cis*-regulatory function of LTR7Y, which potentially contributes to the naive pluripotency.

KLF5 rewires NANOG in the naive transcription network to regulate naive-specific genes and TEs

To further investigate how KLF5 and NANOG coordinately function through LTR7Y loci in the naive transcription network, we generated KLF5 and NANOG genome-wide binding profile and H3K27ac profile in naive hESCs and compared with NANOG genome-wide binding profile and H3K27ac profile in primed hESCs (Figure 3A). Region Associated DEGs (RAD) analysis (Guo et al., 2021b) revealed that genomic locations near KLF5 binding peaks were enriched with naive upregulated genes (Figure S3A), similar to the RAD pattern surrounding naive-specific H3K27ac peaks (Figure S3B). This supported the transcription activation role of KLF5 in naive pluripotency network. KLF5 peaks were enriched with KLF motif and motif of OCT4-SOX2-TCF-NANOG complex (Figure S3C). The majority of KLF5 peaks overlapped (5,944 out of 9,372 peaks) with NANOG peaks in the naive state (Figures 3A and S3D), while these sites lacked NANOG and H3K27ac ChIP-seq signals in primed hESCs (Figures 3A, 3B, and S3E). On the contrary, primed-specific NANOG peaks were minimally bound by KLF5 in naive states (Figures 3B, S3D, and S3F).

Comparing the RAD plots for different types of NANOG peaks (Figure S3D), we found that the naive upregulated genes were more enriched near NANOG-KLF5 overlapped peaks (n = 5,944) than the naive NANOG-only peaks (n = 117,640) (Figure 3C). While the primed NANOG-only peaks were more enriched with primed-specific genes in their neighboring regions (Figure 3C). These results revealed a global rewiring of NANOG targeting affinity from primed to naive state and hinted that KLF5 potentially played a role in the rewiring process. To validate this hypothesis, we performed NANOG CUT&Tag to detect whether the lack of KLFs altered NANOG binding profiles in naive hESCs. Indeed, with the depletion of KLF4/5 in DKO cells, NANOG signal in NANOG-KLF5 overlapped peaks were prominently reduced, and the reduction was stronger compared with other types of NANOG peaks (Figure 3D). Furthermore, KLF5 depletion downregulated genes in the neighboring region of KLF5 peaks (Figure S3G). Reciprocally, when KLF5 was over-expressed in the primed hESCs, the binding of NANOG onto NANOG-KLF5 overlapped peaks increased (Figure 3E), and this was also accompanied with the upregulation of naive-marker genes and downregulation of primed marker genes (Figure S3H and Table S4). We further performed reciprocal co-immunoprecipitation and confirmed the interaction between KLF5 and NANOG in hESCs (Figure 3F). These data support that KLF5 can directly bind to NANOG, rewire its targeting to naive-specific sites, and promote the naive pluripotency transcription.

Interestingly, around two-thirds of KLF5 peaks overlapped with TEs (6,109 out of 9,372 peaks), and naive upregulated genes were also enriched near these KLF5 bound TE sequences (Figure S3I). Among different types of genomic elements and TEs,

(C) RAD bar plots representing the association between different types of NANOG/KLF5 ChIP-seq peaks and the DEGs between naive and primed hESCs. The types of peaks include NANOG-KLF5 overlapped peaks (n = 5,944), naive NANOG-only peaks (n = 117,640), and primed NANOG-only peaks (n = 17,868). Hypergeometric test was used: *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

(D) Metaplot and heatmap for NANOG CUT&Tag signal in naive hESCs (WT) and naive hESCs with *KLF4* and *KLF5* double KO (DKO).

(E) Metaplot and heatmap for NANOG ChIP-seq signals in primed hESCs (primed WT) and primed hESCs with KLF5 over-expression (primed KLF5-OE). For (D) and (E), signals over +/− 5 kb regions around three different types of NANOG peaks in the genome are shown, including NANOG-KLF5 overlapped peaks (n = 5,944, red), naive NANOG-only peaks (n = 117,640, orange), and primed NANOG-only peaks (n = 17,868, blue).

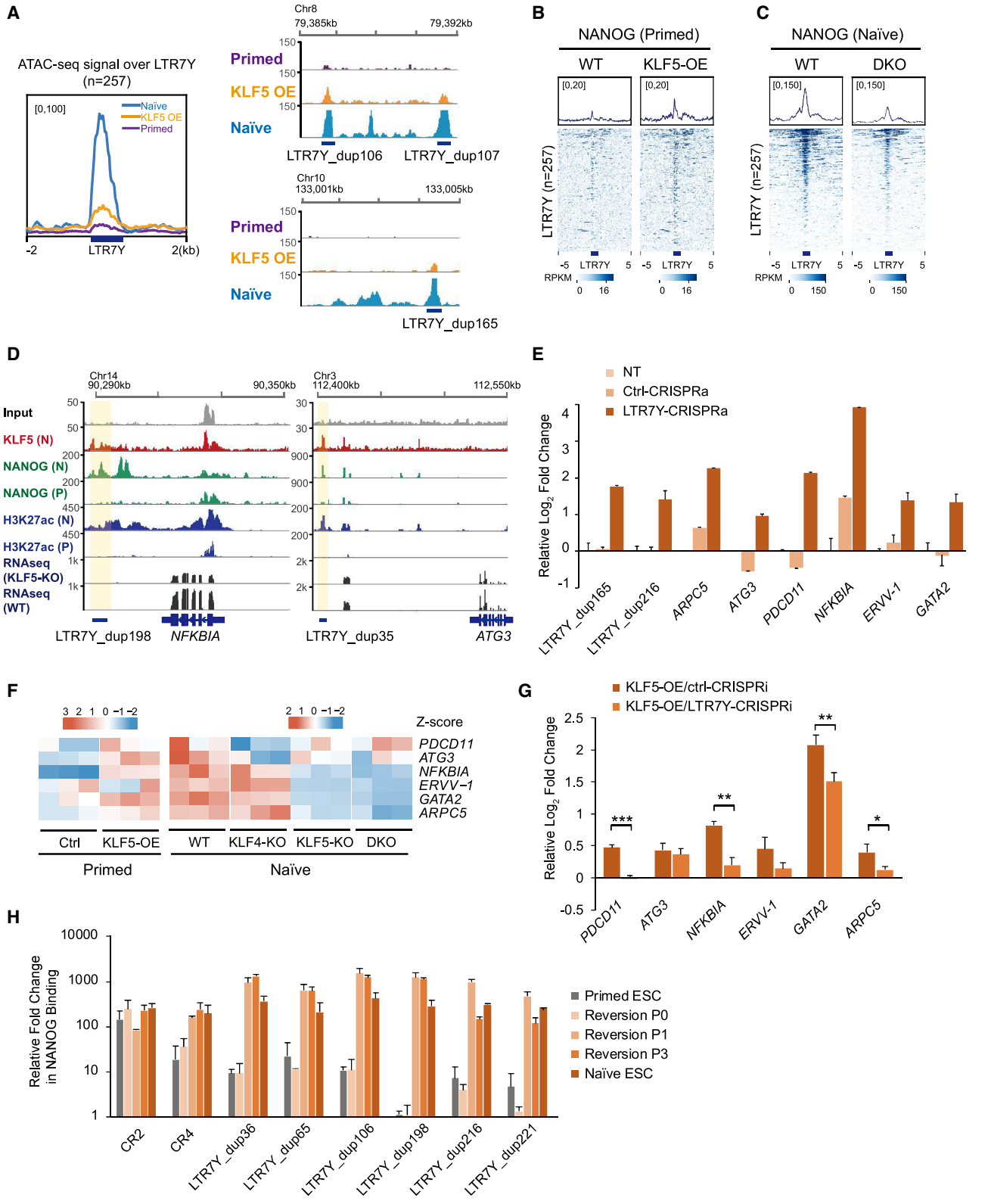
(F) Western blots showing the reciprocal co-immunoprecipitation of HA-tagged KLF5 and NANOG in hESCs (H1). HA or NANOG antibodies were used for co-immunoprecipitation, and NANOG or KLF5 antibodies were used for immunoblotting.

(G) The Log_2 enrichment score of observed over expected frequencies for LTR7Y and LTR7 in NANOG-KLF5 overlapped peaks (n = 5,944), naive NANOG-only peaks (n = 117,640), and primed NANOG-only peaks (n = 17,868). The expected frequency represents the genomic shuffle average.

(H) Heatmap of KLF5, NANOG, and H3K27ac ChIP-seq signal in naive and primed hESCs over +/− 5 kb regions across all the LTR7Y loci (n = 257) in the genome. The distribution of KLF5 and NANOG motifs over +/− 1 kb regions of all LTR7Y copies is shown with the pink shade highlighting the KLF5 motif near the 5' end of LTR7Y sequences.

(I) Heatmap of KLF5, NANOG, and H3K27ac ChIP-seq signal in naive and primed hESCs over +/− 5 kb regions across all the LTR7 loci in the genome. The distribution of KLF5 and NANOG motifs over +/− 1 kb regions of all LTR7 copies are shown. Based on the number of KLF5 motifs present in the LTR7 sequences, LTR7s were divided into two groups: KLF5 motif number $\geq$ 2 (LTR7 n = 693) or number $\leq$ 1 (LTR7 n = 1,725). The pink shade highlights the KLF5 motif at the 5' end of LTR7 sequences.

See also Figure S3.



(legend on next page)

KLF5 was specifically enriched over LTR elements (Figure S3J). This is also consistent with the observation that the top downregulated DE-TEs upon KLF5 depletion were mainly ERV/LTRs (Figure 2E). Interestingly, the primed-specific LTR7s were more enriched with primed NANOG-only peaks, whereas the naive-specific LTR7Ys were more enriched in NANOG-KLF5 overlapped peaks, while rarely bound by primed NANOG-only peaks (Figure 3G). Metaplots and heatmaps further confirmed that majority of LTR7Y loci were co-bound by KLF5 and NANOG and marked by H3K27ac in naive state, but depleted with NANOG binding and H3K27ac in primed state (Figures 3H and S3K). This is consistent with the LTR7Y luciferase assay, in which NANOG-only cannot bind and activate LTR7Y in the absence of KLF5 (Figure 2G). In sharp contrast, LTR7 sub-family had prominent NANOG binding in primed state (Figure 3I and S3L). This could be contributed by the higher abundance of NANOG motifs in the sequences of LTR7s compared with LTR7Ys (Figures 3H, 3I, and S3M). In contrast to LTR7Ys, a majority of LTR7 loci showed reduced NANOG binding in naive state (Figures 3I and S3L), except for a small group of LTR7s ($n \geq 2{,}693$), which exhibited high KLF5 and NANOG binding in naive state (Figure 3I). Interestingly, motif analysis revealed that this LTR7 sub-group possessed an additional KLF5 motif near the 5' end of their sequences (Figure 3I), while the majority of LTR7s only had one or fewer KLF5 motif (Figure S3M). This additional 5' KLF5 motif was also largely present in the sequences of LTR7Y sub-family (Figures 3H and S3M). To check whether the differential affinities toward KLF5 and NANOG relied on KLF motif differences, we constructed luciferase reporter using LTR7 genomic loci with or without this additional 5' KLF5 motif. Interestingly, LTR7-dup647 with the additional 5' KLF5 motif was activated by KLF5 but not by NANOG (Figure S3N), similar to the activation pattern of LTR7Y. On the other hand, the LTR7-dup804, which represents a majority of LTR7s lacking this 5' motif, can only be activated by NANOG but not by KLF5 (Figure S3N). This evidence suggest that the temporal expression dynamics and regulatory differences in LTR7Y and LTR7 sub-families can be routed back to their sequence-mediated differential interaction with cell-state-specific TFs.

KLF5 rewires NANOG on LTR7Ys to promote naive gene transcription

To understand the mechanism of how KLF5 rewires NANOG to activate LTR7Ys, we performed ATAC-seq to look at whether KLF5 influences the chromatin accessibility of LTR7Y loci. LTR7Y loci were largely in the closed state in primed hESCs but had increased accessibility in naive state (Figure 4A). The ectopic expression of KLF5 in primed hESCs led to increased accessibility of LTR7Y loci (Figure 4A). Furthermore, transient KLF5 over-expression also increased NANOG recruitment onto LTR7Y sites (Figure 4B), implying KLF5 may open up LTR7Y loci for subsequent recruitment of NANOG. Consistent with this, depletion of KLF4/KLF5 in naive hESCs rendered dramatic loss of NANOG binding on LTR7Ys (Figure 4C). KLF5 and NANOG also interacted with P300 (Figure S4A), and following KLF5 over-expression, H3K27ac deposition was observed on some of the LTR7Y loci (Figure S4B). Naive-specific genes near LTR7Y loci (Figure 4D) were upregulated upon the expression of CRISPRa targeting LTR7Y conserved sequences (Figures 4E and S4C). This outcome was independent of LTR7Y RNA expression, as simultaneously knocking down LTR7Y by siRNA did not abolish the upregulation of these LTR7Y nearby genes (Figure S4D). Correspondingly, these LTR7Y nearby genes were also upregulated upon KLF5 over-expression in primed hESCs (Figure 4F; Table S4). Further over-expression of NANOG enhanced the activation effect of KLF5 on some of these genes (Figure S4E). CRISPR inactivation (CRISPRi) of LTR7Ys in the KLF5 over-expressing hESCs led to downregulation of these genes to different extents (Figure 4G). Although, it is noted that only KLF5 but not KLF4 depletion can efficiently reduce the expression of LTR7Y nearby naive genes (Figure 4F).

In addition, during the primed to naive reversion process, increased NANOG binding on LTR7Y sites was observed as early as passage 1 (P1) (Figure 4H), when endogenous *KLF5* expression was induced (Figure 1D). These results collectively indicate that KLF5 recruited NANOG and other co-activators to activate LTR7Y loci, promoting the transcription activation of naive-specific genes near LTR7Ys.

Figure 4. The coordinated activation of LTR7Y by KLF5 and NANOG promotes naive pluripotency gene expression

(A) Metaplot on the left showing the ATAC-seq signal over +/− 2 kb regions around LTR7Y copies (n = 257) in primed hESCs, primed hESCs with KLF5-OE, and naive hESCs. Screenshot on the right showing the ATAC-seq signals in primed, KLF5-OE, and naive hESCs over selected LTR7Y loci (dup106, dup107, and dup165).

(B) Metaplot and heatmap for NANOG ChIP-seq signals over +/− 5 kb regions of all LTR7Y copies (n = 257) in WT and KLF5-OE primed hESCs.

(C) Metaplot and heatmap showing NANOG CUT&Tag signals over +/− 5 kb regions of all LTR7Y copies in WT and KLF4/KLF5 double-knockout (DKO) naive hESCs.

(D) Screenshot of KLF5 (red), NANOG (green), and H2K37ac (blue) ChIP-seq peaks from naive and primed hESCs around LTR7Y loci (dup198 and dup35) and their nearby naive genes (*NFKBIA* and *ATG3*). Tracks of RNA-seq reads for WT and KLF5-KO are shown in black.

(E) qRT-PCR showing the Log_2 relative expression fold change of naive-specific LTR7Y nearby genes in untreated primed hESCs (NT), primed ESCs over-expressing CRISPRa targeting scrambled sequence (Ctrl-CRISPRa), or targeting LTR7Y conserved sequences (LTR7Y-CRISPRa, with sg-RNA1/2/3) (mean ± SEM, n = 3, technical replicates).

(F) Expression heatmap for naive-specific LTR7Y nearby genes in primed hESCs control (Ctrl) and primed hESCs with KLF5-OE, or in naive hESCs (WT) and naive hESCs with KLFs knockout (KLF4-KO, KLF5-KO and DKO). RNA-seq was performed for biological triplicates.

(G) The Log_2 relative fold change of LTR7Y nearby genes in KLF5-OE primed hESCs or KLF5-OE primed hESCs with simultaneous CRISPRi-KRAB targeting LTR7Ys (mean ± SEM, n = 3, technical replicates; *p < 0.05, **p < 0.01, ***p < 0.001; unpaired t test).

(H) ChIP-qPCR showing the relative fold change of NANOG binding during primed to naive pluripotency reversion process. NANOG ChIP was performed in P0, P1, and P3 of hESCs during the reversion process as well as in fully converted naive hESCs. The relative NANOG binding on *OCT4* distal (CR4) and proximal (CR2) enhancers, selected naive-specific LTR7Y loci were detected (mean ± SEM, n = 3, technical replicates).

See also Figure S4; Table S1 and S4.

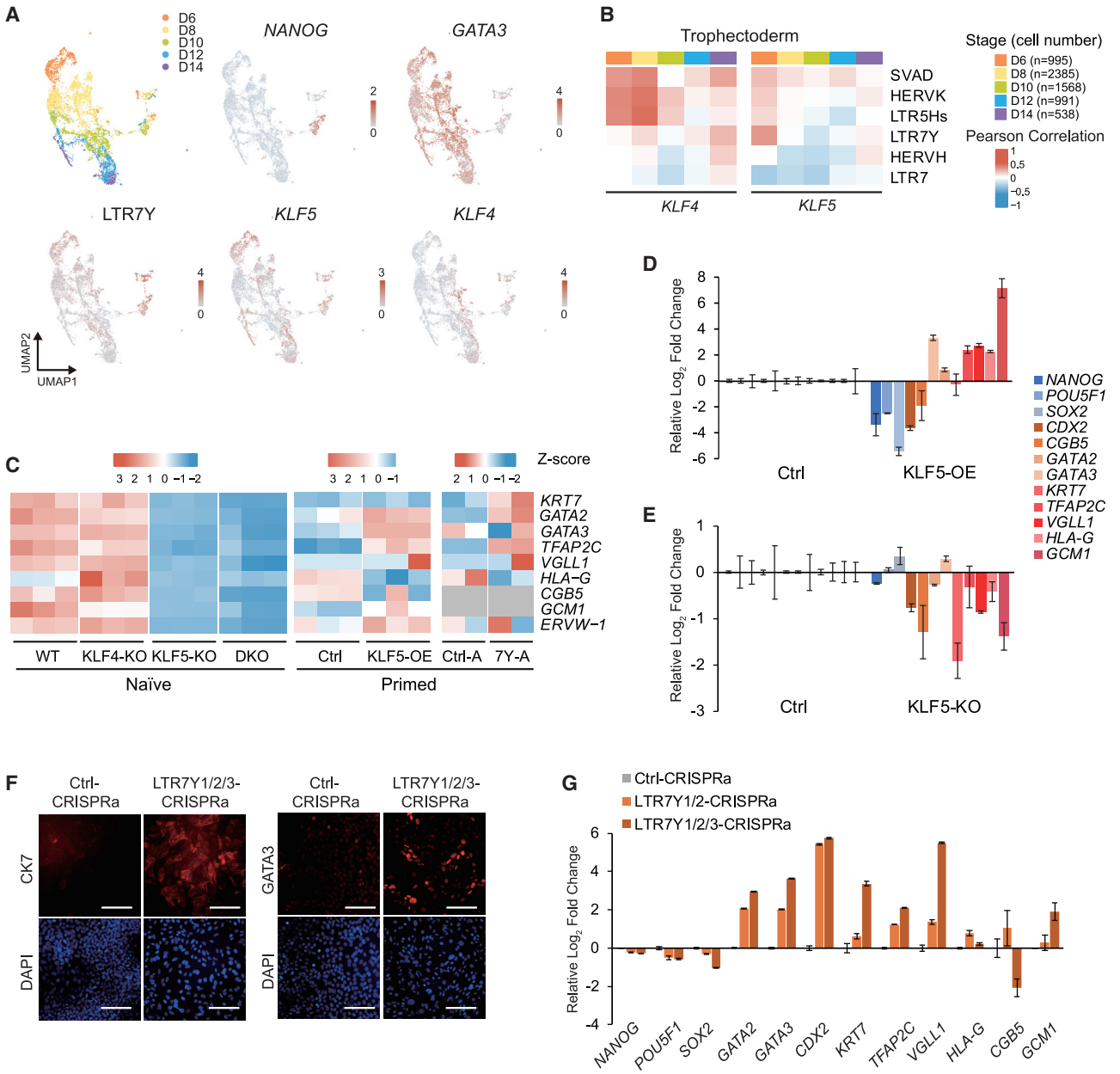


Figure 5. KLF5-LTR7Y capacitates cells with trophectoderm potential

(A) UMAP for single cell transcriptome of post-implantation human embryos (n = 7,924) (GSE109555): top left panel showing the cells colored by developmental stage (day 6, 8, 10, 12, 14). The expressions of *NANOG, GATA3,* LTR7Y*, KLF4*, and *KLF5* are shown. Color code on the right of UMAPs indicated the relative expression level of genes and TEs.

(B) Heatmap showing the expression correlation of *KLF4*, *KLF5*, and pluripotency-associated TEs in trophectoderm lineage at different developmental stages. The developmental stages (day 6, 8, 10, 12, 14) are indicated by different colors, and the cell numbers from different embryonic days used for calculation of correlation are shown in brackets; color code shows the Pearson correlation coefficient.

(C) Heatmap for trophectoderm-related genes in WT, KLF4-KO, KLF5-KO, and DKO naive hESCs, primed hESCs over-expressing empty control or KLF5 (Ctrl and KLF5-OE), and primed hESCs over-expressing CRISPRa targeting scrambled sequence or LTR7Y conserved sequence (Ctrl-A and 7Y-A). RNA-seq was performed in biological triplicates for KLF depletion and KLF5-OE experiments, and in biological duplicates for CRISPRa experiment.

(D) Primed hESCs over-expressing RFP or KLF5 were differentiated toward trophectoderm for 4 days. The relative expression of pluripotency and trophectoderm marker genes was measured by qRT-PCR (mean ± SEM, n = 3, technical replicates).

(E) WT or KLF5-KO naive hESCs were differentiated toward trophectoderm for 4 days. The relative expression for pluripotency and trophectoderm marker genes was measured by qRT-PCR (mean ± SEM, n = 3, technical replicates).

(legend continued on next page)

KLF5-LTR7Y promotes extra-embryonic potential of pluripotent stem cells

TEs are increasingly recognized to be involved in species-specific transcription network (Kunarso et al., 2010). Interestingly, genes near LTR7Y loci seemed to be specifically upregulated in human naive ESCs but not in mouse naive ESCs compared with their respective primed states (Figure S5A) (Ghimire et al., 2018). Human naive pluripotency state was recently found to bear extended developmental potential toward trophectoderm lineages compared with mouse naive counterpart (Cinkornpumin et al., 2020; Dong et al., 2020; Guo et al., 2021a; Io et al., 2021). Notably, single-cell RNA-seq analysis of peri/post-implantation human embryos (Zhou et al., 2019) showed that LTR7Y was not only expressed in epiblast, but it also exhibited extensive expression in primitive endoderm (PrE) and peri-implantation (day 6) trophectoderm (TrE) lineages (Figures 5A and S5B). KLF5 exhibited higher expression correlation with LTR7Y at the single cell level in early trophectoderm as well (Figures 5B and S5B). In addition, genes related to trophectoderm development were also downregulated in KLF5-KO and DKO lines (Figures 5C and S2F; Table S1). Upon the over-expression of KLF5 or CRISPRa targeting LTR7Ys in primed hESCs, some of these trophectoderm marker genes also showed a tendency of upregulation (Figure 5C; Table S4). A closer examination of the human embryo staining further revealed that KLF5 expression was retained in the trophectoderm cells of blastocyst, which were negative for OCT4 staining (Figure S5C). Collectively, these data imply that KLF5 and its activation of the evolutionarily recent LTR7Ys potentially contribute to the expanded cell potency toward trophectoderm.

To test this, we initiated trophectoderm differentiation from hESCs (Figures S5D and S5E). The over-expression of KLF5 hastened the downregulation of pluripotency markers but increased the upregulation of trophectoderm markers (Figure 5D), while knocking out *KLF5* conversely reduced the upregulation of trophectoderm marker genes during the differentiation process (Figure 5E). In addition, activation of LTR7Ys by CRISPRa in primed hESCs during differentiation enhanced the upregulation of key trophectoderm markers (Figures 5F and 5G). To exclude the potential off-targeting effect toward LTR7s, a published sgRNA set for LTR7Y/B sub-families (Pontis et al., 2019) with less LTR7 off-targeting rate was also tested (Figure S5F). Similarly, enhanced trophectoderm differentiation was evident (Figure S5G). Overall, these data collectively indicate an intriguing role of LTR7Ys activated by KLF5 in supporting the expanded potency of hESCs toward trophectoderm lineage.

DISCUSSION

ERV is an important class of TE activated in early embryogenesis and potentially participates in pre-implantation transcription regulation (Macfarlan et al., 2012; Pontis et al., 2019). Interestingly, at different stages of early mouse and human embryos, distinct sub-families of ERVs exhibit stage-specific activation (Goke et al., 2015; Modzelewski et al., 2021). In this study, we identified how the naive and primed-specific transcription machinery are tailored to activate different LTRs of HERVH and how the naive pluripotency-specific LTR7Ys contribute to the naive transcription network and expand the developmental potential toward trophectoderm.

KLF5 rewires NANOG in the naive pluripotency transcription network

The pre- and post-implantation blastocysts correspond to different pluripotency states underlain by distinct transcription networks. Here we identified the role of Krüppel-like factor, KLF5, in regulating and rewiring human naive transcription network through NANOG. NANOG is functionally important for both the primed and naive pluripotent cell states (Chambers et al., 2003, 2007; Chan et al., 2013; Mitsui et al., 2003; Takashima et al., 2014). It binds to different regulatory sequences to control overlapping and distinct targets in primed and naive hESCs. Yet, how NANOG achieves distinct function in these two states remains elusive, as the expression level of NANOG only differs by two to three folds between primed and naive states (Chan et al., 2013; Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014). Our data hereby demonstrated that KLF5 rewires NANOG targeting affinity toward the naive-specific regulatory sequences (Figures 2G, 3C–3H, 4B, 4C, and 4H), and they coordinately promote naive-specific transcription (Figures 4D–4G, S4A, S4B, and S4E). We reckon that the change in NANOG affinity may be mediated through its direct interaction with KLF5 (Figure 3F) and the capacity of KLF5 in de-compacting and epigenetically activating the naive-specific chromatin regions (Figures 4A, S4A, and S4B). On the other hand, the interaction and synergistic effect between KLF4 and NANOG was much weaker (Figures 2G and 2H). Thus, KLF5 potentially acts as pioneer factor to recruit and redirect NANOG transcription preference in naive pluripotency state.

The activation of different ERV sub-families by cell-state-specific transcription machinery

Our study showcased how different ERV sub-families could be specifically activated by cell-state-specific transcription machineries. During early human embryogenesis, different sub-families of LTRs are activated in a stage-specific manner (Goke et al., 2015), but how their regulation specificities are achieved is not clear. We find that the LTR7 sub-family of HERVH is primarily bound and activated by NANOG (Figures 3I, S3L, and S3N), while the LTR7Y sub-family cannot

(F) Primed hESCs over-expressing CRISPRa targeting scrambled sequence or LTR7Y conserved sequence (Ctrl-CRISPRa and LTR7Y1/2/3 CRISPRa) were differentiated toward trophectoderm for 4 days. The pictures show the immunofluorescence staining of DAPI (blue) and trophectoderm-specific markers GATA3 or CK7 (red); scale bar: 50 μm.

(G) The relative expression fold change for pluripotency and trophectoderm marker genes upon 4 days of trophectoderm differentiation of primed hESCs over-expressing CRISPRa targeting LTR7Ys using different sgRNA sets (control sgRNA, LTR7Y sgRNA1/2, and LTR7Y sgRNA1/2/3) (mean ± SEM, n = 3, technical replicates).

See also Figure S5 and Tables S1 and S4.

be activated by NANOG alone but requires KLF5 to rewire NANOG affinity (Figures 2H, 3H, 4B, 4C, S3K, and S3N). Thus in naive state and potentially pre-implantation blastocyst, KLF5 together with NANOG activated LTR7Y; while in the primed hESCs and potentially post-implantation blastocyst, NANOG alone activated LTR7. We therefore propose that from ZGA to blastocyst stage, the observed temporal waves of ERV activation represent a developmental program, whereby stage-specific TFs activate and coerce different ERV sub-families to contribute to the dynamic remodeling of global transcription. The coding genome and retrotransposons coordinate and inter-regulate each other closely to accomplish rapid cell fate transition during early embryogenesis.

KLF5-LTR7Y axis promotes human-specific naive transcription and cell potential

The coordinated binding of LTR7Y by KLF5 and NANOG (Figures 2G, 2H, 3G, 3H, 4A–4C, and 4H) led to upregulation of LTR7Y nearby naive genes (Figures 4F, 4G, and S4C–4E). Interestingly, LTR7Y nearby naive genes seemed to be specifically upregulated in human naive ESCs but not in mouse naive ESCs (Figure S5A). Furthermore, closely correlated expressions of LTR7Y and KLF5 were found in the human peri-implantation trophectoderm lineage (Figures 5A, 5B, and S5B). Further examination of trophectoderm gene expression and trophectoderm differentiation potential upon KLF5 perturbation or LTR7Y activation (Figures 2F, 5C–5G, and S5F–S5G) collectively support the idea that KLF5 activates the *cis*-regulatory function of LTR7Ys to promote trophectoderm differentiation. Besides, among upregulated LTR7Y nearby genes activated by CRISPRa (Figure 4E), a few of them, such as *GATA2* and *ERVV-1*, have been implicated in trophectoderm development (Krendl et al., 2017) Currently, it is not known yet whether LTR7Y has any facilitating roles in the human implantation processes. This will be worthy of future investigation. Human ERV-related early developmental and implantation pathologies cannot be addressed in mouse models. Thus our study potentially offers a different angle to dissect the molecular basis for the numerous undiagnosed early pregnancy failures, which have not been attributed to any coding genes so far.

Limitation of the study

The taxonomy of LTRs for HERVH was based on sequence similarities. However, due to high sequence similarities shared among different LTR sub-families, inaccuracies in classification may exist. As we have identified in this study, sub-groups of LTR7s may bear more regulatory and functional similarities with LTR7Y sub-families. Thus, further refinement of the current classification may be needed in the future. Besides, due to the close sequence similarities, we may not be able to completely rule out off-target effects of LTR7Y sgRNAs toward other HERVH LTRs, like LTR7Bs. In addition, KLF5-LTR7Y axis was found to promote trophectoderm gene induction in this study; however, it is not known whether KLF5 and LTR7Y are essential in trophectoderm development or TSC formation. The functional and *in vivo* significance of KLF5-LTR7Y axis in trophectoderm development warrants further investigation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
  - Human embryo samples and ethics approval
- METHOD DETAILS
  - Reverse transfection of siRNA
  - Generating KLF knockout hESC lines using CRISPR-Cas9
  - Lentivirus production
  - RNA extraction and RT-qPCR
  - Immunofluorescence staining
  - Embryo staining
  - Luciferase assay
  - Co-immunoprecipitation
  - CRISPRa/i experiment
  - Western blotting
  - Reference genome
  - RNA-seq and data analysis
  - Single cell RNA-seq
  - Naive hESC single cell RNA-seq data analysis
  - Pre-implantation human embryogenesis single cell RNA-seq dataset analysis
  - Post-implantation human embryogenesis single cell RNA-seq dataset analysis
  - ChIP-seq, qPCR and data analysis
  - CUT&Tag and data analysis
  - ATAC-seq and data analysis
  - Motif enrichment analysis of transcription factors
  - Region associated DEG
  - sgRNA target sites prediction

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111240.

ACKNOWLEDGMENTS

This project is funded by the National Key Research and Development Program of China (2018YFC1005003), National Natural Science Foundation of China (31871372, 31950410535, 81974224, and 32170551), Zhejiang Provincial Natural Science Foundation of China (LZ22C120002 and LQ20C06004), Key Research and Development Program of Zhejiang Province (2021C03098), and the Singapore Ministry of Health's National Medical Research Council Open Fund Individual Research Grant (OFIRG16nov021). We thank Dr. Jun Ma, Xiaohang Yang, Pengfei Xu, Chih-Hung Hsu, and Xinyi Lu for their discussion and suggestions on this manuscript.

AUTHOR CONTRIBUTIONS

H.L. and Y-S.C. conceived the project. Z.A. coordinated with the experiments; Z.A. and Y.X. performed all of the experiments except for the following: X.Y.X.,

Y.S., and B.G. did the bioinformatics analysis; I.S. did the adaptation and KLF5 ChIP experiments; Y.Q. did the embryo-related experiments; X.X., C.M., and H.S. helped with Co-IP, WB, IF, and RT-qPCR; Y.S. generated the single-cell RNA-seq web analysis tool; D.C. performed the ATAC-seq library preparation; M.N.B.R. helped with the KO line generation. Y.L. and J.H. helped with the luciferase assay. H.H.N., D.Z., Y-S.C., W.L., and H.L. funded the project. H.L. oversaw the whole project. D.Z., W.L., Y-S.C., and H.L. supervised the project and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 10, 2022
Revised: May 6, 2022
Accepted: July 27, 2022
Published: August 23, 2022

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., et al. (2018). Ten things you should know about transposable elements. Genome Biol. *19*, 199.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.

Carter, A.C., Xu, J., Nakamoto, M.Y., Wei, Y., Zarnegar, B.J., Shi, Q., Broughton, J.P., Ransom, R.C., Salhotra, A., Nagaraja, S.D., et al. (2020). Spen links RNA-mediated endogenous retrovirus silencing and X chromosome inactivation. Elife *9*, 54508.

Carter, T.A., Singh, M., Dumbović, G., Chobirko, J.D., Rinn, J.L., and Feschotte, C. (2022). Mosaic cis-regulatory evolution drives transcriptional partitioning of HERVH endogenous retrovirus in the human embryo. Elife *11*, e76257.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chan, Y.S., Göke, J., Ng, J.H., Lu, X., Gonzales, K.A.U., Tan, C.P., Tng, W.Q., Hong, Z.Z., Lim, Y.S., and Ng, H.H. (2013). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. Cell Stem Cell *13*, 663–675.

Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E., and Clark, A.T. (2018). The TFAP2C-regulated OCT4 naive enhancer is involved in human germline formation. Cell Rep. *25*, 3591–3602.e5.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. *23*, 1163–1171.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science *351*, 1083–1087.

Cinkornpumin, J.K., Kwon, S.Y., Guo, Y., Hossain, I., Sirois, J., Russett, C.S., Tseng, H.W., Okae, H., Arima, T., Duchaine, T.F., et al. (2020). Naive human embryonic stem cells can give rise to cells with a trophoblast-like transcriptome and methylome. Stem Cell Rep. *15*, 198–213.

de Souza, F.S.J., Franchini, L.F., and Rubinstein, M. (2013). Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? Mol. Biol. Evol. *30*, 1239–1251.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L.A., Khan, S.A., Park, K.M., Yoon, E.J., Xing, X., et al. (2020). Derivation of trophoblast stem cells from naive human pluripotent stem cells. Elife *9*, 52504.

Fedoroff, N.V. (2012). Presidential address. Transposable elements, epigenetics, and genome evolution. Science *338*, 758–767.

Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. *13*, 283–296.

Fu, B., Ma, H., and Liu, D. (2019). Endogenous retroviruses function as gene expression regulatory elements during mammalian pre-implantation embryo development. Int. J. Mol. Sci. *20*, 790.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282–286.

Geis, F.K., and Goff, S.P. (2020). Silencing and transcriptional regulation of endogenous retroviruses: an overview. Viruses *12*, 884.

Ghimire, S., Van der Jeught, M., Neupane, J., Roost, M.S., Anckaert, J., Popovic, M., Van Nieuwerburgh, F., Mestdagh, P., Vandesompele, J., Deforce, D., et al. (2018). Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. Sci. Rep. *8*, 5884.

Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell *16*, 135–141.

Göke, J., and Ng, H.H. (2016). CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. EMBO Rep. *17*, 1131–1144.

Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221–225.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

Gunne-Braden, A., Sullivan, A., Gharibi, B., Sheriff, R.S.M., Maity, A., Wang, Y.-F., Edwards, A., Jiang, M., Howell, M., Goldstone, R., et al. (2020). GATA3 mediates a fast, irreversible commitment to BMP4-driven differentiation in human embryonic stem cells. Cell Stem Cell *26*, 693–706.e9.

Guo, G., Stirparo, G.G., Strawbridge, S.E., Spindlow, D., Yang, J., Clarke, J., Dattani, A., Yanagida, A., Li, M.A., Myers, S., et al. (2021a). Human naive epiblast cells possess unrestricted lineage potential. Cell Stem Cell *28*, 1040–1056.e6.

Guo, M., Zhang, Y., Zhou, J., Bi, Y., Xu, J., Xu, C., Kou, X., Zhao, Y., Li, Y., Tu, Z., et al. (2019). Precise temporal regulation of Dux is important for embryo development. Cell Res. *29*, 956–959.

Guo, Y., Xue, Z., Yuan, R., Li, J.J., Pastor, W.A., and Liu, W. (2021b). RAD: a web application to identify region associated differentially expressed genes. Bioinformatics *37*, 2741–2743.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017).

Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. *49*, 925–934.

Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., et al. (2021). Ensembl 2021. Nucleic Acids Res. *49*. D884–d891.

Io, S., Kabata, M., Iemura, Y., Semi, K., Morone, N., Minagawa, A., Wang, B., Okamoto, I., Nakamura, T., Kojima, Y., et al. (2021). Capturing human trophoblast development with naive pluripotent stem cells in vitro. Cell Stem Cell *28*, 1023–1039.e13.

Imbeault, M., Helleboid, P.Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature *543*, 550–554.

Jönsson, M.E., Garza, R., Johansson, P.A., and Jakobsson, J. (2020). Transposable elements: a common feature of neurodevelopmental and neurodegenerative disorders. Trends Genet. *36*, 610–623.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat. Cell Biol. *10*, 353–360.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat. Commun. *10*, 1930.

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626–1632.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.

Klein, S.J., and O'Neill, R.J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosome Res. *26*, 5–23.

Krendl, C., Shaposhnikov, D., Rishko, V., and Drukker, M. (2017). GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. PNAS *114*, E9579–E9588. https://doi.org/10.1073/pnas.1708341114.

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. *42*, 631–634.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

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

Lu, X., Sachs, F., Ramsay, L., Jacques, P.É., Göke, J., Bourque, G., and Ng, H.-H. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol. *21*, 423–425.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57–63.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10–12.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Modzelewski, A.J., Shao, W., Chen, J., Lee, A., Qi, X., Noon, M., Tjokro, K., Sales, G., Biton, A., Anand, A., et al. (2021). A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. Cell *184*, 5541–5558.e22.

Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., and Clark, A.T. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. Nat. Cell Biol. *20*, 553–564.

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T.W., Jaenisch, R., and Trono, D. (2019). Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. Cell Stem Cell *24*, 724–735.e5.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.

Senft, A.D., and Macfarlan, T.S. (2021). Transposable elements shape the evolution of mammalian development. Nat. Rev. Genet. *22*, 691–711.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S.Y., Hahn, W.C., Sharp, P.A., et al. (2003 Apr). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA *9*, 493–501.

Sun, Y., Keown, J.R., Black, M.M., Raclot, C., Demarais, N., Trono, D., Turelli, P., and Goldstone, D.C. (2019). A dissection of oligomerization by the TRIM28 tripartite motif and the interaction with members of the krab-ZFP family. J. Mol. Biol. *431*, 2511–2527.

Szczerbinska, I., Gonzales, K.A.U., Cukuroglu, E., Ramli, M.N.B., Lee, B.P.G., Tan, C.P., Wong, C.K., Rancati, G.I., Liang, H., Göke, J., et al. (2019). A chemically defined feeder-free system for the establishment and maintenance of the human naive pluripotent state. Stem Cell Rep. *13*, 612–626.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 524–526.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405–409.

Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nat. Genet. *49*, 935–940.

Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis (New York: Springer-Verlag).

Wu, K., Liu, H., Wang, Y., He, J., Xu, S., Chen, Y., Kuang, J., Liu, J., Guo, L., Li, D., et al. (2020). SETDB1-Mediated cell fate transition between 2C-like and pluripotent states. Cell Rep. *30*, 25–36.e6.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature *500*, 593–597.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

Yang, F., Huang, X., Zang, R., Chen, J., Fidalgo, M., Sanchez-Priego, C., Yang, J., Caichen, A., Ma, F., Macfarlan, T., et al. (2020). DUX-miR-344-ZMYM2-mediated activation of MERVL LTRs induces a totipotent 2C-like state. Cell Stem Cell *26*, 234–250.e7.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS A J. Integr. Biol. *16*, 284–287.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

Zhou, F., Wang, R., Yuan, P., Ren, Y., Mao, Y., Li, R., Lian, Y., Li, J., Wen, L., Yan, L., et al. (2019). Reconstituting the transcriptome and DNA methylome landscapes of human implantation. Nature *572*, 660–664.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-NANOG	R&D Systems	Cat#AF1997; RRID: AB_355097
anti-HA	Sigma-Aldrich	Cat#H3663; RRID: AB_262051
anti-KLF4	Abcam	Cat#ab106629; RRID: AB_10858595
anti-KLF5	Homemade serum	N/A
anti-KLF5	Active Motif	Cat#61099; RRID: AB_2614941
anti-KLF17	Sigma-Aldrich	Cat#HPA024629; RRID: AB_1668927
anti-OCT4	Santa Cruz,	Cat#sc-5279; RRID: AB_628051
anti-GATA3	Thermo Fisher Scientific	Cat#MA1-028; RRID: AB_2536713
anti-CK7	Leica Biosystems	Cat#NCL-L-CK7-560; RRID: AB_563826
anti-GAPDH	Santa Cruz	Cat#sc-25778; RRID: AB_10167668
anti-H3K27ac	Abclonal	Cat#A7253; RRID: AB_2767797
anti-P300	Santa Cruz	Cat#sc-48343; RRID: AB_628075
Mouse Control IgG	ABclonal	Cat#AC011; RRID: AB_2770414
anti-Mouse IgG (H&L) HRP	GenScript	Cat#A00160; RRID: AB_1968937
anti-Rabbit IgG (H&L) HRP	GenScript	Cat#A00098; RRID: AB_1968815
HRP-conjugated AffiniPure Goat Anti-Mouse IgG Light Chain	Abclonal	Cat#AS062; RRID: AB_2864056
Chicken anti-Mouse IgG (H + L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 594	Thermo Fisher Scientific	Cat#A-21201; RRID: AB_2535787
Donkey anti-Rabbit IgG (H + L) ReadyProbes Secondary Antibody, Alexa Fluor™ 488	Thermo Fisher Scientific	Cat#R37118; RRID: AB_2556546
Bacterial and virus strains		
DH5α	TransGen Biotech	Cat#CD201-02
Stbl3	TransGen Biotech	Cat#CD521-02
Biological samples		
Human embryos	This study	N/A
Chemicals, peptides, and recombinant proteins		
FGF2	Gibco	Cat#PHG0023
Activin A	StemCell Technologies	Cat#78001.1
Recombinant Human LIF	PeproTech	Cat#300-05-50UG
PD0325901	Sigma-Aldrich	Cat#PZ0162-5MG
SB590885	Sigma-Aldrich	Cat#SML0501-5MG
DASA	Selleckchem	Cat#S1021
AZD5438	Tocris	Cat#3968
Y-27632	StemCell Technologies	Cat#72308
BMP4	R&D Systems	Cat#314-BP-050
A83-01	Stemgent	Cat#04-0014-10
Thiazovivin	Tocris	Cat#3845/10
BSA	Sigma-Aldrich	Cat#A1933
L-Glutamine	Gibco	Cat#25030081
2-Mercaptoethanol	Gibco	Cat#21985023
MEM Non-Essential Amino Acids Solution	Gibco	Cat#11140050

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DMEM High Glucose	Hyclone	Cat#SH30243-01
basal mTeSR	StemCell Technologies	Cat#05896
mTeSR1 Complete Kit	StemCell Technologies	Cat#85850
Neurobasal Medium	Gibco	Cat#21103049
DMEM/F12	Nacalai Tesque	Cat#08460-95
N2 Supplement	Gibco	Cat#17502048
B27 Supplement	Gibco	Cat#17504044
Matrigel hESC-qualified matrix	Corning	Cat#354277
Reduce Growth factor matrigel	Corning	Cat#354230
Dispase (1U/mL)	StemCell Technologies	Cat#07923
TrypLE™ Express Enzyme	Thermo Fisher Scientific	Cat#12604013
Fetal Bovine Serum	Gibco	Cat#10270-106
Dynabeads™ protein G	Thermo Fisher Scientific	Cat#10003D
Protease inhibitor cocktail	Biotool	Cat#B14001
Paraformaldehyde	Sigma-Aldrich	Cat#P6148-500G
DAPI	Solarbio	Cat#C0065
Virus Concentration Solution	Origene	Cat#TR30026
EndoFectin Max transfection reagent	GeneCopoeia	Cat#EF014
RNAiMAX	Thermo Fisher Scientific	Cat#13778150
Polyethylene Glycol (PEI)	Polysciences	Cat#24765-1
Critical commercial assays		
SimpleChIP® Plus Enzymatic Chromatin IP Kit	Cell signaling Technology	Cat#9005
Dual Luciferase Reporter Gene Assay Kit	Promega	Cat#E1910
HiScript III 1st Strand cDNA Synthesis Kit (+gDNA wiper)	Vazyme Biotech	Cat#R312-01
Fast SYBR Green Master Mix	Thermo Fisher Scientific	Cat#4385612
TruePrep DNA Library Prep Kit V2 for Illumina	Vazyme Biotech	Cat#TD501
Chromium Single Cell 3′GEM library and Gel Bead Kit V3	10x Genomics	Cat#P/N1000075
Truseq RNA Sample Prep Kit v2	Illumina	Cat#RS-122-2001
FastPure DNA Extraction Mini Kit	Vazyme Biotech	Cat#DC301
Hyperactiv In-Situ ChIP Library Prep Kit	Vazyme Biotech	Cat#TD902
TruePrep Index Kit V2	Vazyme Biotech	Cat#TD202
MagicSeq Tn5 DNA Library Prep Kit for Illumina	Magic-Bio	Cat#M3141
MinElute PCR Purification Kit	QIAGEN	Cat#28006
1xNEBnext PCR master mix	New England Biolabs	Cat#M0541S
TransStart FastPfu DNA Polymerase	TransGen Biotech	Cat#AP221-02
RNAios Plus	Takara	Cat#9109
Deposited data		
Naive hESCs: single cell RNA-seq	This study	GEO: GSE169678
Primed hESCs (Ctrl and KLF5-OE): Bulk RNA-seq	This study	GEO: GSE169678
Primed hESCs (CRISPRa-Ctrl and CRISPRa-LTR7Y): Bulk RNA-seq	This study	GEO: GSE169678
Naive hESCs (WT, KLF4-KO, KLF5-KO, KLF4/KLF5-Double KO): Bulk RNA-seq	This study	GEO: GSE169678
Naive hESCs: KLF5 ChIP-seq	This study	GEO: GSE169678

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Naive hESCs: NANOG ChIP-seq	This study	GEO: GSE169678
Naive hESCs: H3K27ac ChIP-seq	This study	GEO: GSE169678
Primed hESCs (Ctrl and KLF5-OE): NANOG ChIP-seq	This study	GEO: GSE169678
Primed hESCs (Ctrl and KLF5-OE): H3K27ac ChIP-seq	This study	GEO: GSE169678
Naive hESCs (WT and KLF4/KLF5-Double KO): NANOG CUT&Tag	This study	GEO: GSE169678
hESCs (Primed Ctrl, Primed KLF5-OE and Naive Ctrl): ATAC-seq	This study	GEO: GSE169678
Experimental models: Cell lines		
HEK-293T	ATCC	Cat#CRL-3216
Primed hESC (the H1 line)	Wicell	Cat#WA01-pcbc
Primed hESCs (the H1 KLF4 KO line)	This study	N/A
Primed hESCs (the H1 KLF5 KO line)	This study	N/A
Primed hESCs (the H1 KLF4/KLF5 double KO line)	This study	N/A
Naive hESCs (the H1 line)	This study	N/A
Naive hESCs (the H1 KLF4 KO line)	This study	N/A
Naive hESCs (the H1 KLF5 KO line)	This study	N/A
Naive hESCs (the H1 KLF4/KLF5 Double KO line)	This study	N/A
Oligonucleotides		
Primers for RT-qPCR and ChIP assay	See Table S5	N/A
Sequences of sgRNA	See Table S5	N/A
Sequences of siRNA	See Table S5	N/A
Recombinant DNA		
psi-CHECK2-LTR7Y_dup216	This study	N/A
psi-CHECK2-LTR7_dup804	This study	N/A
psi-CHECK2-LTR7_dup647	This study	N/A
psPAX2	Gift from Didier Trono	Addgene#12260
pCMV-VSVG	Stewart et al., 2003	Addgene#8454
pLVX-empty control	This study	N/A
pLVX-HABP4	This study	N/A
pLVX-CNOT3	This study	N/A
pLVX-ZNF789	This study	N/A
pLVX-ZNF134	This study	N/A
pLVX-SSX4	This study	N/A
pLVX-SOD2	This study	N/A
pLVX-ILF2	This study	N/A
pLVX-SNRPD1	This study	N/A
pLVX-DMRTC2	This study	N/A
pLVX-TCL1A	This study	N/A
pLVX-BLZF1	This study	N/A
pLVX-GABPB1	This study	N/A
pLVX-RNF138	This study	N/A
pLVX-DPRX	This study	N/A
pLVX-TAF5L	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pLVX-SNAI1	This study	N/A
pLVX-KLF11	This study	N/A
pLVX-KLF2	This study	N/A
pLVX-KLF11	This study	N/A
pLVX-KLF17	This study	N/A
pLVX-KLF4	This study	N/A
pLVX-KLF5	This study	N/A
pLVX-KLF4-3HA	This study	N/A
pLVX-KLF5-3HA	This study	N/A
pLVX-NANOG	This study	N/A
CRISPRa: pAC154-dual-dCas9VP160-sgExpression	Cheng et al., 2013	Addgene#48240
CRISPRi: pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP	Thakore et al., 2015	Addgene#71237
dCas9-VP160-LTR7Y sgRNA 1	This study	N/A
dCas9-VP160-LTR7Y sgRNA 2	This study	N/A
dCas9-VP160-LTR7Y sgRNA 3	This study	N/A
dCas9-VP160-Scrambled sgRNA	This study	N/A
dCas9-KRAB-LTR7Y-sgRNA 1	This study	N/A
dCas9-KRAB-LTR7Y-sgRNA 2	This study	N/A
dCas9-KRAB-LTR7Y-sgRNA 3	This study	N/A
dCas9-KRAB-Scrambled sgRNA	This study	N/A
Software and algorithms		
ImageJ	Open Source/National Institutes of Health	https://imagej.nih.gov/ij/
GraphPad Prism 5.0 software	GraphPad Software, Inc.	http://www.graphpad.com/scientificsoftware/prism/
igv-2.3.72g	Thorvaldsdóttir et al., 2013	http://software.broadinstitute.org/software/igv/
FastQC v0.11.8	https://www.bioinformatics.babraham.ac.uk/projects/	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
STAR v2.7.0e	Dobin et al., 2013	https://github.com/alexdobin/STAR
DESeq2 v1.26.0	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
clusterProfiler v3.12.0	Yu et al., 2012	http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html
cutadapt v2.9	Martin, 2011	https://github.com/marcelm/cutadapt/
FeatureCounts v2.0.0	Liao et al., 2014	http://subread.sourceforge.net/
Seurat v4.0.5	Hao et al., 2021	https://satijalab.org/seurat/
deeptools v3.4.3	Ramírez et al., 2014	https://deeptools.readthedocs.io/en/develop/
ggplot2 v3.3.5	N/A	https://sourceforge.net/projects/ggplot2.mirror/files/v3.3.5/
Cell Ranger 6.0.2	10 X Genomics	https://github.com/10XGenomics/cellranger/
Other		
Human pre-implantation embryos: scRNA-seq	Yan et al., 2013	GEO: GSE36552
Human post-implantation embryos: scRNA-seq	Zhou et al., 2019	GEO: GSE109555

RESOURCE AVAILABILITY

Lead contact

All information and requests should be directed to the lead contact: Hongqing Liang (lianghongqing@zju.edu.cn).

Materials availability

All stable reagents, plasmids, and cell lines generated in this study are available from the lead contact.

Data and code availability

- The raw data of RNA-seq, CUT&Tag, ATAC-seq, ChIP-seq and single cell RNA-seq have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Data referenced in this study are available in Gene Expression Omnibus (GEO) with the references and accession numbers listed in the key resources table.
- The codes of bioinformatics analysis have been deposited at GitHub: https://github.com/XinyuXiang/LTR7YProjectAnalysis and are publicly available as of the date of publication.
- Any additional information required to re-analyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

HEK293T cells were cultured on 0.12% gelatin pre-coated dishes in DMEM high glucose (HyClone, SH30243.01) supplemented with 10% FBS (Gibco, 10,270-106). Primed hESC line H1 were cultured in mTeSR (Stem cell technologies, 85,851) on matrigel (Corning, 354,277) pre-coated plate in normal O_2, 5% CO_2 incubator. H1 was passaged at 1:6 ratios every 5–6 days upon Dispase (Stem Cell Technologies, 07,923) dissociation for 5 min at 37°C. Naive hESC line H1 was adapted from primed H1 lines using feeder-independent naive embryonic (FINE) stem cells culture system on plates pre-coated with diluted reduce growth factor matrigel (Corning, 354230) as described by Szczerbinska (Szczerbinska et al., 2019; Yang et al., 2020). Adapted H1 cells were grown in the FINE medium in hypoxia conditions (5% O_2, 5% CO_2). Every 5–7 days, they were passaged at 1:6 ratio upon digestion with TrypLE Express (Gibco,12604013). Trophoblast differentiation was performed as previously described (Gunne-Braden et al., 2020; Io et al., 2021), briefly hESCs were plated on ColIV coated plate and stimulated with 2 mM PD035901, 2 mM A8301 and 100 ng/mL BMP4 in custom basal TeSR medium (basal mTeSR medium, 05,896) without TGFβ and bFGF.

Human embryo samples and ethics approval

The work performed in the National Research Center for Assisted Reproductive Technology and Reproductive Genetics was conducted under the regulations of the Human Biomedical Research Ethics Guidelines (regulated by the National Health Commission of the People's Republic of China on 1 December 2016), the 2016 Guidelines for Stem Cell Research and Clinical Translation (issued by the International Society for Stem Cell Research, ISSCR) and the Human Embryonic Stem Cell Research Ethics Guidelines (regulated by China National Center for Biotechnology Development on 24 December 2003). Approval number: IRB-20200268-R.

METHOD DETAILS

Reverse transfection of siRNA

1 μM siRNA were mixed with RNAiMAX (Thermo Fisher Scientific, 13778150) diluted with Opti-MEM (Gibco, 31985070), and incubated at room temperature (RT) for 20 min before addition into wells. Subsequently, ESCs were seeded into the wells with medium and 1 μM Thiazovivin (Tocris, 3845). The medium containing transfection mix was replaced on the following day.

Generating KLF knockout hESC lines using CRISPR-Cas9

H1 hESCs cells were transfected with CRISPR-Cas9 with targeting sgRNAs by electroporation. Cells were seeded at single cells density and selected in mTeSR +1 μM Thiazovivin medium with 1 μg/mL Puromycin. Monoclonal lines were picked up and amplified upon single cell seeded grown into a colony. Knockout clones were identified by immunofluorescence staining, PCR, and Sanger sequencing.

Lentivirus production

HEK293T cells were grown to around 70% confluency. The transfer plasmid, psPAX2 (Addgene#12260), and pCMV-VSVG (Addgene#8454) were mixed at the ratio of 4:4:1 and incubated with 1 μg/mL of Polyethylene Glycol (PEI) (Polysciences, 24765-1) diluted in Opti-MEM for 25 min. The mix was added drop by drop to HEK293T cells, and gently mixed. After 48 hours and 72 hours, the supernatant was harvested and filtered. The virus was precipitated using Virus Concentration Solution (Origene, TR30026) by centrifugation at 3500 g. The virus pellet was resuspended in PBS and aliquoted before storing at -80°C.

RNA extraction and RT-qPCR

Total RNA was extracted by RNAios Plus (Takara, 9109). 500 ng of total RNA was used for cDNA synthesis with HiScript III 1st Strand cDNA Synthesis Kit + gDNA wiper (Vazyme Biotech, R312-01). qPCR was performed using Fast SYBR Green Master Mix (Thermo Fisher Scientific, 4385612) in the Roche LightCycler® 480 (LC480) system.

Immunofluorescence staining

Cells grown on plastic coverslips were fixed with 4% paraformaldehyde (PFA) for 15 min, permeabilized with 0.5% Triton X-100/PBS for 30 min and blocked in 2% BSA in 0.1% Triton X-100/PBS for 30 min at RT. Cells were incubated with primary antibody overnight at 4°C, and fluorophore-conjugated secondary antibodies (Thermo Fisher Scientific, A-21201 and R37118) and 5 μg/mL DAPI (Solarbio, C0065) for 2 hours at RT. Antibodies were all diluted in 2% BSA with 0.1% Triton X-100/PBS. Coverslips were mounted onto carrying slides and images were taken under Nikon A2 confocal microscope. Primary antibodies used: anti-NANOG (R&D Systems, AF1997), anti-KLF5 (homemade serum), anti-KLF4 (Abcam, ab106629), anti-KLF17 (Sigma-Aldrich, HPA024629), anti-OCT4 (Santa Cruz, sc-5279), anti-GATA3 (Thermo Fisher Scientific, MA1-028), and anti-CK7 (Leica Biosystems, NCL-L-CK7-560).

Embryo staining

Human embryos used in this study were donated surplus 3PN zygotes generated from *in vitro* fertilization. The embryos were fixed in 4% PFA, washed twice in PBST, and permeabilized with 0.5% Triton X-100 in PBS at RT overnight. Upon two washes in PBST and blocking for 4 hours in blocking buffer (3% BSA in PBS). Embryos were transferred to a new well, washed three times in PBST, and then incubated overnight with primary antibodies in blocking buffer at 4°C. Upon three washes in PBST, embryos were incubated for 16-18 hours with secondary antibodies and 5 μg/mL DAPI. After washing three times with PBST, they were transferred to a glass-bottom dish (JingAn Biological, J40201) with 60% glycerol aqueous solution for confocal imaging. The staining of different marker proteins was quantified by ImageJ: briefly, a nuclear mask using signals from the DAPI channel was generated, and the nuclear average intensity from single nucleus in embryos at different stages were quantified.

Luciferase assay

The LTR7Y (dup 216, Chr17: 34183191-34183623), LTR7 (dup 804, Chr4: 186241957-186242322; dup 647, Chr4: 61764219-61764699) sequences were cloned into luciferase reporter plasmid psi-CHECK-2 (Promega, C8021), and transfected into HEK293T cell using PEI together with pLVX plasmid containing the respective transcription factor at a 1:1 ratio. Forty-eight hours after transfection, luciferase activities were measured by a Dual-luciferase assay system (Promega, E1910) and analyzed on Synergy 2 plate reader (BioTek).

Co-immunoprecipitation

Cells were lysed in Lysis Buffer with protease inhibitor cocktail (Biotool, B14001) on ice for 30 min, and centrifuged at 4°C for 25 min. 500 μg total protein was incubated with 5 μg of antibodies, including mouse IgG as control, HA (Sigma-Aldrich, H3663) and NANOG (R&D Systems, AF1997) antibody at 4°C overnight with rotation. Lysates and antibody mix were incubated with 30 μL protein G Dynabeads (Thermo Fisher Scientific, 10003D) for 2 hours at RT with rotation. Beads were washed with lysis buffer and eluted.

CRISPRa/i experiment

sgRNAs targeting LTR7Y were designed from LTR7Y consensus sequence in Dfam and cloned into dCAS9-VP160 or dCAS9-KRAB vectors modified from Addgene #48240 or #71237. The three sgRNA sequences were as shown below:

TCGCATCCCCTGTGACTTGC;
GCAGGAACAAATCACAATGG;
TTGCTCACACAAAGCCTGTT.

dCAS9-VP160 or dCAS9-KRAB vectors with sgRNAs were transduced into hESCs. hESCs were assayed or directed to trophectoderm differentiation as described in the previous session.

Western blotting

Cells lysate in RIPA buffer with protease inhibitor cocktail or Co-IP samples were boiled for 5 min at 95°C and kept on ice. Protein samples were resolved using 10–12.5% SDS-PAGE and transferred onto a polyvinylidene difluoride (PVDF) membrane (Sigma-Aldrich, 3010040001) under a wet tank transfer system. The PVDF membrane was blocked in 5% non-fat milk in PBST for 1 hours and incubated the primary antibody diluted in blocking buffer at 4°C overnight and subsequently secondary antibody conjugated with HRP (GenScript, A00098). The membrane was visualized using the Azure C300 system.

Reference genome

Human reference genome GRCh38.97 from Ensembl (Howe et al., 2021) was used for alignment. TE annotation file GRCh38 from repeatmasker (http://repeatmasker.org/) and gene annotation file GRCh38.97 from Ensembl (Howe et al., 2021) was applied to all genomics analyses.

RNA-seq and data analysis

The library was constructed with Truseq RNA Sample Prep Kit v2 (Illumina, RS-122-2001) before being subjected to Illumina platform sequencing with a depth of 20M reads per sample. Quality control for raw reads was performed by FastQC (Andrews, 2010) v0.11.8. The first 10 bp of both paired-end reads were trimmed by cutadapt v2.9 (Martin, 2011). STAR v2.7.0e (Dobin et al., 2013) was used for alignment onto human reference genome GRCh38.97. Reads with maximal 1000 multiple mapped sites and no more than 3

mismatches were retained, and only the best hit was kept (–outFilterMultimapNmax 1000, –outFilterMismatchNmax 3, –outSAMmultNmax 1). Quantification for both gene and TE were calculated by FeatureCounts v2.0.0 (Liao et al., 2014). For gene quantification, only unique mapped reads were included (-p -B -P -C), while for TE quantification, multiple mapped reads were also included (-p -B -P -C -M). Differentially expressed genes (DEGs) and transposable elements (DE-TEs) were generated by R package DESeq2 v1.26.0 (Love et al., 2014). Only features with mean RPKM greater than 1 in either control or treatment group were retained. Differentially expressed features were obtained from two-fold change and 0.05 FDR filtering criteria. GO term enrichment for DEGs was conducted by R package clusterProfiler v3.12.0 (Yu et al., 2012) (ont = 'BP', pAdjustMethod = 'BH', pvalueCutoff = 0.05, qvalueCutoff = 0.05).

For down-stream data visualization, PCA plot was plotted using R package DESeq2 v1.26.0 (Love et al., 2014) and ggplot2 v3.3.5 (Wickham, 2016), with all gene or TE features included. To visualize the expression pattern of genes, heatmap of selected genes was plotted using R package ComplexHeatmap v2.0.0 (Gu et al., 2016). Gene expression level was normalized by Z-score of RPKM (Reads Per Kilobase per Million mapped reads). For the analysis of the top 10 up- or down-regulated DE-TE sub-families, barplot was generated by ggplot2 v3.3.5 (Wickham, 2016). The fraction of DE-TE in the sub-family was obtained by dividing the copy number of DE-TEs by the total TE copy number from the same sub-family.

Single cell RNA-seq

The cultured cells in 6-well-plate were digested by TrypLE with collagenase I. Add 1 mL of FINE + TZV media to inactive TrypLE. Cells were gently pipeted up and down. Then the cells were spun down at 1200 rpm for 5 min. The obtained cell pellet was washed twice and resuspended in DPBS. Cell viability was assessed by Propidium Iodide (MedChemExpress, HY-D0815). Single cell suspensions were loaded onto the 10×Genomics to generate single-cell GEMs. Single cell RNA-seq libraries were prepared using Chromium Single Cell 3′GEM library and Gel Bead Kit V3 (10x Genomics, P/N 1000075). All the libraries were sequenced on the Illumina Novaseq platform by Novogene.

Naive hESC single cell RNA-seq data analysis

The reads were aligned and quantified by Cell Ranger v3.1.0 (Zheng et al., 2017) from 10x Genomics to both gene and TE reference genome with default parameters. The cell-by-gene/TE unique molecular identifier (UMI) count matrix was processed in R package Seurat v4.0.5 (Hao et al., 2021) for downstream data visualization. We aggregated reads from individual TE copies to subfamilies due to the limited coverage in single cell RNA-seq data.

We retained cells that expressed over 200 gene features and 100 TE features with fewer than 30% mitochondrial genes. The UMI count matrices for both gene and TE were then Log normalized and scaled (scale. factor = 10,000 for genes, scale. factor = 1000 for TEs). The scaled data with the top 2000 variable genes were used to perform principal component analysis (PCA). UMAPs were calculated by RunUMAP function using top 20 principal components. More information of UMAPs showing the expression of different genes can be found at https://labw.org/hESCTE.

Pre-implantation human embryogenesis single cell RNA-seq dataset analysis

Single cell RNA-seq data of pre-implantation human embryos were downloaded from previous publications:

GEO: GSE36552 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36552].

This TrueSeq derived single cell RNA-seq dataset was aligned and quantified using the same workflow as bulk RNA-seq data. Quality control for raw reads was performed by FastQC (Andrews, 2010) v0.11.8. The first 10 bp and last 40 bp of reads were trimmed by cutadapt v2.9 (Martin, 2011). STAR v2.7.0e (Dobin et al., 2013) was used for alignment onto human reference genome GRCh38.97. Reads with maximal 1000 multiple mapped sites were retained and only the best hit was kept (–outFilterMultimapNmax 1000, –outSAMmultNmax 1). Read quantification for both gene and TE were calculated by FeatureCounts v2.0.0 (Liao et al., 2014). For gene quantification, only unique mapped reads were included, while for TE quantification, multiple mapped reads were also included (-M). The expression pattern of selected genes and TEs was plotted using R package ComplexHeatmap v2.0.0 (Gu et al., 2016), normalized by Log_2(RPKM+1).

Post-implantation human embryogenesis single cell RNA-seq dataset analysis

Single cell RNA-seq data of post-implantation human embryos were downloaded from previous publications:

GEO: GSE109555 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE109555].

This modified STRT-seq derived single cell RNA-seq dataset was aligned and quantified with workflow adapted from Zhou et al. (Zhou et al., 2019). Quality control for raw reads was performed by FastQC v0.11.8 (Andrews, 2010). STAR v2.7.0e (Dobin et al., 2013) was used for alignment onto human reference genome GRCh38.97. Reads with maximal 1000 multiple mapped sites were retained and no more than 3 mismatches were retained, and only the best hit was kept (–outFilterMultimapNmax 1000, –outFilterMismatchNmax 3, –outSAMmultNmax 1). Read quantification for both gene and TE were calculated by HTseq v1.99.2 (Anders et al., 2015) with customized script provided from Zhou et al. (Zhou et al., 2019).

ChIP-seq, qPCR and data analysis

Cells were harvested and fixed with 1% formaldehyde (Sigma-Aldrich, 47,608–250ML-F) for 10 min at RT with rotation; Quenching was done with 0.14 M glycine at RT for 10 min. Cells were lysed by ChIP lysis buffer (10 mM Tris-HCl (pH 8.0), 0.25% Triton X-100,

10 mM EDTA, 100 mM NaCl, protease inhibitor cocktail). The genomic DNA was sonicated into short fragments with an average size of 500 bp. The fragmented DNA was incubated with KLF5 (Active Motif, 61099), NANOG (R&D Systems, AF1997), or H3K27ac (Abclonal, A7253) antibodies respectively overnight at 4°C. The mix was subsequently incubated with 30 μL protein G Dynabeads for 2 hours at RT with rotation. DNA was eluted in elution buffer (50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1% SDS), treated with proteinase K at 60°C overnight. DNA was purified by FastPure DNA Extraction Mini Kit (Vazyme Biotech, DC301) and subjected to qPCR or Illumina sequencing.

For ChIP-seq data analysis, quality control was performed by FastQC (v0.11.8) (Andrews, 2010). First 10 bp and last 90 bp of both paired-end reads were trimmed by cutadapt v2.9 (Martin, 2011). STAR v2.7.0e (Dobin et al., 2013) was used for alignment onto human reference genome GRCh38.97. Reads with maximal 1000 multiple mapped sites and no more than 3 mismatches were retained, and only the best hit was kept (–outFilterMultimapNmax 1000, –outFilterMismatchNmax 3, –outSAMmultNmax 1). By creating STAR index without general feature format file and not allowing intron length (–alignIntronMax 1), the splice junction was neglected. PCR duplicates were removed by Samtools v1.2 (Li et al., 2009) rmdup function.

ChIP-seq peaks were defined using MACS2 v2.2.7.1 (Zhang et al., 2008) callpeaks function with default parameters. Bigwig tracks were generated using deeptools v3.4.3 (Ramírez et al., 2014) by normalizing to RPKM using binsize of 10 bp. ChIP-seq signals over genomic regions were plotted by deeptools v3.2.1 (Ramírez et al., 2014), or visualized by Integrative Genomics Viewer igv-2.3.72g (Thorvaldsdóttir et al., 2013). For TE annotation of peak region, annotation of different TE families/classes over peak regions (Observed) was obtained by annotatePeaks.pl function in Homer v4.7 (Heinz et al., 2010) by using TE annotation file. Randomly distributed regions with the same number and length as the peak regions were generated by bedtools v2.29.2 (Pastor et al., 2018; Quinlan and Hall, 2010) shuffle function as control (Expected). For TE composition, percentages of the annotated TE number from each class over both peak regions and randomly distributed regions were calculated.

CUT&Tag and data analysis

CUT&Tag was performed as previously described (Kaya-Okur et al., 2019) with Hyperactiv In-Situ ChIP Library Prep Kit (Vazyme Biotech, TD902). Briefly, 1 × 10^5 cells were incubated with 10 μL ConA beads for 10 min. Beads-bound cells were resuspended in 50 μL of antibody buffer with 1.25 μg NANOG antibody (R&D Systems, AF1997) and incubated overnight at 4°C with slow rotation. Secondary antibody was diluted in 50 μL Dig-wash buffer and cultured with samples for one hour at room temperature. 0.58 μL pA-Tn5 to 100 μL Dig-300 buffer was added. Samples were incubated with pA-Tn5 at room temperature for one hour and washed twice with dig-wash buffer. 300 μL of tagmentation buffer was added to each reaction, and incubated for 37°C. To terminate tagmentation, 10 μL of 0.5 M EDTA, 1.5 μL of 20% SDS, and 2.5 μL 20 mg/mL proteinase K were added to each reaction. DNA was extracted using the Phenol-Chloroform method. The DNA libraries were amplified by TruePrep Index Kit V2 (Vazyme Biotech, TD202). All libraries were sequenced by Illumina novaseq 6000 according to the manufacturer's instructions.

For CUT&Tag data analysis, quality control was performed by FastQC (v0.11.8) (Andrews, 2010). Tn5 adapter (AGATGTGTATAAGAGACAG) was trimmed by cutadapt v2.9 (Martin, 2011). STAR v2.7.0e (Dobin et al., 2013) was used for alignment onto human reference genome GRCh38.97. Reads with maximal 1000 multiple mapped sites and no more than 3 mismatches were retained, and only the best hit was kept (–outFilterMultimapNmax 1000, –outFilterMismatchNmax 3, –outSAMmultNmax 1). By creating STAR index without general feature format file and not allowing intron length (–alignIntronMax 1), the splice junction was neglected. PCR duplicates were removed by Samtools v1.2 (Li et al., 2009) rmdup function.

CUT&Tag peaks were defined using MACS2 v2.2.7.1 (Zhang et al., 2008) callpeaks function with default parameters. Bigwig tracks were generated using deeptools v3.4.3 (Ramírez et al., 2014) by normalizing to RPKM using binsize of 10 bp. CUT&Tag signals over genomic regions were plotted by deeptools v3.2.1 (Ramírez et al., 2014).

ATAC-seq and data analysis

ATAC-seq was performed as previously described (Chen et al., 2018) using MagicSeq Tn5 DNA Library Prep Kit for Illumina (MagicBio, M3141). hESC pellet was treated transposase at 37°C for 30 min, purified using MinElute PCR Purification Kit (QIAGEN, 28006), and amplified using 1xNEBnext PCR master mix (NEB, M0541S) using custom Nextera PCR primers 1 and 2. Upon purification with MinElute PCR Purification Kit (QIAGEN, 28006). Libraries were subjected to Nova pair-end 150bp sequencing.

For ATAC-seq data analysis, quality control was performed by FastQC (v0.11.8) (Andrews, 2010). Tn5 adapter (AGATGTGTA TAAGAGACAG) was trimmed by cutadapt v2.9 (Martin, 2011). STAR v2.7.0e (Dobin et al., 2013) was used for alignment onto human reference genome GRCh38.97. Reads with maximal 1000 multiple mapped sites and no more than 3 mismatches were retained, and only the best hit was kept (–outFilterMultimapNmax 1000, –outFilterMismatchNmax 3, –outSAMmultNmax 1). By creating STAR index without general feature format file and not allowing intron length (–alignIntronMax 1), the splice junction was neglected. PCR duplicates were removed by Samtools v1.2 (Li et al., 2009) rmdup function.

ATAC-seq peaks were defined using MACS2 v2.2.7.1 (Zhang et al., 2008) callpeaks function with default parameters. Bigwig tracks were generated using deeptools v3.4.3 (Ramírez et al., 2014) by normalizing to RPKM using binsize of 10bp. ATAC-seq signals over genomic regions were plotted by deeptools v3.2.1 (Ramírez et al., 2014).

Motif enrichment analysis of transcription factors

Motif files of transcriptional factors (TFs) were obtained from Homer v4.7 (Heinz et al., 2010) and the position for each motif in GRCh38 genome were calculated by Homer scanMotifGenomeWide.pl function then transformed to bigwig format by bedGraphToBigWig v2.8 (Kent et al., 2010) with log-odds score. Motif enrichment of KLF5 and NANOG over LTR7Ys and LTR7s were calculated by bedtools v2.29.2 (Quinlan and Hall, 2010) intersect function.

Region associated DEG

Website application RAD (https://labw.org/rad) from Guo et al. (Guo et al., 2021b) was used for region associated DEG analysis. Up- and down-regulated DEGs were input as DEGs lists; bed file of peak regions was input as Genomic Regions of Interest (gROI) file. Different types of NANOG and KLF5 ChIP-seq peak regions were generated by bedtools v2.29.0 (Quinlan and Hall, 2010) multiinter function. For submit options, ''GRCh38'' was selected as reference genome, ''1000, 500, 200, 100, 50, 25, 20, 15, 10, 5, 0 kb'' as the customized peak extend distance, and ''hypergeometric test'' as the statistical test.

sgRNA target sites prediction

sgRNA targeting sites prediction for LTR7Y sgRNAs from this paper and LTR7Y/B sgRNAs from Pontis et al. was conducted with customized scripts. The predicted targets with less than 3 mismatches were annotated using gene and TE annotation and then categorized into either promoter, exonic, TE, intronic, or intergenic sites (annotation with multiple categories would be retained with this order).