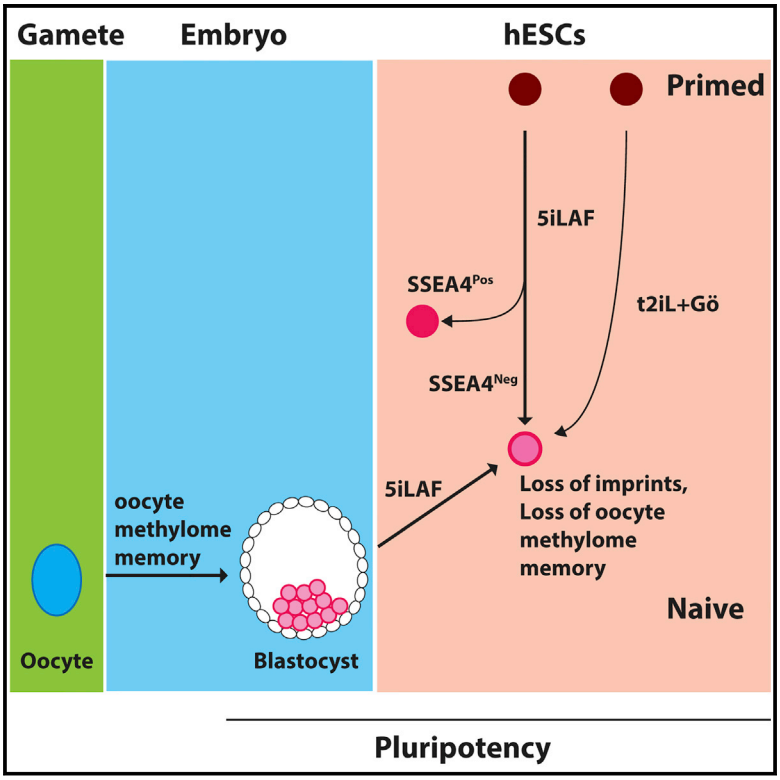
Brief Report

Cell Stem Cell

Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory

Graphical Abstract



Authors

William A. Pastor, Di Chen, Wanlu Liu, ..., Kathrin Plath, Steven E. Jacobsen, Amander T. Clark

Correspondence

jacobsen@ucla.edu (S.E.J.), clarka@ucla.edu (A.T.C.)

In Brief

Pastor and colleagues show that reversion of primed hESCs in 5iLAF, or derivation of hESCs in 5iLAF, results in a population of naive cells characterized by loss of the marker SSEA4. However, these cells have a methylation pattern with little resemblance to blastocyst and near total loss of imprinting.

Highlights

- Reversion or derivation of hESCs in 5iLAF results in SSEA4-negative cells
- SSEA4-negative hESCs show gene expression consistent with naive pluripotency
- Naive hESCs show lost "memory" of gamete and blastocyst methylation
- Imprinting is lost in naive hESCs

Accession Numbers

GSE76970

Pastor et al., 2016, Cell Stem Cell *18*, 323–329
March 3, 2016
http://dx.doi.org/10.1016/j.stem.2016.01.019

Cell Stem Cell
Brief Report

Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory

William A. Pastor,[1,5] Di Chen,[1,5] Wanlu Liu,[1,5] Rachel Kim,[3] Anna Sahakyan,[2] Anastasia Lukianchikov,[1] Kathrin Plath,[2,3] Steven E. Jacobsen,[1,2,3,4,*] and Amander T. Clark[1,3,*]
[1]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA
[3]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, CA 90095, USA
[4]Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA
[5]Co-first author
*Correspondence: jacobsen@ucla.edu (S.E.J.), clarka@ucla.edu (A.T.C.)
http://dx.doi.org/10.1016/j.stem.2016.01.019

SUMMARY

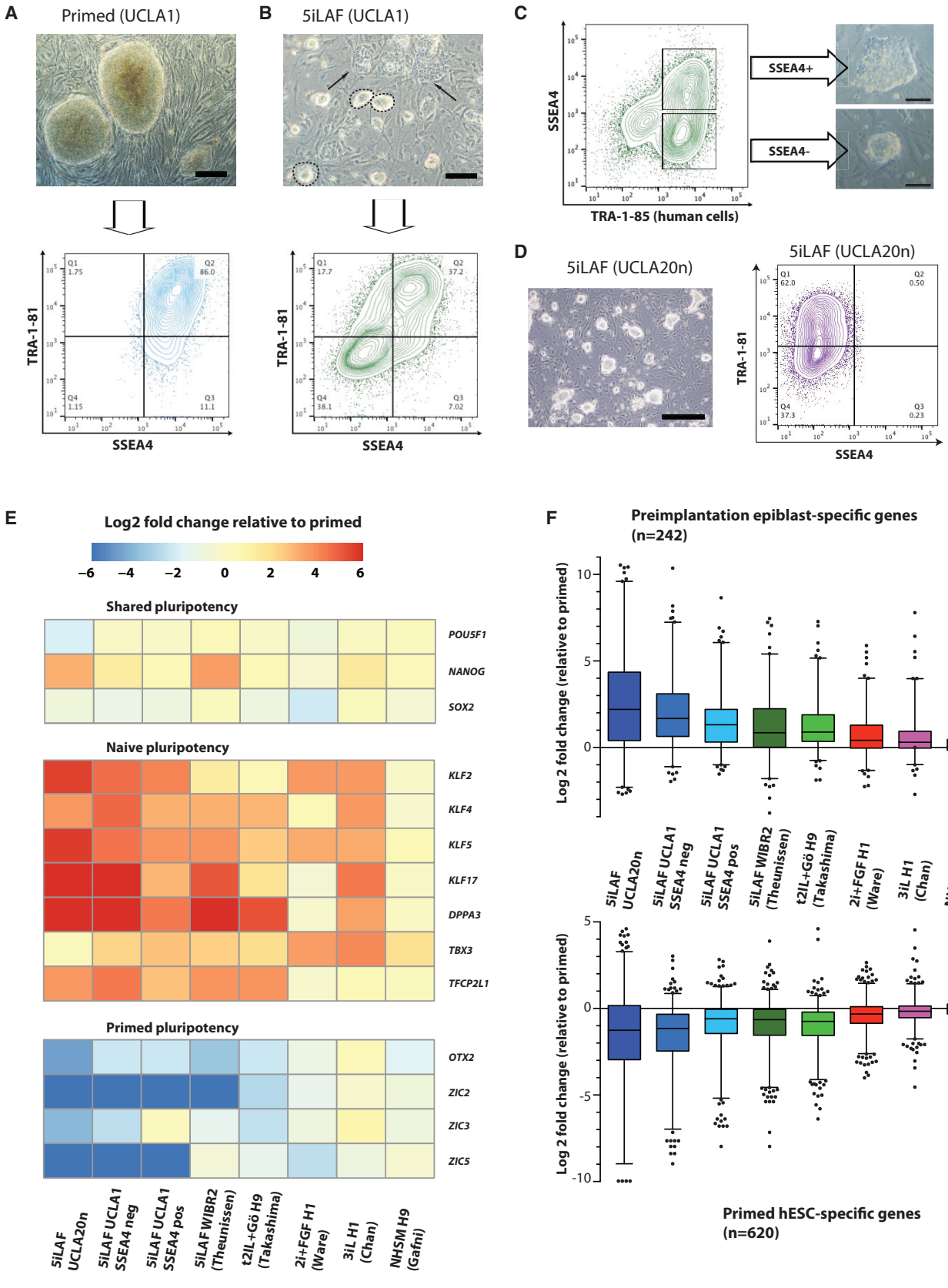
Human embryonic stem cells (hESCs) typically exhibit "primed" pluripotency, analogous to stem cells derived from the mouse post-implantation epiblast. This has led to a search for growth conditions that support self-renewal of hESCs akin to hypomethylated naive epiblast cells in human pre-implantation embryos. We have discovered that reverting primed hESCs to a hypomethylated naive state or deriving a new hESC line under naive conditions results in the establishment of Stage Specific Embryonic Antigen 4 (SSEA4)-negative hESC lines with a transcriptional program resembling the human pre-implantation epiblast. In contrast, we discovered that the methylome of naive hESCs in vitro is distinct from that of the human epiblast in vivo with loss of DNA methylation at primary imprints and a lost "memory" of the methylation state of the human oocyte. This failure to recover the naive epiblast methylation landscape appears to be a consistent feature of self-renewing hypomethylated naive hESCs in vitro.

Human embryonic stem cells (hESCs) are in vitro pluripotent cell types with the capacity for unlimited self-renewal and differentiation, making them critical models for understanding mechanisms required for human embryo development and differentiation. Although hESCs are derived from pre-implantation human blastocysts, they are morphologically and transcriptionally similar to murine epiblast stem cells (EpiSCs), which are derived from post-implantation mouse embryos. As such, hESCs and EpiSCs are said to exhibit a "primed pluripotent state" while mouse ESCs derived from the pre-implantation blastocyst exhibit a "naive pluripotent state" corresponding to an earlier stage of development (Nichols and Smith, 2009).

A number of culture conditions have recently been developed that promote maintenance and self-renewal of naive human pluripotent stem cells (Chan et al., 2013; Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014). Each protocol generates cell types with slightly different molecular characteristics, which may reflect metastable states in the spectrum of naive to primed pluripotency. A recent meta-analysis of sequencing data indicates that two of these protocols generate cells with a close transcriptional resemblance to the human pre-implantation epiblast (Huang et al., 2014). In the first protocol, hESCs are transfected with *KLF2* and *NANOG* and are cultured in media with titrated two inhibitors plus leukemia inhibitory factor and Gö6983 (t2iL+Gö) (Takashima et al., 2014). In the second protocol, primed cells can be reverted by being transferred to a media containing a cocktail of five inhibitors plus LIF, Activin, and/or Fibroblast Growth Factor 2 (5iLAF) (Theunissen et al., 2014). Using t2iL+Gö reversion of the H9 primed hESC line, it was shown that DNA methylation is globally reduced to the average level measured in human pre-implantation epiblasts (Takashima et al., 2014), with additional locus-specific erosion in the 5′ region of the LINE1 human specific (L1HS) retrotransposons (Gkountela et al., 2015). The DNA methylation profile of cells cultured in 5iLAF has never been evaluated.

Before studying the methylation pattern of 5iLAF cultured cells, we first wanted to confirm and characterize the naive phenotype. We performed n = 4 independent reversions of the hESC line UCLA1 (Diaz Perez et al., 2012) using 5iLAF (Theunissen et al., 2014). Upon the reversions we observed a mixture of small, round colonies similar to naive mESCs as well as flat, cobblestone-like colonies (Figures 1A and 1B). We evaluated one reversion using two classic human pluripotency surface markers called SSEA4 and TRA-1-81. Unlike primed UCLA1 hESCs, which are double positive for SSEA4 and TRA-1-81, the 5iLAF-reverted hESCs have a large fraction of double-negative cells (Figures 1A and 1B). Immunofluorescence staining showed that the SSEA4- and TRA-1-81-negative cells were still positive for OCT4 and NANOG (Figures S1A–S1F).

Next, we sorted the 5iLAF-cultured cells into SSEA4-positive and -negative populations using fluorescence-activated cell sorting (FACS) and re-plated the sorted cells onto MEFs in 5iLAF media (Figure 1C). We discovered that SSEA4-positive cells yielded mostly flat colonies, whereas SSEA4-negative cells



(legend on next page)

yielded mostly round colonies. One passage after sorting, the SSEA4-negative population remained SSEA4 negative, indicating that this is a relatively stable state (Figure S1G). We then reverted two additional lines called UCLA4 and UCLA5 (Diaz Perez et al., 2012) and found that small, round colony morphology was always enriched in the SSEA4-negative fraction whereas the SSEA4-positive cells yielded mostly flat, cobblestone colonies (Figures S1H–S1K).

In order to determine whether the heterogeneity in SSEA4 expression was also observed when deriving hESC lines completely under naive 5iLAF conditions, we derived n = 2 new hESC lines, which we have called UCLA19n and UCLA20n after thawing n = 7 day 5 vitrified human blastocysts. Colonies were uniformly round and flow cytometry revealed that UCLA19n was 85% SSEA4 negative (results not shown), whereas UCLA20n was almost completely SSEA4 negative (Figure 1D). In contrast, TRA-1-81 was expressed on a significant portion of SSEA4-negative cells in UCLA20n as well as reverted UCLA4 and UCLA5 hESC lines (Figure 1D, Figure S1I, Figure S1K). Therefore, loss of TRA-1-81 is not a consistent marker of naive morphology, whereas absence of SSEA4 is a highly correlated feature of naive round colony morphology. In summary, reversion of primed hESCs in 5iLAF generates a heterogeneous mixture of colonies, with SSEA4-negative hESCs correlating with small round colony morphology similar to naive hESCs derived from the human pre-implantation blastocyst.

On the basis of morphology, we speculated that 5iLAF SSEA4-negative hESCs are the naive population and thus transcriptionally resemble the cells of the human pre-implantation epiblast. To address this, we performed RNA-seq of 5iLAF-cultured SSEA4-positive or SSEA4-negative fractions of UCLA1, and we compared them to SSEA4-positive primed UCLA1 hESCs at equivalent passages (Table S1). We also performed RNA-seq of UCLA20n at passage 20 after derivation. We did not analyze UCLA19n as it was found to be 70% polyploid by passage 15. Consistent with the expression patterns of genes associated with naive pluripotency in mice, the 5iLAF SSEA4-negative cells and UCLA20n had elevated levels of *NANOG* as well as a dramatic upregulation of *KRUPPLE-LIKE FACTOR* (*KLF*) family transcription factors and reduced expression of primed state master regulators such as *ZINC FINGER OF THE CEREBELLUM* (*ZIC*) family transcription factors and *OTX2* (Buecker et al., 2014; Tang et al., 2011; Yang et al., 2014) (Figure 1E, Table S1). To further confirm the similarity of 5iLAF SSEA4-negative and UCLA20n hESCs to the human pre-implantation blastocyst, we used the previously published single-cell expression data from late pre-implantation epiblast and primed hESCs (Yan et al., 2013). We defined a set of "pre-implantation epiblast-specific" and "primed-specific" genes, which showed >4-fold difference in expression between these two cell types (Table S1). Using these genes as a reference, we found that the 5iLAF SSEA4-negative hESCs and UCLA20n had global upregulation of naive epiblast-specific genes and downregulation of primed-specific genes (Figure 1F).

In contrast, the SSEA4-positive hESCs sorted from 5iLAF cultures had an intermediate expression pattern between primed and naive, suggesting that SSEA4-positive cells that stably self-renew in 5iLAF are partially reverted to the naive state (Figures 1E and 1F). Comparing to published datasets, we found that the SSEA4-negative population in UCLA1 and the new hESC line UCLA20n is analogous to the original 5iLAF hESC lines created by reverting WIBR2 (Theunissen et al., 2014) and to t2iL+Gö-cultured hESCs created by reverting H9 (Takashima et al., 2014). In contrast, other published naive methods showed a less pronounced shift toward the naive state and failure to repress primed markers (Chan et al., 2013; Gafni et al., 2013; Ware et al., 2014). Interestingly, lines generated by these methods are also reported to be SSEA4 positive. Given these results, we focused our methylation analysis on the 5iLAF and t2iL+Gö conditions.

To determine the methylation pattern of hESCs in 5iLAF, we performed whole-genome bisulfite sequencing (WGBS) on two to four independent sorts of SSEA4-negative or SSEA4-positive reverted UCLA1 cells, SSEA4-negative UCLA20n cells, and primed UCLA1 cells that had been in culture a similar length of time to the reverted lines. We discovered that, similar to the levels observed in t2iL+Gö (Gkountela et al., 2015; Takashima et al., 2014), 5iLAF-cultured SSEA4-negative hESCs and UCLA20n had an average CG methylation level that resembled that of the human blastocyst (Figure 2A) (Okae et al., 2014).

In mammals the methylation pattern of the blastocyst is shaped by events during gametogenesis and early embryogenesis. The male pronucleus is selectively demethylated in early embryonic development, with only a few regions such as paternally methylated imprinted loci protected from DNA demethylation (Okae et al., 2014; Smith et al., 2012, 2014). Thus in humans the methylation pattern of the blastocyst strongly resembles that of the oocyte (Figures 2B and 2C). In contrast, the

Figure 1. 5iLAF SSEA4-Negative Subpopulation Recapitulates Naive Expression Pattern

(A) Upper: brightfield image of primed UCLA1 hESCs. Lower: flow cytometry plot of primed UCLA1 hESCs stained for SSEA4 and TRA-1-81. Scale bar indicates 200 μm.

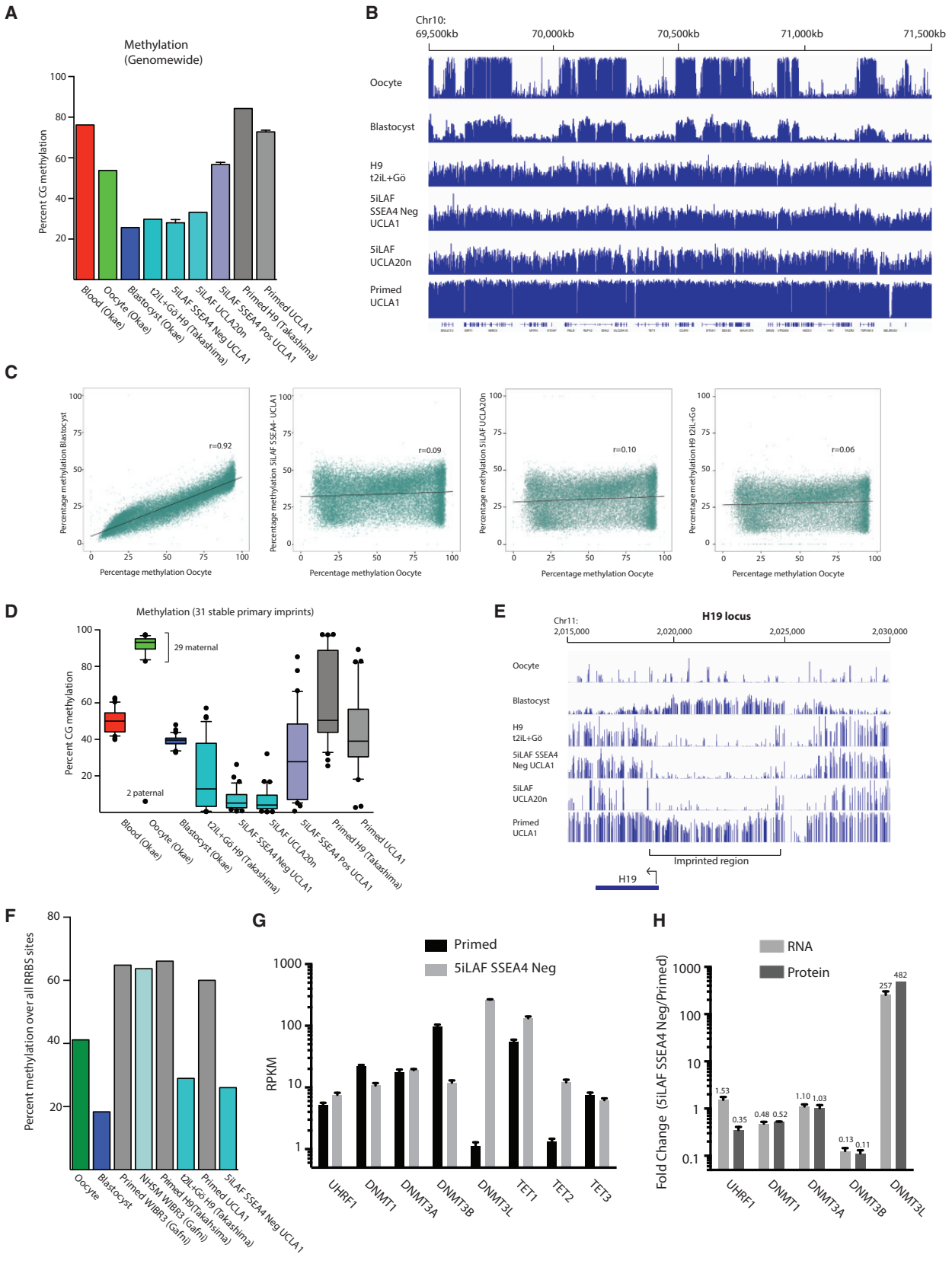
(B) Upper: UCLA1 hESCs reverted in 5iLAF. A mixture of round (indicated with dotted circle) and flat colonies (indicated with arrow) are observed. Lower: flow cytometry plot of 5iLAF cultured UCLA1 hESCs stained for SSEA4 and TRA-1-81. Scale bar indicates 200 μm.

(C) 5iLAF cells were sorted into SSEA4-positive and SSEA4-negative populations. Upon re-plating, the SSEA4-positive cells formed flat colonies and the SSEA4-negative cells formed round colonies (n = 2 biological replicates). Scale bar indicates 100 μm.

(D) hESC line UCLA20n, derived from a 5-day human blastocyst in 5iLAF. Left: brightfield image. Scale bar indicates 200 μm. Right: flow cytometry plot of TRA-1-85^{+} (human) UCLA20n hESCs stained for SSEA4 and TRA-1-81.

(E) Expression of genes identified by others as associating with the naive and primed states in mice. Expression level is determined by RNA-seq. For 5iLAF SSEA4-negative (neg) and primed hESCs, n = 4. For 5iLAF SSEA4-positive (pos), n = 2. Other data comes from published RNA-seq or microarray datasets. Methodology, cell type, and citation are indicated.

(F) A set of "pre-implantation epiblast" and "primed" specific genes were defined based on published data. Expression of these genes is shown for various methodologies, relative to primed controls from the same dataset. UCLA20n was normalized to a primed UCLA1 library generated and sequenced at the same time.



(legend on next page)

methylation landscapes of SSEA4-negative 5iLAF-cultured hESCs, UCLA20n, and t2iL+Gö-cultured cells are only weakly correlated with the human blastocyst and human oocyte (Figures 2B and 2C). Naive cells, even if cultured by different methodologies or derived directly from the human blastocyst, converge toward a methylation pattern that is different from that of the pre-implantation human blastocyst (Figures 2B and 2C). A striking example of this trend is observed at 332 CpG islands identified previously as ''transient maternal imprints:'' sites that are highly methylated in oocytes and the maternal chromosomes of blastocyst that lose methylation upon implantation (Smith et al., 2014). We discovered that reversion does not regenerate methylation at these sites, nor is methylation retained at these transient maternal imprints in the UCLA20n hESC line (Figure S2A).

An additional, striking deviation from oocyte and blastocyst methylation patterns in 5iLAF and t2iL+Go cultured cells occurred at stable imprints. These are regions where DNA methylation is established exclusively during germ-cell development. These methylated sites are protected from DNA demethylation during pre-implantation embryo development, remaining differentially methylated in somatic cells through the life of the organism and promoting a parent-of-origin-specific expression pattern in the neighboring genes. We examined DNA methylation at 29 maternally methylated stable primary imprints and 2 paternally methylated stable primary imprints (Okae et al., 2014) (Figure 2D, Table S2). There is roughly 50% methylation in somatic tissue and slightly below 50% methylation in blastocysts, as expected. In the primed UCLA1 hESCs used in our study, the median methylation of these imprinted sites was close to 50%, though some imprints were hyper- or hypomethylated, similar to what has been observed previously for other hESC lines (Rugg-Gunn et al., 2007). Strikingly, the 5iLAF SSEA4-negative hESCs and UCLA20n had near complete loss of methylation from all 31 primary stable imprints evaluated in our study, with loss over many imprints also found in t2iL+Gö (Figures 2D and 2E). Taking advantage of single nucleotide polymorphisms (SNPs) present in the UCLA1 hESC line, we observed a shift upon reversion in 5iLAF from monoallelic to biallelic expression of several imprinted genes including *H19* and *SNRPN*. (Figure S2B, Table S2). In order to determine whether methylation could be restored at imprinted genes by reverting the naive hESCs back to a primed state, we cultured 5iLAF SSEA4-negative and primed UCLA1 cells in primed epiblast-like cell (EpiLC) media (Hayashi et al., 2011) for 16 days. During this time, we discovered that 5iLAF SSEA4-negative cells showed a global shift toward expression of primed-specific genes and gained DNA methylation genome-wide (Figure S2C). However, increased methylation over imprinted regions was very modest, and biallelic expression was still observed (Figure S2D, Table S2). Thus, when lost, imprinting is not re-established in cells cultured in primed conditions, a similar scenario to the rescue of global DNA methylation, but not imprint methylation, in Dnmt1 knockout ESCs by the re-expression of Dnmt1 (Holm et al., 2005). Furthermore, consistent with data observed in hESCs cultured in t2iL+Gö (Gkountela et al., 2015), young LINE elements also show dramatic promoter hypomethylation in 5iLAF (Figure S2E).

Given the problem with maintenance of imprint methylation in naive cells, we considered the possibility that the 5iLAF SSEA4-positive cells may represent a useful intermediate. However, we discovered that the SSEA4-positive cells showed intermediate levels of global and imprint methylation loss (Figures 2C and 2D), with biallelic expression of *SNRPN* and *H19* (Table S2). We also analyzed the methylation loss imparted by naive human stem cell media (NHSM) (Gafni et al., 2013), which shows the smallest transcriptional shift toward naive pluripotency (Figures 1E and 1F). In order to directly compare our methylome data to findings of Gafni et al. (2013), we modified all whole-genome datasets to simulate the Reduced Representation Bisulfite Sequencing (RRBS) approach used by Gafni to measure DNA methylation. We discovered that imprint methylation was unperturbed in NHSM (Figure S2F), but very little global change in methylation was observed either (Figure 2F).

Consistent with an initial report of karyotypic instability in 5iLAF culture (Theunissen et al., 2014), we discovered that, 24 passages after reversion, the 5iLAF UCLA1 hESCs developed widespread karyotypic abnormalities, which was not observed in the first 13 passages following reversion (Figure S2G). Similarly, UCLA20n had evidence of trisomies at chromosomes 3, 7, 12, and 20 by passage 14 and as discussed above, UCLA19n was 70% polyploid at passage 15 (Figures S2G and S2H). Therefore, karyotypic instability may also be a frequent consequence of naive hESC culture.

To determine the cause of the cells' failure to maintain DNA methylation, we analyzed changes in RNA and protein levels of DNA methylation and demethylation machinery. We found that the RNA and protein levels of the de novo DNA methyltransferase DNMT3B dropped sharply in the 5iLAF SSEA4-negative cells,

Figure 2. Naive hESCs Fail to Recapitulate Naive-Specific Methylation Pattern

(A) Average genomewide-CG methylation level in primed and 5iLAF UCLA1 hESCs, shown in comparison with published datasets. For 5iLAF SSEA4-negative (neg) and primed hESCs, n = 3. For 5iLAF SSEA4-positive (pos), n = 2.

(B) DNA methylation is shown for a region of chromosome 10. Each bar indicates a single CG, and the height of the bar indicates the percentage of CG methylation. Where multiple CGs are too close to be visually rendered separately, an average value is shown.

(C) Correlation plots relative to human oocyte using 100 kb genome bins.

(D) DNA methylation over stable primary imprints. The average methylation level of each imprint in a given sample is represented as one point in the box and whisker point.

(E) DNA methylation over the paternally imprinted *H19* locus. Each bar indicates a single CG, and the height of the bar indicates the fraction of CG methylation. Where multiple CGs are too close to be visually rendered separately, an average value is shown.

(F) Total DNA methylation for three competing approaches for culturing naive cells. Because the Gafni et al. (2013) data was generated by RRBS, only CGs that had coverage in that dataset are included in this analysis to make the data comparable.

(G) Expression (RPKM) of DNA methyltransferases, DNMT cofactors, and Tet-family oxidases as measured by RNA-seq (n = 4).

(H) RNA and protein levels of DNA methyltransferases in 5iLAF SSEA4-neg UCLA1 hESCs relative to primed. RNA level is determined from RNA-seq data (n = 4) and protein level from quantitative westerns (UHRF1, n = 6 western blots; DNMT1, DNMT3A, and DNMT3B, n = 2; DNMT3L, n = 1).

while DNMT3A was unchanged and DNMT3L increased dramatically relative to primed hESCs. *UHRF1* RNA levels were slightly elevated in naive hESCs. However, at the protein level, we observed a 65% loss of UHRF1, and both *DNMT1* RNA and protein levels were reduced by 50% in the naive state. Furthermore, expression of the 5mC oxidases TET1 and TET2 increased substantially in the naive state (Figures 2G and 2H).

In the current study we have shown that naive hESCs have a transcriptional program enriched in human pre-implantation-specific genes but with a global DNA methylation landscape that is distinct from the normal state of the human pre-implantation blastocyst. The negative effect of the loss of ''transient imprints'' and the failure to recapitulate the oocyte-like methylation pattern is unclear. However, the loss of stable primary imprints is potentially serious in human pluripotent stem cell research. Correct imprinting is necessary for organism survival, and a number of rare human medical disorders have been linked to aberrant imprinting (Butler, 2009). Of note, murine embryonic germ cell lines are transcriptionally similar to murine ESCs but have widespread loss of imprints and contribute poorly to chimeras (Leitch et al., 2013; Oliveros-Etter et al., 2015; Tada et al., 1998), demonstrating the importance of imprints in correct differentiation of pluripotent cells in vivo. We also observed, in accompaniment to the loss of imprints, extensive karyotypic abnormalities in cells after their prolonged culture in 5iLAF. Loss of DNA methylation has been linked to karyotypic instability (Haaf, 1995).

We note that methylation at the imprinted loci is clearly depressed relative to that of surrounding regions. This may reflect the observation that many imprinted loci are promoters or regulatory elements that are active in the blastocyst (Rugg-Gunn et al., 2007). Thus if methylation is partially eroded at the imprint, the relevant transcription factors bind and cause further demethylation (as is generally the case at these genetic elements). In other words, methylation may be a very weak barrier to locus activation in 5iLAF. Similar dynamics may be at work at L1HS elements.

Although we observed a reduction in DNMT3B protein in the naive cells, we propose that this has only modest effects on creating the 5iLAF methylome given that *DNMT3A*$^{-/-}$ *DNMT3B*$^{-/-}$ DKO primed hESCs maintain primary imprints and show only modest DNA demethylation even after extended culture (Liao et al., 2015). We therefore propose that a combination of impaired maintenance methylation and increased TET activity could explain the majority of the 5iLAF hypomethylation phenotype. In a cell type with impaired maintenance and some continuous de novo methylation (imparted by DNMT3A and the remaining DNMT3B), DNA methylation levels will reach a steady state, but memory of previous methylation will be lost with DNA replication.

ACCESSION NUMBERS

The RNA-seq and BS-seq data in this paper are available under the GEO accession number GEO: GSE76970.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes two figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.01.019.

AUTHOR CONTRIBUTIONS

The reversion and culture of hESCs in naive conditions was conducted by R.K., D.C., and A.S. hESC derivation from human blastocyst was undertaken by R.K. Experiments and data interpretation were conducted by D.C., W.A.P., A.L., and R.K. Computation analysis was conducted by W.L. and W.A.P. Conceiving and directing research were conducted by K.P., S.E.J., and A.T.C. Maintenance of University Compliance, including ESCRO, IRB, and Biological Safety, was overseen by A.T.C. The manuscript was written by W.A.P. and A.T.C.

ACKNOWLEDGMENTS

We thank the UCLA Broad Stem Cell Research Center (BSCRC) Flow Cytometry core for flow and FACS assistance, the UCLA BSCRC High Throughput Sequencing Core, Sriharsa Pradhan from N.E.B. for donating anti-DNMT1 antibody, and Steven Peckman from the UCLA BSCRC for consenting couples for embryo donation for hESC derivation. We thank Colin Shew, Beatrice Sun, and Tiasha Shafiq for help with experiments and the Fall 2015 UCLA Biomedical Research Minor 5HB undergraduate class for useful discussions. W.A.P. was supported by the Jane Coffin Childs Memorial Fund for Medical Research and a UCLA BSCRC Postdoctoral Training Fellowship. D.C. is supported by a UCLA BSCRC Postdoctoral Training Fellowship. W.L. is supported by the Philip J. Whitcome Fellowship from the UCLA Molecular Biology Institute and a scholarship from the Chinese Scholarship Council. This work was supported by the NIH R01 HD079546 (A.T.C.), CIRM RB4-06133 (K.P.), and P01 GM099134 (K.P.). Funds for human embryo banking and derivation of new hESC lines were provided by the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research. No federal grant funding was used for work with human embryo's or derivation of new hESC lines. No payment was provided to embryo donors for their generous gift of surplus embryos to stem cell research. S.E.J. is an investigator of the Howard Hughes Medical Institute.

Received: September 16, 2015
Revised: October 2, 2015
Accepted: January 15, 2016
Published: February 4, 2016

REFERENCES

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

Butler, M.G. (2009). Genomic imprinting disorders in humans: a mini-review. J. Assist. Reprod. Genet. *26*, 477–486.

Chan, Y.S., Göke, J., Ng, J.H., Lu, X., Gonzales, K.A., 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.

Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., and Clark, A.T. (2012). Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. Hum. Mol. Genet. *21*, 751–764.

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.

Gkountela, S., Zhang, K.X., Shafiq, T.A., Liao, W.W., Hargan-Calvopiña, J., Chen, P.Y., and Clark, A.T. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. Cell *161*, 1425–1436.

Haaf, T. (1995). The effects of 5-azacytidine and 5-azadeoxycytidine on chromosome structure and function: implications for methylation-associated cellular processes. Pharmacol. Ther. *65*, 19–46.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519–532.

Holm, T.M., Jackson-Grusby, L., Brambrink, T., Yamada, Y., Rideout, W.M., 3rd, and Jaenisch, R. (2005). Global loss of imprinting leads to widespread tumorigenesis in adult mice. Cancer Cell *8*, 275–285.

Huang, K., Maruyama, T., and Fan, G. (2014). The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. Cell Stem Cell *15*, 410–415.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., et al. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat. Genet. *47*, 469–478.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., and Arima, T. (2014). Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet. *10*, e1004868.

Oliveros-Etter, M., Li, Z., Nee, K., Hosohama, L., Hargan-Calvopina, J., Lee, S.A., Joti, P., Yu, J., and Clark, A.T. (2015). PGC Reversion to Pluripotency Involves Erasure of DNA Methylation from Imprinting Control Centers followed by Locus-Specific Re-methylation. Stem Cell Reports *5*, 337–349.

Rugg-Gunn, P.J., Ferguson-Smith, A.C., and Pedersen, R.A. (2007). Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. Hum. Mol. Genet. *16* (Spec No. 2), R243–R251.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611–615.

Tada, T., Tada, M., Hilton, K., Barton, S.C., Sado, T., Takagi, N., and Surani, M.A. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. Dev. Genes Evol. *207*, 551–561.

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.

Tang, F., Barbacioru, C., Nordman, E., Bao, S., Lee, C., Wang, X., Tuch, B.B., Heard, E., Lao, K., and Surani, M.A. (2011). Deterministic and stochastic allele specific gene expression in single mouse blastomeres. PLoS ONE *6*, e21208.

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*, 471–487.

Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., et al. (2014). Derivation of naive human embryonic stem cells. Proc. Natl. Acad. Sci. USA *111*, 4484–4489.

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, S.H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A.D. (2014). Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. Cell Rep. *7*, 1968–1981.

Cell Stem Cell, Volume *18*

Supplemental Information

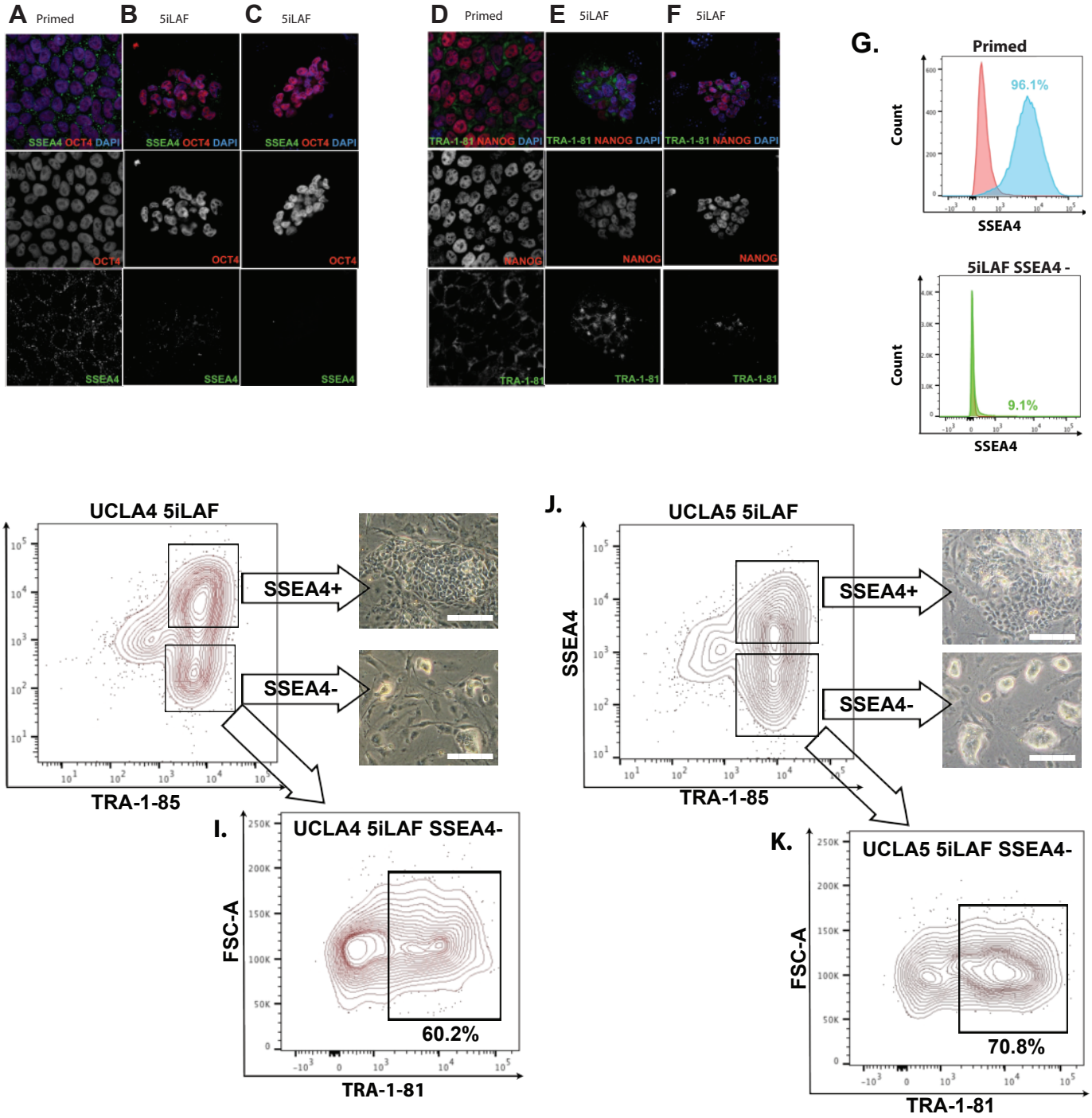
Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory

William A. Pastor, Di Chen, Wanlu Liu, Rachel Kim, Anna Sahakyan, Anastasia Lukianchikov, Kathrin Plath, Steven E. Jacobsen, and Amander T. Clark

Supplementary materials inventory

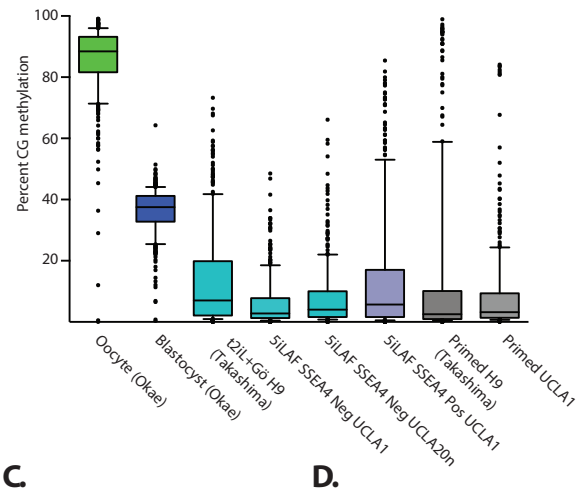
I. Supplementary Figures

Supplementary Figure 1

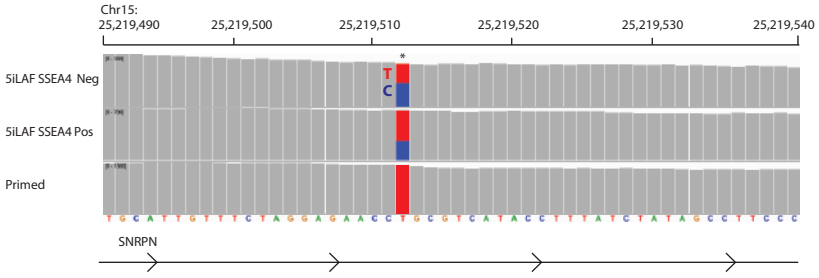


Supplementary Figure 2

A.



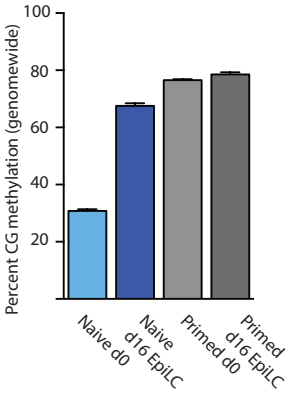
B.



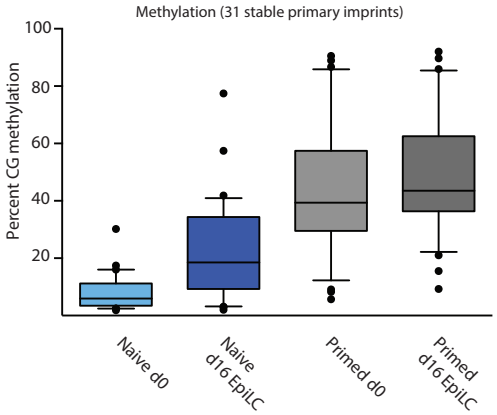
SNRPN (base 25,219,512)

	Allele	Reads
5iLAF SSEA4 Neg	**T**	**132**
	C	158
5iLAF SSEA4 Pos	**T**	**407**
	C	259
Primed	**T**	**1435**
	C	4

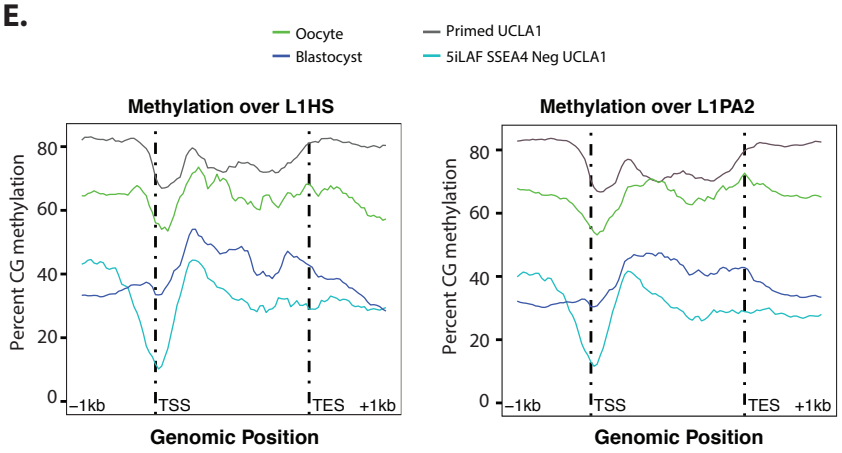
C.



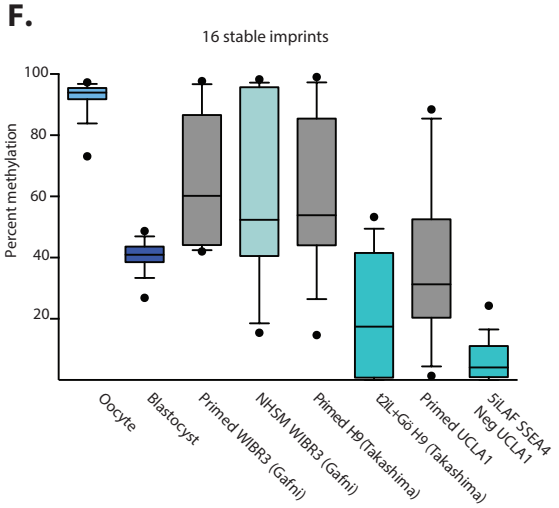
D.



E.



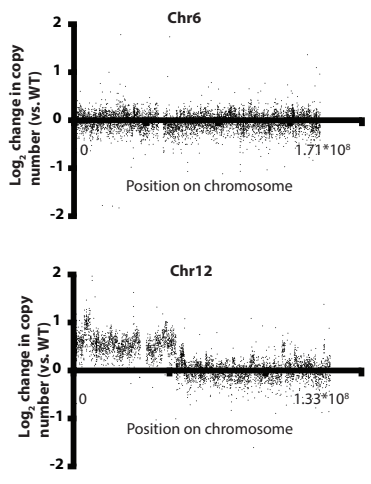
F.



G.

	Origin	Karyotype
UCLA1 5iLAF p13	Derived under primed conditions and reverted to naive in 5iLAF. Grown in 5iLAF 13 passages	G-band karyotyping. 18/20 cells 46,XX. 2/20 non-clonal abberations.
UCLA1 5iLAF p24	Derived under primed conditions and reverted to naive in 5iLAF. Grown in 5iLAF 24 passages	G-band karyotyping. 18/20 cells abnormal female karyotype with 13/20 = 46, XX,add(8)(q24.1). 30% polyploid based on n=70 metaphases
UCLA19n	Derived from blastocyst in 5iLAF. Grown in 5iLAF 15 passages.	G-band karyotyping. 70% polyploid.
UCLA20n	Derived from blastocyst in 5iLAF. Grown in 5iLAF 14 passages.	Array comparative genomic hybridization Trisomy over regions of Chr3, Chr7, 12 and 20.

H.



II. Supplementary Figure captions

Supplementary Figure 1 (related to Figure 1): Properties of 5iLAF SSEA4 negative and SSEA4 positive cells generated by reversion of primed hESCs. **A-C)** Immunofluorescence for SSEA4 and OCT4. Note that all colonies are OCT4 positive. **A)** A colony of SSEA4 positive primed UCLA1 hESCs. **B)** A colony of 5iLAF SSEA4 positive UCLA1 hESCs. **C)** A colony of 5iLAF SSEA4 negative UCLA1 hESCs. **D-F)** Immunofluorescence for TRA-1-81 and NANOG. Note that all populations are NANOG positive. **D)** A colony of TRA-1-81 positive UCLA1 hESCs. **E)** A colony of 5iLAF TRA-1-81 positive UCLA1 hESCs. **F)** A colony of 5iLAF TRA-1-81 negative UCLA1 hESCs. **G)** Flow cytometry of control primed and re-plated 5iLAF SSEA4 negative UCLA1 hESCs grown for one passage. Fluorescence of unstained cells is indicated in red and percentage showing positive staining is indicated. After re-plating, the vast majority of sorted SSEA4 negative cells remain SSEA4 negative. **H-K)** Similar to UCLA1, SSEA4 positive 5iLAF cells from UCLA4 (**H**) and UCLA5 (**J**) yield flat colonies upon re-plating while the SSEA4 positive 5iLAF cells yield round colonies. **I,K)** Unlike UCLA1, most cells in the SSEA4 negative subpopulation have high TRA-1-81 expression (compare to Figure 1B).

Supplementary Figure 2 (related to Figure 2): A) DNA methylation over transiently imprinted CG islands. The average methylation level of each imprint in a given sample is represented as one point in the box and whisker point. **B)** Reads mapped over an annotated SNP in the maternally imprinted SNRPN locus. Reads over each base are plotted, and the SNP sequence is indicated by color. Only one allele is expressed in the parent primed UCLA1 line, but both alleles are expressed in 5iLAF cells. **C)** Global DNA methylation in naïve and primed cells before and after sixteen days of culture in EpiLC-like conditions to restore the primed state. **D)** DNA methylation over imprints in naïve and primed cells before and after sixteen days of culture in EpiLC-like conditions to restore the primed state. Each imprint is represented as a single point in the box plot. Note the modest increase in methylation at imprints as the naïve cells are converted to primed conditions, whereas the global increase in methylation is much greater. **E)** Hypomethylation of young LINE elements including L1 human specific (L1HS) and its descendent L1PA2 in 5iLAF SSEA4 negative UCLA1 hESCs, as shown by metaplot. Note the dramatic loss of methylation in the vicinity of the element promoter. **F)** DNA methylation over imprints for three alternate approaches for culturing naïve cells. Because the Gafni 2013 data was generated by RRBS, only CGs that had coverage in the Gafni 2013 dataset are included in this analysis to make the data comparable. Only sixteen stable imprints had sufficient coverage for robust analysis. **G)** Karyotyping results from reverted UCLA1 lines and new lines derived from blastocyst in 5iLAF. **H)** Comparative Genomic Hybridization (CGH) data is shown over two chromosomes for the UCLA20n line cultured in 5iLAF. Most chromosomes showed normal karyotype (e.g chromosome 6, left), but several showed regions of elevated DNA content consistent with aneuploidy (e.g. chromosome 12, right).

III. Supplementary Table Captions

Supplementary Table 1 (related to Figure 1): Description of all cell populations used for RNA-seq library generation (panel 1). RPKM values for all genes in these RNA-seq samples (panel 2). A the list of of “Epiblast-specific” and “hESC specific” genes described in Figure 1F (panel 3).

Supplementary Table 2 (related to Figure 2): Description of all cell populations used for WGBS library generation (panel 1). Methylation level of samples over 332 transient imprints (defined using data taken from (Smith et al., 2014)) (panel 2) and 31 primary stable imprints (identified in (Okae et al., 2014)) (panel 3). Also shown is the number of reads corresponding to SNPs found in imprinted genes identified as reactivated in the 5iLAF SSEA4$^-$ UCLA1 hESCs (panel 4). The imprinted allele is indicated below the expressed allele, except for *PEG3* where the imprinted allele could not be identified because there is no expression in the primed cells. Major SNPs are those with >1% abundance in the human population according to dbSNP build 142.

IV. Supplementary Experimental Protocols

Cell culture

Reversion and culture of cells was adapted from the published 5iLAF protocol (Theunissen et al., 2014). UCLA-derived human embryonic stem cell lines were routinely maintained in DMEM/F-12 (Life Technologies), 20% KSR (Life Technologies), 10ng/mL bFGF (Peprotech), 1% nonessential amino acids (Life Technologies), 2mM GlutaMAX (Life Technologies), penicillin-streptomycin (Life Technologies) and 0.1mM beta-mercaptoethanol (Sigma) and passaged with 1mg/mL collagenase type IV (Life Technologies). During maintenance, they were passaged once every seven days. To achieve reversion, two days post passage, medium was changed to DMEM/F-12, 15% FBS (Omega Scientific), 5% KSR (Life Technologies), 4ng/mL bFGF (Peprotech), 1% nonessential amino acids (Life Technologies), 1mM GlutaMAX (Life Technologies), penicillin-streptomycin (Life Technologies) and 0.1mM beta-mercaptoethanol (Life Technologies). On day 7 post-passage, cells were washed once with 1x dPBS (Life Technologies) and treated for 3 min. with 0.05% trypsin-EDTA (Life Technologies). Cells were dissociated into a single cell suspension, passed through a 40µm cell strainer and plated at a density of $2x10^5$ cells per 9.5 cm^2 in the 15% FBS containing medium with the addition of 10µM Y-27632 (Stemgent). Subsequent media changes were in the absence of Y-27632. Two days post plating, medium was changed to 5iLAF with daily changes thereafter. 5iLAF medium contained a 50:50 mixture of DMEM/F-12 (Life Technologies) and Neurobasal (Life Technologies), with 1x N2 supplement (Life Technologies), 1x B27 supplement (Life Technologies), 8ng/mL bFGF (Peprotech), 1% nonessential amino acids (Life Technologies), 1mM GlutaMAX (Life Technologies), penicillin-streptomycin (Life Technologies), 0.1mM beta-mercaptoethanol (Life Technologies), 50µg/mL BSA (Sigma), 1µM PD0325901 (Stemgent), 1µM IM-12 (Enzo), 0.5µM SB590885 (R&D Systems), 1µM WH-4-023 (A Chemtek), 10µM Y-27632 (Stemgent), 20ng/mL Activin A (Peprotech), 20ng/mL rhLIF (Millipore) and 0.5% KSR (Life Technologies). At about 11 to 12 days post plating, cells were passaged using a 3 min. treatment with StemPro Accutase (Life Technologies) and replated after passing through a 40µm cell strainer in 5iLAF medium. Round naïve colonies could be seen at this subsequent passage. Cultures were maintained in 5iLAF and passaged every 5-6 days using Accutase. All cultures were grown on a MEF layer seeded at a density of $1.5x10^6$ – $2.5x10^6$ cells per 6-well plate. Cells were cultured in ambient oxygen and 5% CO_2.

Analysis of surface markers

In flow cytometry or FACS experiments, cells were detached by Accutase and centrifuged and washed with 1xFACS buffer (1xPBS 1% BSA). Antibodies used for staining include: PE-conjugated TRA-1-85 (R&D systems, FAB3195P), APC-conjugated

anti-SSEA4 (R&D systems, FAB1435), and Alexa488-conjugated anti-TRA-1-81 (Cell Technologies, 60065). DAPI was added immediate prior to flow cytometry or FACS.

To generate stable SSEA4 negative or positive lines from 5iLAF cultures, live (DAPI negative) human (TRA-1-85^{+}) cells were sorted into SSEA4^{-} and SSEA4^{+} subpopulations. The sorted cells were then centrifuged at 1200RPM, re-suspended in 5iLAF media, and plated at a density of 300k/well of a 6-well plate.

To generate material for sequencing or Western blots, TRA-1-85^{+} SSEA4^{-} cells were sorted from the 5iLAF SSEA4^{-} culture, and TRA-1-85^{+} SSEA4^{+} cells were sorted from 5iLAF SSEA4^{+} or primed cultures. The cells were then centrifuged at 1200RPM five minutes and frozen, and DNA, RNA or protein extracted as described below.

Human hESC derivation in 5iLAF

Derivation of UCLA19n and UCLA20n were performed with vitrified day five human blastocysts under hypoxic conditions (5% O_2, 3%CO_2). A total of seven human blastocysts were used for these experiments. Blastocysts were received vitrified from the in vitro fertilization clinic following informed consent and thawed using Vit Kit-Thaw (Irvine Scientific) according to manufacturer protocol. The embryos were cultured in drops of Continuous Single Culture media (Irvine Scientific) supplemented with 20% Quinn's Advantage SPS Serum Protein Substitute (Sage Media) under mineral oil (Irvine Scientific) overnight at 37°C, 6% CO_2 and 5% O_2. The zona pellucida was removed using Tyrode's solution acidified (Irvine Scientific) before plating onto inactivated MEFs in 5iLAF media at passage (P) 0. Derivation success rate involved 5/7 blastocysts attaching to the MEFs at P0, and 2/7 giving rise to naïve (n) hESC lines capable of self-renewal for at least 15 passages. Accutase was used to harvest the P0 blastocyst outgrowths at day 6 (UCLA19n) and day 9 (UCLA20n). UCLA20n was supplemented from P0-P3 with a 50:50 mix of MEF conditioned media (20% knockout serum replacer and 4ng/ml FGF2) and 5iLAF to promote colony outgrowth. Starting at P4 UCLA20n was maintained exclusively in 5iLAF on inactivated MEFs under normoxic conditions according to methods described above for reverted hESC lines. UCLA19n was cultured from P0-P14 in 5% O_2, 3%CO_2 in 5iLAF on MEFs. Pluripotent stem cell identity for UCLA19n was confirmed by round dome-shaped colony morphology at all passages and positive immunofluorescence staining for TRA-1-81, OCT4 and NANOG at passage 5. Flow cytometry was performed at passage 7 revealing 60% Tra-1-81 positive and 85% SSEA4 negative cells. UCLA19n cultures were sent for karyotyping by Cell Line Genetics Inc. (Madison, WI) at passage 15 resulting in the discovery that UCLA19n was 70% polyploidy. UCLA20n was characterized by a round, dome-shaped colony morphology at all passages, together with SSEA4 negative staining. Array comparative genomic hybridization (aCGH) was performed at passage 14 by Cell Line Genetics Inc revealing gains at chromosome 3, 7 and 12. 5iLAF ESC derivations were also attempted under normoxic conditions using n=5 day five vitrified human blastocysts. Under these conditions, 4/5 blastocysts attached at P0, however none resulted in ESC lines.

Human embryo studies were approved by the full UCLA Institutional Review Board (IRB#11-002027) and the UCLA Embryonic Stem Cell Research Oversight (ESCRO) Committee (2007-005).

Culture in Epiblast like-cell (EpiLC conditions)

MEF-depleted hESCs were plated at 200k/well in Human Plasma Fibronectin (Invitrogen)-coated 12-well-plate for 16 days in EpiLC media. Media were changed everyday and EpiLCs were split every 4 days. EpiLC media is a 50:50 ratio of DMEM-F-12 (Life Technologies) and Neurobasal media (Life Technologies) with 1x N2 supplement (Life Technologies), 1x B27 supplement (Life Technologies), 1% KSR (Life

Technologies), 10ng/mL bFGF (Peprotech), 20ng/mL Activin A (Peprotech), 10μm Y-27632 (Stemgent).

Western quantitation

Sorted cells were centrifuged 800g 5 minutes, then resuspended in 1xLaemmli buffer at 5k cells/μL and denatured for five minutes at 99°C. Samples were run a 10% Bis-Tris gel (ThermoFisher), transferred at 60-70V for 3 hours, and blocked with 1xOdyssey Blocking Buffer overnight (Licor). Primary and secondary antibody incubation was conducted in 1xOdyssey Blocking buffer 0.15% Tween. The following antibodies were used to stain and quantify DNA methyltransferase levels:

Antigen	Catalog Number	Manufacturer	Concentration
UHRF1	373750	Santa Cruz	1:500
DNMT1	20701	Santa Cruz	1:500
DNMT1	Gift from S. Pradhan	New England Biolabs	1:5000
DNMT3A	20703	Santa Cruz	1:1000
DNMT3A	13888	Abcam	1:1000
DNMT3B	2851	Abcam	1:1000
DNMT3L	39908	Active Motif	1:500

Because fluorescently labeled anti-mouse and anti-rabbit antibodies can be used simultaneously (provided they are conjugated to different fluorophores), multiple proteins were stained simultaneously. Both DNMT3A antibodies were used simultaneously to allow identification of the correct band. For other proteins, costaining with anti-H3 antibody (either Abcam 1791 or Abcam 10799) at 1:5000 was performed to confirm similar loading or to provide relative concentration. After antibody staining, the cells were washed four times with 1xPBS 0.1% tween. Fluorescently conjugated anti-mouse and anti-rabbit secondary antibody (Licor) were used at 1:20,000 concentration in 1xOdyssey Blocking buffer 0.15% Tween and incubated for 45 minutes. The blots were again washed 4x5minutes with 1xPBS 0.1% tween and then rinsed quickly with 1xPBS to remove detergent. The blots were then dried and imaged on an Odyssey Infrared Imager (Licor). Band quantitation was performed using the instrument software.

Immunofluorescence

Colonies of primed or naïve cells were dissociated with Collagenase IV then fixed in 4% paraformaldehyde, embedded in paraffin, and then sectioned and mounted on slides. Slides were deparaffinized by successive treatment with xylene and 100%, 95%, 70% and 50% ethanol, and antigen retrieval was performed by incubation with 10mM Tris pH 9.0, 1mM EDTA, 0.05% Tween 95°C for 40 minutes. The slides were cooled and washed with 1xPBS and 1xTBS 0.05% tween. The samples were permabilized with 0.5% Triton X-100 in 1xPBS, then washed with 1xTBS 0.05% Tween and blocked with 10% donkey serum in 1xTBS-tween. Primary antibody incubation was conducted overnight in 10% donkey serum, using these antibodies:

Antigen	Catalog Number	Manufacturer	Concentration
Oct4	8628	Santa Cruz	1:100
Nanog	AF1997	R&D Systems	1:20
SSEA4	MC-813-70	DSHB	1:100
Tra-1-81	14-8883-82	eBiosciences	1:100

Samples were again washed with 1xTBS-tween and incubated with fluorescent secondary at 1:250 for 45 minutes, then washed and mounted using with ProLong Gold Antifade Mountant with DAPI (ThermoFisher). Slides were imaged on an LSM 780 Confocal Instrument (Zeiss).

DNA preparation.
DNA for bisulfite sequencing was extracted using the Quick gDNA Mini-Prep Kit (Zymo) and quantified using the Qubit dsDNA High Sensitivity Kit (Life Technologies).

RNA preparation.
RNA for RNA-seq was extracted using the RNeasy Micro Kit (Qiagen) and quantified using a Nanodrop ND-1000 (Nanodrop).

Library preparation
RNA sequencing libraries were prepared using the Nugen RNA-seq System V2 with 5-100ng starting material. Bisulfite sequencing libraries were prepared using the Ovation Ultralow Methyl-Seq Library System (Nugen). Unmethylated Lambda phage DNA (NEB) was spiked in at 0.25% input DNA quantity to determine conversion efficiency, which was 99.3%-99.5% for all libraries.

Sequencing
Libraries were sequenced on Illumina HiSeq instruments (Illumina).

RNAseq analysis
Analysis of individual gene expression
Reads were first aligned to hg19 gene annotation using Tophat (Trapnell et al., 2009) by allowing up to two mismatches and only keeping reads that mapped to one location. When reads did not mapped to the annotated genes, the reads were mapped to hg19 genome. Number of reads mapping to genes were calculated by HTseq (Anders et al., 2015) with default parameters. Expression levels were determined by RPKM (reads per kilobase of exons per million aligned reads). For RNAseq of published datasets (Chan et al., 2013), raw reads were processed exactly the same as described above.

Analysis of published array data
Different datasets were processed slightly differently. For Gafni et al, Ware et al and Theunissen et al, processed gene expression levels were downloaded from Gene Expression Omnibus (GEO) or European Bioinformatics Institute (EBI). Microarray probe ID were converted to gene symbol using Bioconductor packages in R. Different probes corresponding to same gene were randomly chosen for future processing. For the gene expression level of Takashima et al, raw expression datasets were downloaded from EBI database. Raw data were processed using Bioconductor packages in R. Affymetrix arrays were normalized using the RMA method, and genes with multiple probes were represented by the arithmetic mean value.

Heatmap on pluripotency genes
RPKM were obtained for each pluripotency genes. Heatmap was plotted over log2 fold changes of 5iLAF SSEA4 negative and 5iLAF SSEA4 positive comparing to Primed cells in R.

Analysis of "pre-implantation blastocyst epiblast" and "primed hESC" up-regulated genes

Pre-implantation blastocyst epiblast and primed hESC expression level (RPKM) as well as differential expressed genes list were obtained from published data (Yan et al., 2013). Genes with greater than 4 fold change as well as a FDR less than 0.05 in epiblast compared to primed hESC are defined as "pre-implantation blastocyst epiblast" up-regulated genes and vice versa.

Whole Genome Bisulfite Sequencing Analysis

Reads were split into 50 bp reads before mapping. Reads were aligned to the hg19 genome using BSMAP (Xi and Li, 2009) by allowing up to 2 mismatches and only retaining reads mapped to one location. Methylation ratio are calculated by #C/(#C+#T) at CG sites.

Metaplot of WGBS data

Metaplot of WGBS data were made using custom Perl and R scripts. Briefly, regions of interest were broken into 50 bins while flanking 1kb regions were each broken into 25bins. CG methylation level in each bin was then determined. Metaplots were then generated with R.

Analysis on imprints

Coordinates for stable primary imprints were obtained from published data (Okae et al., 2014). Transient maternal imprints were defined as CpG islands having higher methylation in blastocyst than sperm (>20% absolute difference), no substantial evidence of *de novo* methylation in blastocyst (<20% absolute increase between cleavage and blastocyst) and low methylation in later development (<20% methylation in brain), using methylation data from (Smith et al., 2014). Percent methylation over imprints was called using data from CG methylation levels were then calculated on those imprints by custom Perl scripts.

Repeat analysis on L1HS, L1PA2, L1PA3

Repeat annotation file of hg19 was downloaded from UCSC genome browser (http://genome.ucsc.edu/). For the metaplot of L1HS and L1PA2, only repeats longer than 6kb were retained for plotting.

Comparison to RRBS data

To compare WGBS sets to published RRBS data (Gafni et al., 2013), we used a custom Python script to filter mapped WGBS data and eliminate data from any CG that was not covered at least once in the RRBS sets. Any imprints that did not have at least one hundred methylation calls for CGs (e.g., if there is tenfold coverage of one CG, that is ten calls), were excluded from further analysis, so only sixteen of the thirty-one possible imprints were analyzed.

Alterations to images

Two brightfield microscopy images (Figure 1A, 1C lower image) were brightened using Adobe Photoshop in order to improve the visibility of the printed figures. Brightness was increased uniformly across the image.

V. Supplementary 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.

Chan, Y.S., Goke, J., Ng, J.H., Lu, X., Gonzales, K.A., 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.
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.
Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M.*, et al.* (2014). Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet *10*, e1004868.
Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611-615.
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*, 471-487.
Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*, 232.