Reprogramming epiblast stem cells into pre-implantation blastocyst cell-like cells

Kiichiro Tomoda,1,2,3 Haiming Hu,4 Yoshiki Sahara,4,5,6 Hashimita Sanyal,4 Minoru Takasato,4,5 and Cody Kime4,*

1Gladstone Institutes, San Francisco, CA 94158, USA
2Center for iPS Cell Research and Application, Kyoto 606-8507, Japan
3Osaka Medical College, Osaka 569-8686, Japan
4RIKEN Center for Biosystems Dynamics Research, Kobe 650-0047, Japan
5Laboratory of Molecular Cell Biology and Development, Department of Animal Development and Physiology, Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan
6Department of Renal and Cardiovascular Research, New Drug Research Division, Otsuka Pharmaceutical Co. Ltd., Tokushima 771-0192, Japan

*Correspondence: cody.kime@riken.jp
https://doi.org/10.1016/j.stemcr.2021.03.016

SUMMARY

Recently, a new wave of synthetic embryo systems (SESs) has been established from cultured cells for efficient and ethical embryonic development research. We recently reported our epiblast stem cell (EPISC) reprogramming SES that generates numerous blastocyst (BC)-like hemispheres (BCLH) with pluripotent and extraembryonic cell features detected by microscopy. Here, we further explored the system over key time points with single-cell RNA-sequencing analysis. We found broad induction of the 2C-like reporter MERVL and RNA velocities diverging to three major cell populations with gene expression profiles resembling those of pluripotent epiblast, primitive endoderm, and trophectoderm. Enrichment of those three induced BC-like cell fates involved key gene-regulatory networks, zygotic genome activation-related genes, and specific RNA splicing, and many cells closely resembled in silico models. This analysis confirms the induction of extraembryonic cell populations during EPISC reprogramming. We anticipate that our unique BCLH SES and rich dataset may uncover new facets of cell potency, improve developmental biology, and advance biomedicine.

INTRODUCTION

Early embryonic development research was the basis for developmental biology and subsequent stem cell biology. In recent decades, much has been learned from mammalian embryology animal models that are not subject to the ethical considerations for human embryos (Hyun et al., 2020; Rossant and Tam, 2009, 2017). Early embryos were used to derive various pluripotent and multipotent stem cell lines with characteristics and chimera-integrating potential analogous to the assumed embryonic origin (Evans and Kaufman, 1981; Martin, 1981). However, the ability to form synthetic embryos from cultured cells had been elusive until recent advances enabled in vitro synthetic embryo systems (SESs) (Tomoda and Kime, 2021).

SESs are part of a newly emerging generation of models akin to organoids, but reflecting early embryology through “embryoids” that are far more convincing (Harrison et al., 2017; Kime et al., 2018, 2019; Rivron et al., 2018; Shahbazi and Zernicka-Goetz, 2018; Tomoda and Kime, 2021; Zheng et al., 2019). Several SESs exist, and some focus on modeling the blastocyst (BC) and its three layers: trophoderm (TE), primitive endoderm (PE), and pluripotent pre-implantation epiblast (EPI). Some SESs utilize embryonic stem cell (ESC) and trophoblast stem cell (TSC) aggregations to model the pre-/post-implantation embryo in vitro. Other SESs, including those produced by our group, involve cell reprogramming or unique cell plasticity states that give rise to BC-like cysts from single cultures. Various approaches building on BC-like cyst formation in vitro continue to be developed and explored as each SES uniquely broadens embryology (Tomoda and Kime, 2021).

In a related field, reprogramming cells with exogenous factors (e.g., transcription factors, small molecules, cytokines, nutrients) pioneered new dimensions in cell biology by inducing donor cells to adopt desirable or unforeseen synthetic states (Davis et al., 1987; Kime et al., 2016, 2019; Takahashi et al., 2007; Woogeng et al., 2020). Indeed, analogs of early embryonic BC-lineage cells have been induced by many scientists (Benchetrit et al., 2019; Kubaczka et al., 2015; Parenti et al., 2016; Takahashi and Yamanaka, 2006). Our past epiblast stem cell (EPISC) reprogramming induced high-quality chimera-forming naive-like cells with X chromosome reactivation (Kime et al., 2016). We recently showed that the same reprogramming generated plates of BC-like hemispheres (BCLHs) with KRT8+ (TROMA-I) TE-like cells surrounding the Xa/Xa EPI-like naive ESC region, which also had a PE-like GATA4+/GATA6+/PDGFRA+ population at its inner face toward the putative blastocoeI (Kime et al., 2018, 2019). However, detailed single-cell gene expression and regulation of the converting BCLH cells to resemble the three BC-lineage cells was previously unknown.

The BCLH SES can be easily set up and generates BCLHs efficiently from EPISC cultures. We established EPISCs with
the 2C-reporter MERVL and saw broad expression prior to BCLH cyst-like formation. We applied single-cell RNA sequencing (scRNA-seq) and saw that on day 5, three distinct regions of cells branched out with gene expression resembling the BC’s TE, PE, and EPI lineages. The three regions each had RNA velocity toward day-7 cells that extended the regions and further enriched convincing cells of the postulated BC-cell identities. Furthermore, RNA splicing regulation and gene-regulatory networks implied that significant cell reprogramming had occurred with germ and zygotic genome activation (ZGA) signature genes. Here, we detail these observations and anticipate a welcome interest in the relatively poorly explored aspect of EPISC SES reprogramming into early embryonic cells.

RESULTS

Naive ESCs in 2iLIF may stabilize 2C/MERVL+ reporter expression

For this study we required control naive ESCs to model early pluripotency. We therefore integrated MERVL::RFP reporters into B6N ESCs, which we cultured as previously reported (Kime et al., 2019). These cells are passaged with Accutase to single cells every few days on iMatrix511 in 6-well plates (see experimental procedures). Two distinct populations of ESCs stabilized within the same culture: one with the traditional dome-like morphology of naive ESCs with transient MERVL::RFP expression (Macfarlan et al., 2012) and the other with a unique larger cell morphology with sustained MERVL::RFP expression (Figure 1A). Our scRNA-seq analysis of the culture confirmed our suspicion of a “duality” because cells clustered into two distinct groups (Figure 1B) that we bisected at the origin of UMAP_1: the left cluster termed “ESC” and the right cluster with much more MERVL::RFP “2C-like” reporter expression we termed “ESC2CL” (figures 1B and 1C). Although the cells were technically cultured and processed for scRNA-seq as one sample, the ESC2CL cells had much higher mean scRNA-seq features (ESC: 7,277; ESC2CL: 7,998; p value: 1.88711 \times 10^{-119}) and counts (ESC: 53,285; ESC2CL: 69,568; p value: 2.576207 \times 10^{-108}) shown by two-sample t test and violin plot (Figure S1A). These differences may come from ESC2CL cells being visibly larger or from harboring more mature transcripts accessible to the 3’ poly(A) priming of the scRNA-seq kit. When compared, both populations generally retained similar core pluripotency features (Figure 1D), and although differential gene markers could also be identified (Tables S1 and S2), there were few outstanding genes in each group’s top 20 (Figure S1B). Because of their technical similarity in scRNA-seq data and our interest in 2C/MERVL regulation, we used both ESC and ESC2CL clusters throughout this study.

The BCLH SES induces 2C/MERVL reporter, XGFP, and three regions of blastocyst-like lineage cells

We previously generated (Kime et al., 2018, 2019) EPISCs with XGFP and MERVL::RFP reporters that are completely off when viewed by fluorescence microscopy (Figure 2A). We induced BCLH reprogramming with these cells and sampled them for scRNA-seq at day 5 and day 7, along with the starting EPISCs and the dual ESC/ESC2CL population, using a standard workflow including SkewC (Abugessaisa et al., 2020) to select high-quality cells for analysis (Figure S2A). On day 4 we rarely spotted XGFP expression while many cells showed MERVL::RFP activation that continued through day 5 and day 7 as XGFP became activated (Figure 2B). Activation of MERVL across full colonies was also visible with a rapidly degrading MERVL::D2nRFP reporter (Kime et al., 2018; Li et al., 1998) (Figure S2B).

We prepared our scRNA-seq samples with Seurat’s SCTransform (Butler et al., 2018) and found a clear trend of three regions of cells on day 5 that clustered closely to similar expanded surrounding regions on day 7 (Figure 2C) across comparable ranges of features and counts with control cells (Figure S2C). In general, cells in the three regions often clustered together and the day-7 cells had more distinguished separation (Figure 2C). The EPISCs, the ESCs, and the ESC2CL cells clustered separately (Figure 2C). Surprisingly, many reprogramming cells in all regions had significant MERVL::RFP expression (Figures 2D and 2E), and mostly one specific region of day-7 cells expressed the XGFP reporter (Figure 2F) that may indicate naive pluripotent cells (Bao et al., 2009; Kime et al., 2016).

Checking gene expression revealed that common late BC-lineage cell markers were enriched and relatively focused in three separate regions characterized by Pluripotent/Epiblast-like, TE-like, and PE-like genes (Figures 2H–2J). The Epiblast-like region colocalized with the XGFP expression (Figure 2F) and was enriched for pluripotent genes (Pou5f1, Zfp42, Tdgf1, Nr0b1, Klf2) (Figure 2H). Importantly, XGFP was detected in the cells where Klf2, Klf4, and Prdm14 were expressed (Figures 2F and 2H), consistent with the previously reported role of these genes in reactivating the inactive X chromosome (Gillich et al., 2012; Kime et al., 2016). In the TE-like region, we observed numerous important TE-establishing genes (Ets2, Tfaq2c, Gata2, Gata3, Elf4) (Figure 2I) and remarkable expression of Krt8 and Krt18, which were recently reported to organize extraembryonic fate determination in the compacting and polarizing embryo (Lim et al., 2020). We also found markers of more specialized TE-lineage cells (Hughes et al., 2013; Latos and Hemberger, 2016) enriched within the TE-like region (Figure S2D), suggesting some TE-lineage differentiation therein. Also, the smaller PE-like region had milder yet focused enrichment
for important PE genes (Pdgfra, Gata4, Gata6, Fgfr1, Lifr, Lama1) (Figure 2J).

The 2C-like MERVL reporter in ESCs has been used in several studies, yet its use in naive ESCs has been limited due to an unclear relationship to ZGA early embryonic-like plasticity. In our SESs we found utility with this reporter related to heightened cell plasticity (Kime et al., 2018, 2019). We therefore checked the expression of numerous recently reported ZGA-like regulators and ZGA signature genes derived from powerful screens (Alda-Catalinas et al., 2020). Many ZGA-like regulators were induced in reprogramming cells (Figure 2K), and ZGA signature genes were often highly expressed broadly or regionally in day-5 and day-7 reprogramming cells (Figure S2E), while many were not expressed in the stem cell controls including the ESC2CL cells (Figure S2E). As such, we hypothesize that ESCs may have a low threshold for MERVL reporter activation that reflects lesser 2C-like/ZGA gene

Figure 1. Mouse ESCs and MERVL reporter expression
(A) Duality of ESC culture in 2iLIF viewed with brightfield imaging (top) and MERVL::RFP expression (bottom). Scale bars, 200 μm.
(B) UMAP-based gene expression feature plot for transgenic MERVL::RFP.
(C) UMAP plot with labeled ESC and ESC2CL populations.
(D) UMAP-based gene expression feature plots for pluripotency genes.
differences while de novo MERVL reporter activation in reprogramming cells might indicate a greater composition of ZGA-like genomic remodeling.

**Three regions of blastocyst likeness enrich over time**

To investigate the state of cells on day 5 and day 7, we employed RNA velocity with Velocyto (La Manno et al., 2018) to determine the "direction" of cell-state change and view RNA splice variation. The EPISC and the ESC/ESC2CL had RNA velocities pointing inward, demonstrating stable RNA splice variation. The EPISC and the ESC/ESC2CL to determine the "direction" of cell-state change and view (K) UMAP-based gene expression feature plots for ZGA-like regulators.

Figure 2. BCLH cell induction and regional gene expression

(A) EPISC culture with MERVL::RFP and XGFP reporters showing brightfield imaging (top) MERVL::RFP expression (middle), and XGFP expression (bottom). Scale bars, 200 μm.

(B) BCLH reprogramming from EPISC with MERVL::RFP and XGFP reporters imaged on day 4, day 5, and day 7, for brightfield (top), MERVL::RFP expression (middle), and XGFP expression (bottom). Scale bars, 100 μm (Day4) and 200 μm (Day5, Day7).

(C) UMAP plot clustering of EPISC, ESC, ESC2CL, day-5, and day-7 samples.

(D) Violin plot of transgenic MERVL::RFP reporter expression in EPISC, ESC, ESC2CL, day-5, and day-7 samples.

(E) UMAP-based gene expression feature plot for transgenic MERVL::RFP.

(F) UMAP-based gene expression feature plot for transgenic XGFP.

(H-J) UMAP-based gene expression feature plots for Pluripotent/Epiblast genes (H, green), TE genes (I, purple), and PE genes (J, burgundy) associated for BC-like regional likeness.

(K) UMAP-based gene expression feature plots for ZGA-like regulators.
The gene regulation and RNA splice differences raised uncanny distinction of the regions. We investigated numerous mouse RNA spliceosome genes (Kanehisa and Goto, 2000) and found some with discrete differences among the day-specific regions (Figure S4A). The Mbnl splice factors that repress naive pluripotency-specific splicing (Han et al., 2013) were only active in the TE-like and PE-like region cells (Figure S4A). Also, Mbnl3, a core trophoblast gene induced by Gata3 and Cdx2 (Ralston et al., 2010), was neatly expressed in the TE-like region. Interestingly, Mbnl2 was one of the top 20 markers for the TE-like region among other TE genes (Figure S4B).

Taken together, the BCLH SES reprogramming cells formed three diverging regions over time with specific epigenetic splicing, expression, and downstream regulation that grossly reflected the three BC-cell lineages; such divergence over time may explain the controlled order and development of BCLHs observed in culture (Kime et al., 2019).

Some BCLH cells adopt the regulatory networks of established models

BCLH SES reprogramming induces many cells organized spatially like BCs (Kime et al., 2019). To better isolate and characterize the most BC-like cells in silico, we selected the induced cells that were most ICM/Epiblast-like (iEPI), TE-like (iTE), and PE-like (iPE) based on critical gene expression criteria (see experimental procedures). We also included comparable numbers of the ESCs/ESC2CL cells and starting EPISCs. For established TE and PE model cell data, we sourced a loom file (Posfai et al., 2021) that was built from established reports’ scRNA-seq data. Using the 17,242 intersecting features of all samples, we then integrated with 5,000 anchors to correct for batch differences (see experimental procedures).

Seurat UMAP clustering from the integrated gene expression generally showed distinct populations based on type, although there was a general clustering of extraembryonic (iTE/TE/iPE/PE) samples (Figure 4A), similar to the TE/PE analysis in the model cell study (Posfai et al., 2021). Many TE and PE cells clustered together with iTE and iPE cells, likely from common extraembryonic expression. Excitingly, analysis with SCENIC showed distinct trends among 152 regulons that distinguished the putative similar cells (Figure 4B), demonstrated by the single-cell level total regulon activity plot (Figure S4C). SCENIC binarized regulon analysis revealed, again, the distinction between the pluripotency and extraembryonic cells with high similarity to established models (Figure S4D). Both SCENIC regulon plots (Figures 4B and S4D) had more clarity than average gene expression alone (Figure S4E), which highlighted the importance and power of gene-regulatory networks to determine a cell identity (Aibar et al., 2017). As expected, the iEPI cells were regulated alike to the ESCs/ESC2CL cells, having lost nearly all EPISC-specific regulons and acquiring regulons of naive pluripotency (Figure 4B) more clearly than the whole Epiblast-like cluster (Figure S3D). iPE/iTE samples had lost most EPISC-specific regulons that were also not found in the PE/TE cell models. In fact, the iPE/iTE and PE/TE model cells mostly had similar regulon enrichment profiles (Figure 4B) despite minor differences that appeared to come from a batch effect. Importantly, the iPE enriched key PE regulons (e.g., Gata4, Gata6, Sox17) and the iTE enriched many TE regulons (e.g., Cdx2, Gata2, Klf6).

We imported the regulon activity tables to the Seurat object and used FindMarkers to identify the top 10 regulon markers of the naive pluripotent, PE, and TE samples and plot their regulon activity (Figure S4F). Indeed, many top markers discovered were highly reported genes that define the correlating EPI/TE/PE cell states (e.g., Klf2, Klf4, Nanog::Cdx2, Ets2, Klf6::Sox7, Gata4, Gata6) and were regulated relatively neatly among induced and model embryonic/extraembryonic populations (Figure S4F). In general, each of the three induced BC-like cell populations in BCLHs appeared to be regulated by the critical transcription factors of their putative equivalents.

Reclustering cells based on SCENIC regulon activity can provide clearer identity-based clustering (Aibar et al., 2017; Posfai et al., 2021). Upon doing so in t-distributed stochastic neighbor embedding (tSNE) map space, the

---

**Figure 3. RNA velocity and gene regulation in the three BCLH regions**

(A) UMAP-based RNA velocity plot for EPISC, ESC, ESC2CL, day-5, and day-7 samples with BC-like regions labeled based on regional gene expression (Figure 2).

(B) UMAP-based RNA splicing plots for spliced reads (s) and unspliced reads (u), with Cell Type coloring (A) and residual (resid) unspliced expression shown. RNAs detected are regionally labeled by color for pluripotency (green), TE (purple), and PE (burgundy).

(C) UMAP plot of downsampled day-specific cells and control cells for use in gene heatmap (D) and SCENIC regulon analysis (E and F; see also Figure S3D).

(D) Heatmap (Do heatmap) plot of the top 100 variable features of all cells, ordered by day-specific regions of cells and control cells.

(E) SCENIC binarized regulons with heatmap (pheatmap) clustering and binarized regulon frequency (red scale) within the population.

(F) Heatmap (pheatmap) of the row-matched transcription factor average expression (log-transformed) for the regulons of E.
(legend on next page)
distinction of iTETE and iPPE marginally improved among better distinction between the pluripotent and extraembryonic states (Figure 4C). As seen with the regulon heatmaps (Figure 4B), the ESC and ESC2CL cell regulon activity caused those cells to nearly share the same tSNE space, strengthening the notion that these cells were more similar at the gene-regulatory level despite the stable MERVL reporter expression only in ESC2CL (Figures 4B and 4C). As expected, the iEPI cells now clustered closer to the ESCs/ESC2CL cells (Figure 4C). To view critical regulon activity in individual cells, we prepared three tSNE plots with BC-lineage-specific regulons on each plot that explained the regulatory activity responsible for the diverse cell types (Figure 4C, top right and bottom).

To learn more about the iTETE and iPPE identity reprogramming, we compared the sample gene expression with the model cells (TE and PE), with in vitro stem cell counterparts derived from natural BCs (TSC and XEN), and with in vitro reprogrammed cells (GETSM_TSC, cXEN, iXEN) from previously reported bulk RNA-seq data (Benchetrit et al., 2019; Gao et al., 2018; Huang et al., 2020; Parenti et al., 2016) (see experimental procedures). We based the gene list on the 152 regulon gene names that the SCENIC analysis learned while neatly separating our cells. Indeed, a core TE/TSC signature could be identified from the TE and TSCs, which was generally reflected in both reprogrammed iTETE and GETSM_TSC. However, the iTETE was closer to TE than to TSCs, possibly reflecting the differences between TE and TSCs, or batch effects (Figure 4D, left). Similarly, a core PE/XEN signature could be identified from the PE and XENs, which was well reflected in the iPPE and iPXEN samples and less clear for cXEN samples (Figure 4D, right).

Overall, the BCLH SES demonstrated remarkable cell reprogramming from primed pluripotent EPISCs that engaged sophisticated gene regulation and RNA splicing to produce regions and cell output (iEPI, iTETE, and iPPE) with the critical gene expression of pre-implantation embryo cells (TE/PE) and established cell lines (ESC/TSC/XEN).

FIGURE 4. Select BCLH SES cells reprogram meaningfully close to model cells

(A) UMAP-based plot for EPISC, ESC, ESC2CL, iEPI, iPPE, iTETE, PE, and TE samples.
(B) SCENIC total AUC regulon activity for EPISC, ESC, ESC2CL, iEPI, iPPE, iTETE, PE, and TE samples.
(C) Top-left panel: SCENIC tSNE plot based on AUC regulon activity. Top-right and bottom panels: average regulon activity at single-cell level in RGB color for pluripotency regulons (red), PE regulons (green), and TE regulons (blue) across the tSNE plot.
(D) Heatmap (pheatmap) of row-clustered TE/TSC signatures and PE/XEN signatures of reprogramming and model cells from this study and others (Benchetrit et al., 2019; Gao et al., 2018; Huang et al., 2020; Parenti et al., 2016; Posfai et al., 2021). Yellow outline surrounds the model cells (e.g., TE/TSC and PE/XEN), and cells outside the box are experimental (e.g., iTETE/GETSM_TSC and cXEN/iPPE/iXEN).

**DISCUSSION**

Until now we had seen self-assembly and order that resembled BCs in the BCLH SES (Kime et al., 2018, 2019) that emerged in previous reprogramming works (Kime et al., 2016). In addition to confirming those observations, this study provides numerous insights into gene expression, RNA splicing regulation, and gene-regulatory networks that greatly strengthened our hypothesis that EPISCs can reprogram to represent BCs. The numerous ZGA signature and germ program related genes lead us to wonder how the cell reprogramming is engaged, although anticipated Dux and Zscan4 expression was not detected at these time points (Hendrickson et al., 2017).

The BCLH SES is unique for its cell origin, defined culture conditions, and output. Given that the MERVL reporter activated early on, and day-5 cells neatly branched toward day 7, we anticipate that an earlier-time-point reprogramming precursor cell may exist. It will be important to dissect the reprogramming process with earlier time points and shorter intervals to fully understand how the initial reprogramming and lineage divergence occurs. Clues in unpublished aspects of this study continue to suggest germ program features as previously hypothesized (Kime et al., 2019). Discovery and optimization of a unique precursor, reprogrammed from EPISCs, may further advance our understanding of the distinctive properties of this SES. We anticipate that lineage tracing or cell selectivity based on the iEPI/iPPE/iTETE data may help discern key precursors and emerging populations. We would also like to include in scRNA-seq analysis early BC ICM cells and BC Epiblast-specific cells that we could not presently access.

The MERVL reporter has had significant utility in our SESSs, and in the BCLH system is more broadly induced than in the induced BC-like cysts (iBLCs) (Kime et al., 2019), where the defined conditions are altered in different phases. Conversely, BCLH SES cells proceeded more rapidly to form less organized BC-like hemispheres instead of puckering from the plate as floating self-organizing cysts. Since the MERVL reporter was highly active in all three regions of this study across day 5 and day 7, it provided some clues about unique 2C-like early embryonic programs that may be engaging the genome. Surprisingly, the ESC2CL cells in our base condition had sustained MERVL::RFP expression and had interesting differences from ESCs in gene expression despite overwhelming similarity at the gene-regulatory network level. We speculate that the MERVL reporter may have significant utility in our cell
reprogramming that had numerous ZGA-related genes, yet wonder if our data here challenge the MERVL reporter’s value in 2iLIF conditions.

**Cell reprogramming**

In general, the reprogramming cells lost their donor cell state and took on the programs of early BC-like cell lineages, which reflected prior observations in great detail. The XGFP\(^\star\) iEPI population was previously shown as readily potentiated for high contribution in chimeric embryos (Kime et al., 2016). Although TE-like and PE-like regions could be identified and harbored cells with convincing TE or PE properties, we wonder whether such cells could seed TSC or XEN cell cultures if transferred in selective conditions. In the BCLH SES, the TE-like region generally had infrequent Cdx2 detection despite the specific enrichment of the Cdx2 regulon, perhaps related to abundant keratin expression and TE-related transcription factor involvement. Interestingly, the TE can be specified independent of Cdx2 (Wu et al., 2010). Although Cdx2 is not required for BC formation (Meissner and Jaenisch, 2006), its role in implantation is important, and distinct Cdx2\(^+\) TE is roundly regarded (Strumpf et al., 2005). The TE could be specified in 2006), its role in implantation is important, and distinct Cdx2\(^+\) TE is roundly regarded (Strumpf et al., 2005). The PE-like region was less distinct and had significant overlap with the Epiblast-like and TE-like regions; perhaps the molecular distinction of the emerging iPE is as complicated as its natural ICM-to-extraembryonic transition at the late BC stage. Throughout this study we observed traces of gene expression suggesting that iPE and PE-like region cells shared a pluripotent-like origin with the iEPI population, reminiscent of the GATA4\(^*,\) GATA6\(^*,\) and PDGFRA\(^*\) cells arising at the inner face of BCLH pluripotent cells (Kime et al., 2019). Consideration for PE cells has weighed heavily on various SEs (Tomoda and Kime, 2021), and we suspect that correct hypoblast formation will remain a hinge point for healthy embryoid development.

**EXPERIMENTAL PROCEDURES**

**EPISC culture and BCLH reprogramming**

EPISC culture and reprogramming was performed as previously (Kime et al., 2016, 2019), and reprogramming included sodium pyruvate.

**ESC culture and MERVL reporter integration**

CS7BL/6N (B6N) ESCs were converted from 3iLIF conditions and cultured in 2iLIF conditions with CTSFES basal medium as described previously (Kime et al., 2019) on iMatrix511-coated 6-well plates. Cells were integrated with and selected for MERVL::RFP reporters (mCherry and D2nmCherry) in piggyback vectors, the same as with EPISCs in our previous study (Kime et al., 2019). After several passages, the two different populations of ESC/ESC2CL cells became obvious and stabilized.

**Microscopy**

Brightfield and live-cell RFP and GFP fluorescence was imaged with an Olympus IX71 microscope.

**scRNA-seq sampling and processing**

Cells were dissociated and passed through cell screen cuvettes to isolate mostly healthy single cells that were prepared with 10X Genomics Chromium Single Cell 3’ Library & Gel Read Kit V3.0. Sample libraries were finalized and sequenced on one HiSeq X lane (150 bp PE; MacroGen) for each. The standard Cell Ranger protocol detected sample chemistry and produced “possorted” BAM files from which the subsequent Primary Analysis workflow in Figure S2A was performed to filter high-quality cells. For analyses shown in Figures 4 and S4, to match loom data counts for TE and PE control cells (Posfai et al., 2021) we prepared our data tables similarly with DESeq2 size factor normalization and log transformation. We integrated samples with intersecting features and SerumCCA integration across 5,000 anchors. The iEPI, iTE, and iPE cell barcodes were selected from reprogramming cells with the following criteria:

- iEPI: Zfp42 > 0.1 & Klf2 > 0.1 & EGFp > 0 & Prdm14 > 0.001
- iTE: Cdx2 > 0 & Gata2 > 0
- iPE: Gata6 > 0 & (Sox7 > 0 | Gata4 > 0 | Sox17 > 0)

**Bulk external RNA-seq data processing**

FPKM (fragments per kilobase per million mapped reads) tables from Gene Expression Omnibus accession numbers (GEO: GSE98124, GSM2805965, GSM2805966, GSE154398, GSM2054370, GSM20 54371, GSM2054375, GSM2054376) were imported and log1p transformed. The first author’s name of each selected external sample’s related study was appended to the sample name for clarity (see Figure 4D). Genes of interest were reduced to the 152 gene names of the SCENIC regulons (Figure 4B). Our sample (iTE/iEPI/iPE) relative count tables were added with log1p transformation. Each combined table was scaled together and generated in heatmap with row-clustering and model cells at the center to clarify the signature patterns.

**Data and code availability**

All analysis code used in this study is available upon request. Raw and processed sequence data are accessible with accession number GEO: GSE166066.

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2021.03.016.

**AUTHOR CONTRIBUTIONS**

Conceptualization, C.K.; Methodology, C.K. and K.T.; Experimentation, C.K., H.H., H.S., and Y.S.; Formal analysis, C.K. and K.T.; Investigation, C.K.; Resources, C.K., M.T., Y.S., and K.T.; Writing – original draft, C.K.; Writing – revision & editing, C.K. and K.T.; Visualization, C.K.; Project supervision, C.K.; Bioinformatics, C.K.; Project administration and funding, C.K.
CONFLICTS OF INTEREST

C.K. and K.T. have patents related to this technology and extended works.

ACKNOWLEDGMENTS

We are grateful for the support from RIKEN Center for Integrative Medical Sciences for specific training and research environment with support from Erik Arner, Piero Carninci, Imad Abugessaisa, Akira Hasegawa, and Teruaki Kitakura. We also thank Osamu Nishimura and the Shigehiro Kuraku lab at RIKEN for providing an HPC for Cell Ranger processing. We thank Eszter Posfai of Princeton University, and both Vincent Pasque and Adrian Janiszewski of Katholieke Universiteit Leuven for providing important control data and thoughtful discussions. We thank Ivo Woogeng Ngundu for specific support during this study. We also thank Francoise Chanut for copy-editing the manuscript.

This study was primarily funded by the RIKEN Center for Biosystems Dynamics Research Organoid Project. K.T. was supported by the iPS Cell Research Fund from Center for iPS Cell Research and Application, Kyoto University, Core Center for iPS Cell Research (19bm0104001h0007), Research Center Network for Realization of Regenerative Medicine from AMED (Japan Agency for Medical Research and Development), and the L.K. Whittier Foundation.

Received: September 30, 2020
Revised: March 16, 2021
Accepted: March 16, 2021
Published: May 11, 2021

REFERENCES

Abugessaisa, I., Noguchi, S., Cardon, M., Hasegawa, A., Watanabe, K., Takahashi, M., Suzuki, H., Katayama, S., Kere, J., and Kasukawa, A., Noguchi, S., Cardon, M., Hasegawa, A., and Watanabe, M. (2020). Quality assessment of single-cell RNA sequencing data by coverage skewness analysis. BioRxiv https://doi.org/10.1101/2019.12.31.890269.

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Huiselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods 14, 1083–1086.

Alda-Catalinas, C., Bredikhin, D., Hernando-Herraiz, I., Santos, E., Kubinyezcz, O., Eckersley-Maslin, M.A., Stegle, O., and Reik, W. (2020). A single-cell transcriptomics CRISPR-activation screen identifies epigenetic regulators of the zygotic genome activation program. Cell Syst 11, 25–41.e9.

Bao, S., Tang, E., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature 461, 1292–1295.

Benchetrit, H., Jaber, M., Zayat, V., Sebban, S., Pushett, A., Make-donski, K., Zakheim, Z., Radwan, A., Maoz, N., Lasry, R., et al. (2019). Direct induction of the three pre-implantation blastocyst cell types from fibroblasts. Cell Stem Cell 24, 983–994.e7.

Buecker, C., Srivivasan, R., Wu, Z., Calo, E., Acampora, D., Fial, 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.

Burton, A., Muller, J., Tu, S., Padilla-Longoria, P., Guccione, E., and Torres-Padilla, M.-E. (2013). Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. Cell Rep. 5, 687–701.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51, 987–1000.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–156.

Gao, H., Gao, R., Zhang, L., Xiu, W., Zang, R., Wang, H., Zhang, Y., Chen, J., Gao, Y., and Gao, S. (2018). Esrrb plays important roles in maintaining self-renewal of trophoblast stem cells (TSCs) and reprogramming somatic cells to induced TSCs. J. Mol. Cell Biol. 11, 463–473.

Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W.B., Pasque, V., Magnusdottir, E., and Surani, M.A. (2012). Epiblast stem cell-based system reveals reprogramming synergy of germline factors. Cell Stem Cell 10, 425–439.

Han, H., Irimia, M., Ross, P.J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature 498, 241–245.

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

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.

Home, P., Kumar, R.P., Ganguly, A., Saha, B., Milano-Foster, J., Bhattacharya, B., Ray, S., Gunewardena, S., Paul, A., Camper, S.A., et al. (2017). Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage ensures the progression of preimplantation and postimplantation mammalian development. Development 144, 876–888.

Huang, X., Bashkenova, N., Yang, J., Li, D., and Wang, J. (2020). ZFP281 recruits polycomb repressive complex 2 to restrict extraembryonic endoderm potential in safeguarding embryonic stem cell pluripotency. Protein Cell 12, 213–219.

Hughes, M., Natale, B.V., Simmons, D.G., and Natale, D.R.C. (2013). Lyve expression is restricted to syncytiotrophoblast cells of the mouse placenta. Placenta 34, 831–835.

Hyun, I., Munsie, M., Pera, M.F., Rivron, N.C., and Rossant, J. (2020). Toward guidelines for research on human embryo models formed from stem cells. Stem Cell Reports 14, 169–174.
Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30.

Kime, C., Sakaki-Yumoto, M., Goodrich, L., Hayashi, Y., Sami, S., Derynick, R., Asahi, M., Panning, B., Yamanaka, S., and Tomoda, K. (2016). Autotaxin-mediated lipid signaling intersects with LIF and BMP signaling to promote the naïve pluripotency transcription factor program. Proc. Natl. Acad. Sci. U S A 113, 12478–12483.

Kime, C., Kiyonari, H., Ohotsuka, S., Kohbayashi, E., Asahi, M., Yamanaka, S., Takahashi, M., and Tomoda, K. (2018). Implantation-competent blastocyst-like structures from mouse pluripotent stem cells. BioRxiv https://doi.org/10.1101/309542.

Kime, C., Kiyonari, H., Ohotsuka, S., Kohbayashi, E., Asahi, M., Yamanaka, S., Takahashi, M., and Tomoda, K. (2019). Induced 2C expression and implantation-competent blastocyst-like cysts from primed pluripotent stem cells. Stem Cell Reports 13, 485–498.

Krendl, C., Shaposhnikov, D., Rishko, V., Ori, C., Ziegenhain, C., Sass, S., Simon, L., Müller, N.S., Straub, T., Brooks, K.E., et al. (2017). GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. Proc. Natl. Acad. Sci. U S A 114, E9579–E9588.

Kubaczka, C., Senner, C.E., Cierlitza, M., Araújo-Bravo, M.J., Kuckartz, C., Peitz, M., Hemberger, M., and Schorle, H. (2016). From the stem of the placental tree: trophoblast stem cells and their progeny. Development 143, 3650–3660.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C.-C., and Kain, S.R. (1998). Generation of destabilized green fluorescent protein as a transcription reporter. J. Biol. Chem. 273, 34970–34975.

Lim, H.Y.G., Alvarez, Y.D., Gasnier, M., Wang, Y., Tetlak, P., Bissiere, S., Wang, H., Biro, M., and Plachta, N. (2020). Keratins are asymmetrically inherited fate determinants in the mammalian embryo. Nature 585, 404–409.

Lo Nigro, A., de Jaime-Soguero, A., Khoueiry, R., Cho, D.S., Ferrazza, G.M., Pertini, I., Abon Escalona, V., Aranguren, X.L., Chuvah de Sousa Lopes, S.M., Koh, K.P., et al. (2017). PDGFRα* cells in embryonic stem cell cultures represent the in vitro equivalent of the pre-implantation primitive endoderm precursors. Stem Cell Reports 8, 318–333.

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, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U S A 78, 7634–7638.

Meissner, A., and Jaenisch, R. (2006). Generation of nuclear transfer-derived pluripotent ES cells from cloned Cdx2-deficient blastocysts. Nature 439, 212–215.

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

Parenti, A., Haltibon, M.A., Wang, K., Latham, K., and Ralston, A. (2016). OSMK induce extraembryonic endoderm stem cells in parallel to induced pluripotent stem cells. Stem Cell Reports 6, 447–455.

Pfister, S., Steiner, K.A., and Tam, P.P.L. (2007). Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development. Gene Expr. Patterns 7, 558–573.

Posfai, E., Schell, J.P., Janiszewski, A., Rovic, I., Murray, A., Bradshaw, B., Yamakawa, T., Pardon, T., El Bakkali, M., Talon, L., et al. (2021). Evaluating totipotency using criteria of increasing stringency. Nat. Cell Biol. 23, 49–60.

Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, S., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophectoderm development downstream of Tead4 and in parallel to Cdx2. Development 137, 395–403.

Rivron, N.C., Frias-Aldeguer, J., Vrij, E.J., Boisset, J.-C., Korving, J., Vivié, J., Truckenmüller, R.K., van Oudenaarden, A., van Blitterswijk, C.A., and Geijts, N. (2018). Blastocyst-like structures generated solely from stem cells. Nature 557, 106–111.

Rossant, J., and Tam, P.P. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development 136, 701–713.

Rossant, J., and Tam, P.P.L. (2017). New insights into early human development: lessons for stem cell derivation and differentiation. Cell Stem Cell 20, 18–28.

Shahbazi, M.N., and Zernicka-Goetz, M. (2018). Deconstructing and reconstructing the mouse and human early embryo. Nat. Cell Biol. 20, 878.

Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development 132, 2093–2102.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Takahashi, K., Tanabe, K., Ohmuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872.

Tomoda, K., and Kime, C. (2021). Synthetic embryology: early mammalian embryo modeling systems from cell cultures. Dev. Growth Diff. https://doi.org/10.1111/dgd.12713.

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat. Rev. Mol. Cell Biol. 17, 155–169.

Wooeng, I.N., Abugessaisa, I., Tahibana, A., Sahara, Y., Hon, C.-C., Hasegawa, A., Kaczkowski, B., Sakai, N., Nishida, M., Hu, H., et al. (2020). Inducing human retinal pigment epithelium-like cells
from somatic tissue. BioRxiv https://doi.org/10.1101/2020.07.27.215103.

Wu, G., Gentile, L., Fuchikami, T., Sutter, J., Psathaki, K., Esteves, T.C., Araúzo-Bravo, M.J., Ortmeier, C., Verberk, G., Abe, K., et al. (2010). Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. Development 137, 4159–4169.

Yang, Y., JIN, Y., MARTYN, A.C., LIN, P., SONG, Y., CHEN, E., HU, L., CUI, C., LI, X., LI, Q., et al. (2013). Expression pattern implicates a potential role for luman recruitment factor in the process of implantation in uteri and development of preimplantation embryos in mice. J. Reprod. Dev. 59, 245–251.

Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M., Gumucio, D.L., et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. Nature 573, 421–425.
Supplemental Information

Reprogramming epiblast stem cells into pre-implantation blastocyst cell-like cells

Kiichiro Tomoda, Haiming Hu, Yoshiki Sahara, Hashimita Sanyal, Minoru Takasato, and Cody Kime
# SUPPLEMENTAL INFORMATION

## INVENTORY

| SUPPLEMENTAL FIGURES                                      | Pages |
|----------------------------------------------------------|-------|
| Figure S1: Related to Figure 1                           | 2     |
| Figure S2: Related to Figure 2                           | 3     |
| Figure S3: Related to Figure 3                           | 4-5   |
| Figure S4: Related to Figure 3, Figure 4                 | 6     |
| SUPPLEMENTAL TABLE LEGENDS                               | 7     |
Figure S1: Related to Figure 1
A) Violin plots for features and counts of ESC and ESC2CL samples.
B) Heatmap (pheatmap) of the average gene expression of the ESC top 20 markers combined with ESC2CL top 20 markers.
Figure S2: Related to Figure 2
A) An overview of the Primary and Secondary Analysis of scRNA-seq samples in this study.
B) BCLH reprogramming from EPISC with MERVL::D2nRFP and XGFP reporters imaged on Day4 for brightfield (top), MERVL::D2nRFP expression (mid), and XGFP expression (bot). Scale bar = 100μm.
C) Violin plots for features and counts of Day 5, Day 7, EPISC, ESC, and ESC2CL samples.
D) UMAP based gene expression feature plots for TE-lineage differentiation genes.
E) UMAP based gene expression feature plots for ZGA signature genes.
Figure S3: Related to Figure 3
A) UMAP based RNA splicing plots for spliced (s), unspliced (u) reads, with Cell Type coloring (Figure 3A) and residual (resid) unspliced expression shown for germ program factors.
B) UMAP based RNA splicing plots for spliced (s), unspliced (u) reads, with Cell Type coloring (Figure 3A) and residual (resid) unspliced expression shown for ZGA-like regulators.
C) UMAP based RNA splicing plots for spliced (s), unspliced (u) reads, with Cell Type coloring (Figure 3A) and residual (resid) unspliced expression for TE (purple) and PE (burgundy) genes.
D) SCENIC total AUC regulon activity by day-specific cell clusters and control cells.
Figure S4: Related to Figure 3, Figure 4

A) Heatmap (DoHeatmap) plot of notable RNA splicing (spliceosome) factors of all cells, ordered by the day-specific cell clusters and control cells.

B) Heatmap (DoHeatmap) plot of the combined top20 markers per region (2: TE-like, 5: PE-like, and 4: EPI-like) ordered by day-specific cell clusters and control cells.

C) SCENIC total single-cell AUC regulon activity for iTE, iPE, iEPI, TE, PE, ESC, ESC2CL, and EPISC.

D) SCENIC Binarized regulons with heatmap (pheatmap) clustering and binarized regulon frequency (red scale) within the population.

E) Heatmap (pheatmap) of the row-matched transcription factor average expression for the gene names of the regulons in Figure S4D.

F) Pluripotent, PE, and TE, combined Top 10 markers based on regulon activity.
SUPPLEMENTAL TABLE LEGENDS

**Table S1: ESC Markers**
ESC markers discovered from FindMarkers in Seurat.

**Table S2: ESC2CL Markers**
ESC2CL markers discovered from FindMarkers in Seurat.