TITLE

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 Kime[4,*]

1 The Gladstone Institutes, San Francisco, CA 94158, US.

2 Center for iPS Cell Research and Application, Kyoto 606-8507, Japan.

3 Osaka Medical College, Osaka 569-8686, Japan.

4 RIKEN Center for Biosystems Dynamics Research, Kobe 650-0047, Japan.

5 Laboratory of Molecular Cell Biology and Development, Department of Animal Development and Physiology, Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan.

6 Department of Renal and Cardiovascular Research, New Drug Research Division, Otsuka Pharmaceutical Co. Ltd., Tokushima 771-0192, Japan.

* Lead Contact

Corresponding Author: Correspondence should be addressed to C.K. (cody.kime@riken.jp).

ABSTRACT

Recently, a new wave of synthetic embryo systems (SESs) have been established from cultured cells toward 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 microscopically. Here, we further explored the system over key time points with unprecedented single-cell RNA sequencing (scRNA-seq) analysis and revealed broad induction of the 2C-like reporter *MERVL* and RNA velocity diverging three major population regions with genetic expression resembling pluripotent epiblast (EPI), primitive endoderm (PE), and trophectoderm (TE). Enrichment of those three BC-like cell fates involved key regulons, zygotic genome activation (ZGA) related genes, specific RNA splicing, and select cells meaningfully distinguished critical regulons of model cells. This analysis confirms the induction of the extraembryonic cell populations during the reprogramming and we anticipate that our unique BCLH SES and rich data may uncover new facets of cell potency, improve developmental biology, and help biomedicine advance.

KEYWORDS

synthetic embryology; stem cells; reprogramming; early embryo; synthetic biology

INTRODUCTION

Researching early embryonic development was the basis for developmental biology and subsequent stem cell biology. In recent decades embryology have shed light on the mammalian embryo with animal models for broad significance and to address an ethically bound human embryology (Hyun et al., 2020; Rossant and Tam, 2009, 2017). At first, early embryos were used to derive various pluripotent and multipotent stem cell lines cultures with characteristics and chimeric potential analogous to the assumed 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).

SESs are part of a newly emerging field 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; Zheng et al., 2019). Several SESs exist and some focus on modeling the blastocyst (BC) and it's three layers of trophectoderm (TE), primitive endoderm (PE), and pluripotent preimplantation epiblast (EPI). Some SESs utilize embryonic stem cell (ESC) and trophoblast stem cell aggregations to model the pre/post-implantation embryo *in vitro.* Others SESs, including work from our group, involve cell reprogramming or unique cell plasticity states that give rise to BC-like cysts from single cultures. Various SES approaches building on BC-like cyst formation *in vitro* continue to be developed and explored as each system pioneers to widen embryology at large.

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 desirable and generally unforeseen synthetic states (Davis et al., 1987; Kime et al., 2016, 2019; Takahashi et al., 2007; Woogeng et al., 2020). Indeed, cell analogs of early embryonic BC-lineage cells have been induced (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 naïve-like cells with X-chromosome reactivation (Kime et al., 2016). We recently showed that the same reprogramming generated plates of BC-like hemispheres (BCLH) with KRT8+ (TROMA-I) TE-like cells surrounding the Xa/Xa EPI-like naïve ESC region which also had a PE-like

GATA4+/GATA6+/PDGFRA+ population at its inner-face toward the putative blastocoel (Kime et al., 2018, 2019). However, detailed cellular 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 BCLH efficiently from EPISC cultures. We established EPISC with the 2C-reporter *MERVL* and saw broad early expression prior to BCLH cyst-like formation. We applied single cell RNA sequencing (scRNA-seq) and saw that on Day5, three distinct regions of cells branched with gene expression resembling the blastocyst's TE, PE, and EPI lineages. The three regions each had RNA velocity toward Day7 cells that extended those diverging regions and further enriched convincing cells of the postulated BC-cell identities. Furthermore, RNA splicing regulation and gene regulatory networks implicated significant cell reprogramming had occurred with germ and zygotic genome activation (ZGA) signature genes. Herein we detail these observations and anticipate a welcome interest in the relatively poorly explored aspect of EPISC SES reprogramming into much earlier embryonic cells.

RESULTS

Naïve ESC in 2iLIF may Stabilize MERVL+ Reporter Expression

For this study, we found TE and PE scRNA-seq data from previous reports, shown later, and required control naïve embryonic stem cells (ESCs). We therefore integrated BL6 ESCs with *MERVL::*RFP reporters that were cultured in our modified media on laminin as previously reported (Kime et al., 2019). Two distinct populations of ESCs stabilized; one with traditional naïve ESC dome-like morphology and transient *MERVL::*RFP expression (Macfarlan et al., 2012), and the other with a unique larger cell morphology and consistent range of *MERVL::*RFP expression (Figure 1A). Our scRNA-seq sample 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 (Figure 1B,C) termed 'ESC2CL'. Although technically cultured and sampled as one, the ESC2CL had higher scRNA-seq features and counts (Figure S1A). When compared, both populations generally retained similar core pluripotency features (Figure 1D) although differential gene markers could also be identified (Table S1, Table S2) with little outstanding genes each group's top 20 (Figure S1B). For technical similarity in scRNA-seq sampling of naïve ESC controls and a common interest in ESC *MERVL* reporter activity, we used both ESC and ESC2CL clusters through this study.

The BCLH SES Induces 2C-reporter, XGFP, and Three Regions of Blastocyst-like Lineage Cells

We previously generated (Kime et al., 2018, 2019) EPISC with XGFP and *MERVL::*RFP reporters that are completely off when viewed in fluorescence microscopy (Figure 2A). We induced BCLH reprogramming with and sampled the Day5 and Day7 reprogramming cells along with the starting EPISC and the duality ESC/ESC2CL for scRNA-seq with as standard workflow including SkewC (Abugessaisa et al., 2020) to select high quality cells for analysis (Figure S2A). On Day4 we rarely spotted XGFP while many cells showed *MERVL::*RFP activation that continued through Day5 and Day7 as XGFP activated (Figure 2B). The trend of full colonies expressing the MERVL reporter was also visible with a rapidly degrading D2nRFP (Kime et al., 2018; Li et al., 1998) (Figure S2B).

We prepared our scRNA-seq samples in Seurat (Butler et al., 2018) with SCTransform and found a clear trend of three regions of cells on Day5 that clustered closely to similar expanding and surrounding regions on Day7 (Figure 2C). In general, cells in the three regions often clustered together while Day7 cells had more distinguished separation (Figure 2C). The EPISC, the ESC, and the ESC2CL cells clustered separately (Figure 2C). Surprisingly, many reprogramming cells in all regions had enriched the *MERVL::*RFP expression (Figure 2D,E), while mostly one specific region of Day7 cells expressed the XGFP reporter (Figure 2F) grossly related to naïve pluripotency and the ICM.

Checking gene expression revealed that common late BC-lineage cell markers were enriched and relatively focused in the cells of three separated regions with Pluripotent/Epiblast-like, TE-like, and PE-like genes (Figure 2H,I,J). The Epiblast-like region colocalized with the XGFP expression (Figure 2F) and 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 (Figure 2F,H) relating the importance of those genes to reactivate the inactive X chromosome as previously reported (Gillich et al., 2012; Kime et al., 2016). In the TE-like region, we observed numerous important TE establishing genes (*Ets2*, *Tfap2c*, *Gata2*, *Gata3*, *Elf4*) (Figure 2I) with remarkable expression of *Krt8* and *Krt18* recently reported to organize extraembryonic fate determination in the compacting and polarizing embryo (Lim et al., 2020). Also, the smaller PE-like region there 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 naïve ESCs has been limited due to an unclear relationship to ZGA early embryonic-like plasticity. In our SESs we found utility with this reporter (Kime et al., 2018, 2019) related to heightened cell plasticity and therefore checked 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 Day5 and Day7 reprogramming cells (Figure S2C), while many were not expressed in the stem cell controls including the ESC2CL cells (Figure S2C). As such, MERVL reporter ESCs may have a narrower

threshold to activate and reflect smaller 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 Day5 and Day7, we employed RNA Velocity with Velocyto (La Manno et al., 2018) to determine the 'direction' of cell change and view RNA splice variation. The EPISC and the ESC/ESC2CL had RNA velocities pointing inward, demonstrating stable states (Figure 3A). As we anticipated there was an obvious trend of RNA Velocity from Day5 toward Day7 among the three regions, all three of which pointed away from each other (Figure 3A). Perhaps reflecting the PE formation of a BC, the ratio of cells driving into the PE-like region was fewer and distributed between the Epiblast-like and extraembryonic-like regions (Figure 2H,J, Figure 3A).

Discriminating unspliced (u) pre-mRNA vs spliced (s) mRNAs ratios with scRNA-seq and Velocyto allows for a comprehensive understanding of important state-specific splice mechanisms beyond simple RNA detection. Tdgf1 is an early BC ICM gene (Pfister et al., 2007) often detected in ESCs, however *Tdgf1* was mostly unspliced in the ESC/ESC2CL while strongly spliced and enriched in the Epiblast-like region of reprogramming cells (Figure 3A,B) where other expression seems to implicate more ICM-like expression. The germ programming genes *Prdm1 (Blimp1)* and *Smc1b* were also splice-enriched in reprogramming cells (Figure S3A) supporting previous consideration of the importance of germ genes toward higher plasticity in our SES reprogramming (Kime et al., 2018, 2019). As seen in base expression (Figure 2K), the ZGA-like regulators *Dppa2*, *Dppa4*, *Tsc22d4, Smarca5,* and *Smad1*, were induced across the reprogramming cells with splicing preference (Figure S3B) and the ESC2CL cells had spliced *Dppa2* more efficiently than the ESC. Similar region-specific splice mechanism preference was found with TE and PE genes *Fgfr2* and *Lama1 (*Figure 3B) for translation in putative cells*. Pard6b* and *Lifr* followed a similar trend while enriched in reprogramming cells' regions (Figure S3C). Taken together, the regulation of splice mechanisms and time-based BC-like cell region specification is apparent and related to pre-BC stage regulators.

To investigate reprogramming cells downstream gene regulation, we randomly downsampled the three

regions to similar cell numbers after Seurat clusters 2:TE-like, 5:PE-like, and 4:Epiblast-like with Day-specific labeling (Figure 3C). A heatmap of the top 100 global variable genes confirmed that each cluster had a unique pattern of expression (Figure 3D). We then performed SCENIC (Aibar et al., 2017) analysis to determine cell regulons. Expectedly, each region had distinct patterning that generally enriched from Day5 to Day7 (Figure S3D). The regulons were then binarized to clarify interpretation. The pluripotent Epiblast-like population had largely lost primed pluripotency EPISC-specific regulons (e.g., Pou3f1) (Buecker et al., 2014) and reprogrammed with remarkably similar regulation to the ESC/ESC2CL with well-known naïve pluripotency regulons (e.g. Mybl2, Esrrb, Klf2, Klf4) and shared broader pluripotency continuum regulons (e.g., Sox2, Nanog, Pou5f1) with EPISC (Figure 3E) (Nichols and Smith, 2009; Weinberger et al., 2016). Not surprisingly, the ESC and ESC2CL populations had few differences at the regulatory level.

As anticipated, the PE-like and TE-like cells shared important extraembryonic regulons (e.g., Klf6, Kdm5a, Creb3, Elf1, Elf4) (Figure 3E) (Burton et al., 2013; Krendl et al., 2017; Rivron et al., 2018; Yang et al., 2013). Furthermore, the TE-like region had enriched significant TE-specific regulon activity (e.g, Gata2, Gata3, Ascl2, Pparg) (Figure 3E) including Cdx2 regulon (Figure S3D) (Home et al., 2017; Krendl et al., 2017; Ralston et al., 2010). Also, the PE-like region had some enriched PE-specific regulon activity including Gata6 and Hnf1b (Figure S3D) (Lo Nigro et al., 2017). Average gene expression for the same transcription factors of regulon analysis reflected a similar pattern (Figure 3F), although far less specific, highlighting the value of specific downstream regulation analyses when comparing cells (Aibar et al., 2017; Woogeng et al., 2020).

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 naïve pluripotency-specific splicing (Han et al., 2013) were only active in the TE-like and PE-like region cells Figure S4A). Also, Mbnl3 is a core trophoblast gene induced by Gata3 and Cdx2 (Ralston et al., 2010), and was neatly expressed in the TE-like region. Interestingly, *Mbnl2* was one of the top20 markers

for the TE-like region among other TE genes (Figure S4B).

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

Some BCLH Cells Adopt the Regulatory Networks of Established Models

BCLH SES reprogramming induces many cells on the plate regulated spatially to resemble BCLHs (Kime et al., 2019). To better isolate BC-like cells *in silico*, we selected the induced cells that were most ICM/Epiblast-like (iEPI), TE-like (iTE), and PE-like (iPE) cells based on critical gene expression criteria (see methods). We also included comparable numbers of the ESCs/ESC2CL cells and starting EPISC. For established TE and PE model cell data, we sourced a loom file (Posfai et al., 2020) that was built from established reports' scRNA-seq data. We then merged samples and normalized features and counts to integrate the data fairly (see methods).

Seurat UMAP clustering from gene expression generally showed distinct populations based on type although some batch effects were obviated separating TE samples that had different origins (Figure 4A). Some TE and PE cells clustered together, as was seen with some iTE and iPE cells, likely based on common extraembryonic expression. Excitingly, analysis with SCENIC showed distinct trends among putative similar cells (Figure S4C) and SCENIC binarized regulon analysis revealed, again, the distinction of three regions of BC-like lineage cells that resembled putative models but with more established BC-like cell factors (Figure 4B). The iEPI cells were regulated alike the ESC/ESC2CL, having lost nearly all EPISC-specific regulons and acquiring those of naïve pluripotency. The iPE were regulated to have PE regulons at similar or low levels and the iTE were regulated to have many of the TE regulons. Expectedly, the induced extraembryonic-like cells (iPE/iTE) had many regulons shared with the extraembryonic PE/TE cell models (Figure 4B) and some differences that appeared to come from batch effect. It was clear that the iPE and iTE samples had lost many EPISC regulons that were also not found in the PE and TE model data, although some remained (Figure 4B). Checking average gene expression for the transcription

factors of the regulons expectedly showed a similar pattern with less distinction (Figure 4C), highlighting the importance and power of gene regulatory networks to determine a cell identity (Aibar et al., 2017). We imported the binarized regulon activity tables to the Seurat object and used FindMarkers to identify pluripotent, PE, and TE markers, select the top 10 of each, and plot the associated activity (Figure 4D). Indeed, nearly all top markers discovered were highly reported genes for their correlating cell states (e.g., Klf2, Mybl2, Nanog, Prdm14 :: Cdx2, Ets2, Gata2, Gata3 :: Sox17, Sox7, Gata4, Gata6), and were regulated relatively neatly among induced and model populations (Figure 4D). Notably, the iPE population had more iTE/TE extraembryonic regulons than the PE population (Figure 4B,D). In general, each of the three BC-like regions in BCLH appeared to be regulated by the critical transcription factors of their putative embryonic equivalents.

As shown previously, re-clustering cells based on SCENIC regulon activity can provide clearer identity-based clustering (Aibar et al., 2017; Posfai et al., 2020). Upon doing so in tSNE map, a pattern of cells reflecting the reprogramming and cell states had clarified (Figure 4E) and most of the batch effects between the TE cells was reduced, although some model PE and TE cells still mixed. As seen with the regulon heatmaps, the ESC and ESC2CL cells nearly shared the same tSNE space than when based on gene expression, strengthening the notion that these cells were more similar at the gene regulatory level when cultured in 2iLIF (Figure 4E); expectedly, the iEPI cells now clustered very close to the ESC/ESC2CL cells (Figure 4E). The iTE and iPE cells had dimensionally moved toward the TE and PE models with some iPE cells remarkably close to the PE model cluster (Figure 4E). To view individual BC-lineage critical regulon activity (Figure 4D), we prepared three tSNE plots with lineage-specific regulons on each plot that explained the regulatory activity responsible for the diverse cell types (Figure 4F).

At last, select cells of the three major regions of cells on the plate had iEPI, iPE, and iTE, cells present that had undergone remarkable cell reprogramming from primed pluripotent EPISC toward model cells of earlier embryos with sophisticated gene regulation at the regulon and RNA splicing levels.

DISCUSSION

Until now we had seen self-assembly and order to resemble BCs in the BCLH SES (Kime et al., 2018, 2019). In addition to confirming those observations, this study provided numerous aspects of gene expression, RNA splicing regulation, and gene regulator network level that greatly strengthened our understanding that EPISC can reprogram to represent BCs. The numerous ZGA signature and germ program related genes provide further inspiration to wonder how the cell reprogramming is engaged although anticipated Dux and Zscan4 expression was not seen at these time points.

The BCLH SES is unique from its cell origin, defined conditions, and output. Given that the MERVL reporter activated early on, and Day5 cells neatly branched toward Day7, we anticipate that an earlier time-point of a unified reprogramming precursor cell may exist for exploit. Discovery and optimization of that cell, which reprograms from EPISC, may further advance the clarity and distinction of this SES. We anticipate that lineage tracing based on the Day5 data may help discern key precursors and emerging populations for study and optimization. We would also like to include early BC ICM cells and BC Epiblast-specific cells for scRNA-seq analysis that we could not presently access.

The MERVL reporter has had significant utility in our SESs and in the BCLH system it is more broadly induced than in the induced BC-like cysts (iBLC) (Kime et al., 2019) where the defined conditions are modified to different in phases. Conversely, BCLH SES cells proceed more rapidly to 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 Day5 and Day7, it provided some clues about unique early embryonic programs that may be engaging the genome. Surprisingly, the ESC2CL cells in our base condition with apparent *MERVL::*RFP expression had some interesting differences from ESCs in gene expression, but at the regulatory level they near-completely collapsed to the same cluster. We speculate that the MERVL reporter may have significant utility in cell reprogramming that had numerous ZGA related genes and we wonder if our ESC data challenges the reporter's value in 2iLIF.

Cell Reprogramming

In general, the reprogramming cells grossly lost their donor cell state and took on the programs of early

BC-like cell lineages which reflected prior observations with high detail. The XGFP+ iEPI population was previously shown as readily potentiated for high contribution in chimeric embryos (Kime et al., 2016). Although the TE-like and PE-like regions could be identified and had convincing cells therein, we wonder if such cells could seed trophoblast stem cell or XEN cell cultures if transferred to appropriate conditions. In the BCLH SES, the TE-like region generally had low *Cdx2* expression 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 blastocyst formation (Meissner and Jaenisch, 2006), its role in implantation is important and distinct CDX2+ TE is roundly regarded (Strumpf et al., 2005). The PE-like region was less distinguished and had significant overlap between 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. Traces of gene expression throughout this study suggested that iPE and PE-like region cells shared a pluripotent-like origin with the iEPI population reminiscent to the GATA4+, GATA6+, PDGFRA+ positive cells arising at the inner-face of BCLH pluripotent cells (Kime et al., 2019). Consideration for PE cells has weighed heavily on various SESs and we suspect that correct hypoblast formation will remain a hinge point for healthy embryoid development.

AUTHOR CONTRIBUTIONS

Conceptualization, C.K.; Methodology, C.K., K.T.; Experimentation, C.K., H.H., H.S., Y.S.; Formal Analysis, C.K., K.T.; Investigation, C.K.; Resources, C.K., M.T., Y.S., K.T.; Writing – Original Draft, C.K.; Writing – Revision & Editing, C.K., K.T.; Visualization, C.K.; Project Supervision, C.K.; Bioinformatics: C.K.; Project Administration and Funding: C.K.

ACKNOWLEDGEMENTS

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 a HPC for Cell Ranger processing. We thank Eszter Posfai of Princeton University, and both Vincent Pasque and Adrian Janiszewski of Katholieke Universiteit (KU) Leuven for providing important control data and thoughtful discussions.

Funding Sources

This study was primarily funded by the RIKEN Center for Biosystems Dynamics Research Organoid Project.

LEGENDS

Figure 1: Mouse ESCs and MERVL Reporter Expression

A) A duality of ESC culture in 2iLIF viewed with brightfield imaging (top) and *MERVL::*RFP expression (bottom). *Scale bars = 200μm.*

B) UMAP based gene expression feature plot for transgenic *MERVL::*RFP.

C) UMAP plot with newly labeled ESC and ESC2CL populations.

D) UMAP based gene expression feature plots for pluripotency genes.

Figure 2: BCLH Cell Induction and Regional Gene Expression

A) EPISC culture with *MERVL::*RFP and XGFP reporters viewed with brightfield imaging (top) *MERVL::*RFP expression (mid), and XGFP expression (bot). *Scale bars = 200μm.*

B) BCLH reprogramming from EPISC with *MERVL::*RFP and XGFP reporters imaged on Day4, Day5, and Day7, for brightfield (top), *MERVL::*RFP expression (mid), and XGFP expression (bot). *Scale bars = 100μm/Day4, and 200μm/Day5:Day7.*

C) UMAP plot clustering of EPISC, ESC, ESC2CL, Day5, and Day7 samples.

D) Violin plot of transgenic *MERVL::*RFP reporter expression in EPISC, ESC, ESC2CL, Day5, and Day7 samples.

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

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

H,I,J) Multiple colored 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.

Figure 3: BCLH Three Region RNA Velocity and Gene Regulation

A) UMAP based RNA Velocity plot for EPISC, ESC, ESC2CL, Day5, and Day7 samples with BC-like regions labeled based on regional gene expression (Figure 2).

B) UMAP based RNA splicing plots for spliced (s), unspliced (u) reads, with Cell Type coloring (Figure 3A) 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 (Figure 3D) and SCENIC regulon analysis (Figure 3E,F, Figure S3D).

D) Heatmap(DoHeatmap) 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 regulon activity (red scale).

F) Heatmap(pheatmap) of the row-matched transcription factor average expression (log-transformed) for the regulons of Figure 3E.

Figure 4: Select BCLH SES Cells Reprogram Meaningfully Close to Model Cells

A) UMAP based plot for EPISC, ESC, ESC2CL, iEPI, iPE, iTE, PE, and TE samples.

B) SCENIC Binarized regulons with heatmap(pheatmap) regulon activity (red scale).

C) Heatmap(pheatmap) of the row-matched transcription factor average expression (log-transformed) for the regulons of Figure 4B.

D) Pluripotent, PE, and TE, combined Top 10 markers based on Binarized Regulons.

E) SCENIC tSNE plot based on regulon activity for each sample.

F) Average Regulon activity in RGB color for pluripotency regulons (red), TE regulons (green), and PE regulons (blue) across SCENIC tSNE plot (Figure 4E).

Figure S1: Related to Figure 1

A) Seurat 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 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.*

K) 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 germ 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 cells 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 cells 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 cells and control cells.

C) SCENIC total AUC regulon activity for select iTE, iPE, and iEPI, and control TE, PE, ESC/ESC2CL, and EPISC.

Table S1: ESC Markers

ESC markers discovered from FindMarkers in Seurat.

Table S2: ESC2CL Markers

ESC2CL markers discovered from FindMarkers in Seurat.

EXPERIMENTAL PROCEDURES

EPISC Culture and BCLH Reprogramming

EPISCs culture and reprogramming was performed as described previously (Kime et al., 2016, 2019), and reprogramming included Sodium Pyruvate.

ESC Culture and MERVL Reporter Integration

BL6 ESCs were converted from 3iLIF conditions and cultured in 2iLIF conditions with CTSFES basal media as described previously (Kime et al., 2019) on iMatrix511 coated 6-well plates. Cells were integrated and selected for *MERVL::*RFP reporters (mCherry and D2nmCherry) in piggyback vectors the same as with EPISC in our previous study (Kime et al., 2019). After several passages, the two different populations of cells became obvious and stabilized.

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 Chromium Single Cell 3' Library & Gel Bead Kit V3.0. Sample libraries were finalized and sequenced on one Hiseq X lane (150bp PE; Macrogen) for each. 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 match loom data counts for TE and PE control cells (Posfai et al., 2020), we prepared our data tables similarly with DESeq2 size factor normalization prior to merging samples. The iEPI, iTE, and iPE cells were selected 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)

Microscopy

Brightfield and live cell RFP and GFP fluorescence was imaged with a Olympus IX71 Microscope.

REFERENCES

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

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, 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-Herraez, I., Santos, F., Kubinyecz, 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.

Benchetrit, H., Jaber, M., Zayat, V., Sebban, S., Pushett, A., Makedonski, 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.

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.

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.

Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W.B., Pasque, V., Magnúsdóttir, 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.

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. Dev. Camb. Engl. *144*, 876–888.

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 Rep. *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., Derynck, R., Asahi, M., Panning, B., Yamanaka, S., and Tomoda, K. (2016). Autotaxin-mediated lipid signaling intersects with LIF and BMP signaling to promote the naive pluripotency transcription factor program. Proc. Natl. Acad. Sci. *113*, 12478–12483.

Kime, C., Kiyonari, H., Ohtsuka, S., Kohbayashi, E., Asahi, M., Yamanaka, S., Takahashi, M., and Tomoda, K. (2018). Implantation-Competent Blastocyst-Like Structures from Mouse Pluripotent Stem Cells. BioRxiv 309542.

Kime, C., Kiyonari, H., Ohtsuka, 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 Rep. *0*.

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. *114*, E9579–E9588.

Kubaczka, C., Senner, C.E., Cierlitza, M., Araúzo-Bravo, M.J., Kuckenberg, P., Peitz, M., Hemberger, M., and Schorle, H. (2015). Direct Induction of Trophoblast Stem Cells from Murine Fibroblasts. Cell Stem Cell *17*, 557–568.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

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 1–6.

Lo Nigro, A., de Jaime-Soguero, A., Khoueiry, R., Cho, D.S., Ferlazzo, G.M., Perini, I., Abon Escalona, V., Aranguren, X.L., Chuva 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 Rep. *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 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., Halbisen, M.A., Wang, K., Latham, K., and Ralston, A. (2016). OSKM Induce Extraembryonic Endoderm Stem Cells in Parallel to Induced Pluripotent Stem Cells. Stem Cell Rep. *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., Pardon, T., Bakkali, M.E., Talon, I., Geest, N.D., et al. (2020). Defining totipotency using criteria of increasing stringency. BioRxiv 2020.03.02.972893.

Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast 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., Oudenaarden, A. van, Blitterswijk, C.A. van, and Geijsen, 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., Ohnuki, 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.

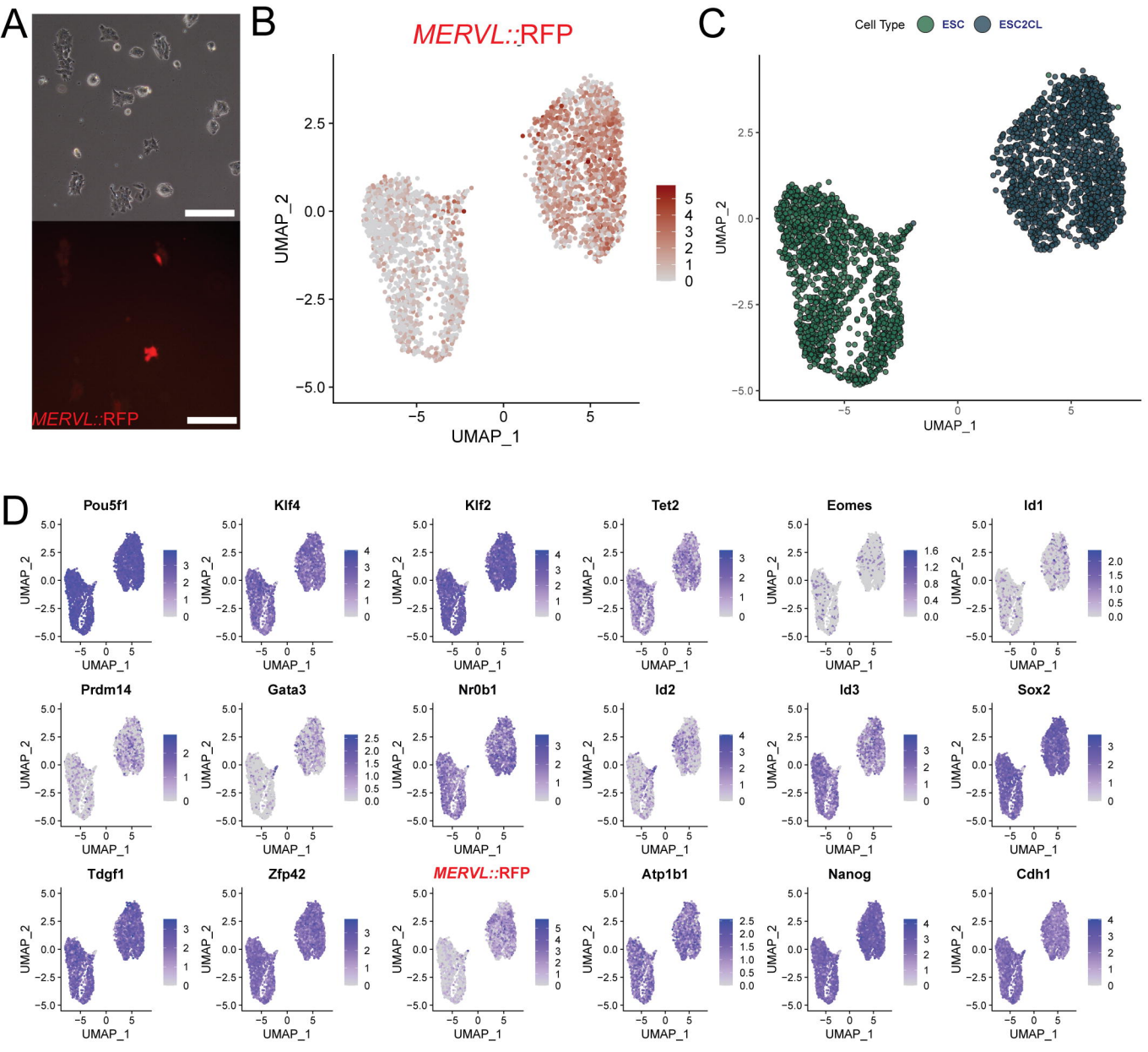
Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat. Rev. Mol. Cell Biol. *17*, 155–169.

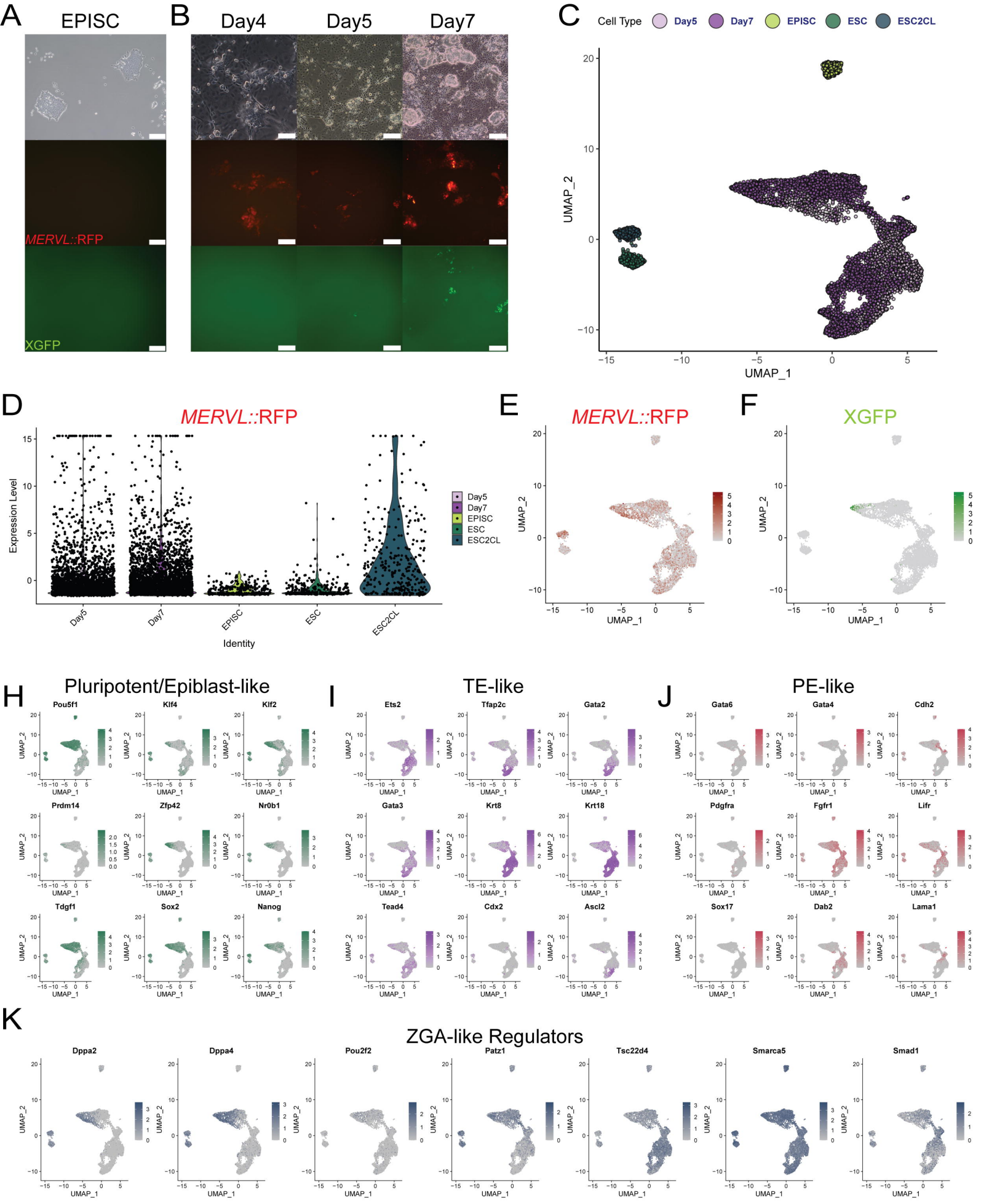
Woogeng, I.N., Abugessaisa, I., Tachibana, A., Sahara, Y., Hon, C.-C., Hasegawa, A., Kaczkowski, B., Sakai, N., Nishida, M., Hu, H., et al. (2020). Inducing Human Retinal Pigment Epithelium-like Cells from Somatic Tissue. BioRxiv 2020.07.27.215103.

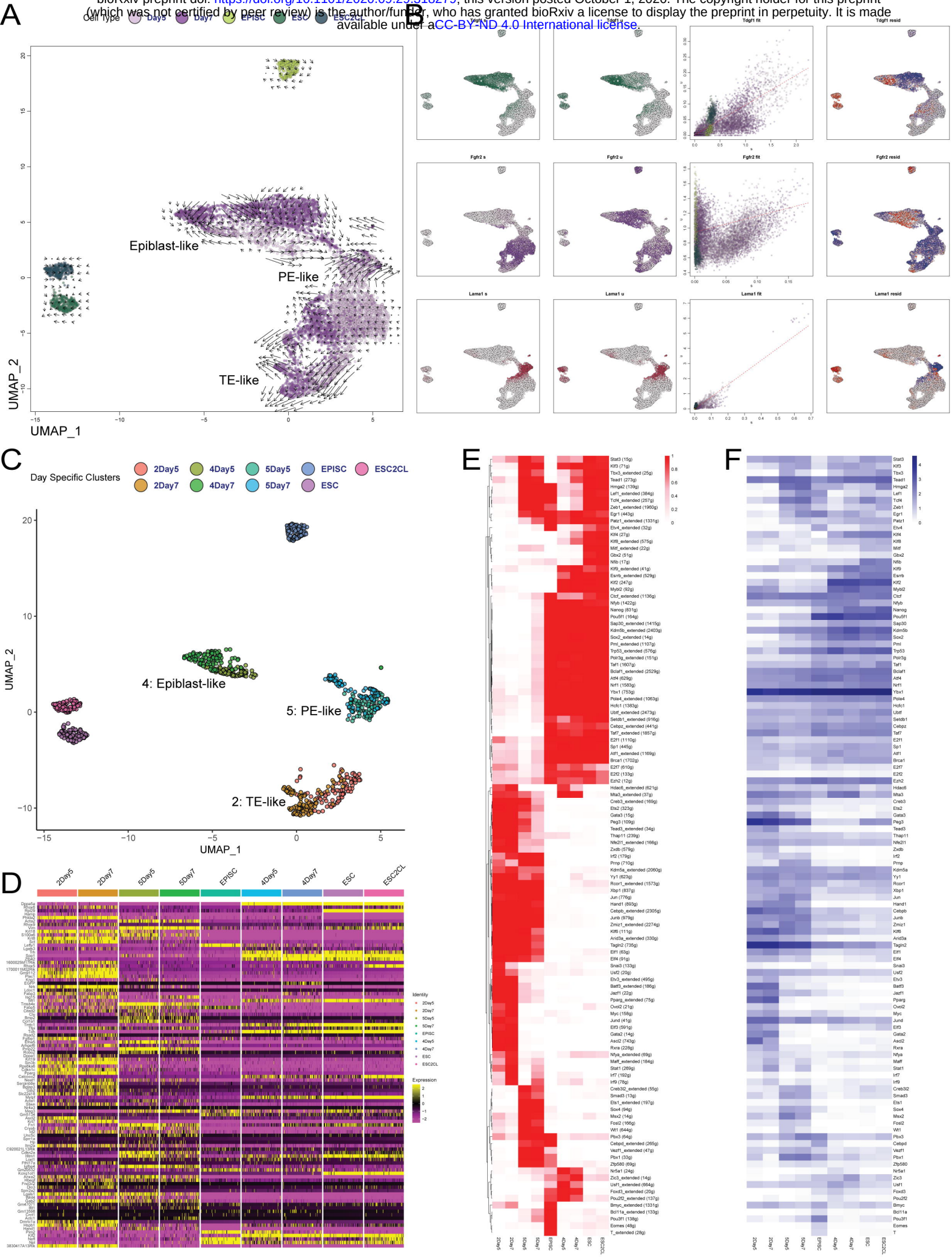
Wu, G., Gentile, L., Fuchikami, T., Sutter, J., Psathaki, K., Esteves, T.C., Araúzo-Bravo, M.J., Ortmeier, C., Verberk, G., Abe, K., et al. (2010). Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. Development *137*, 4159–4169.

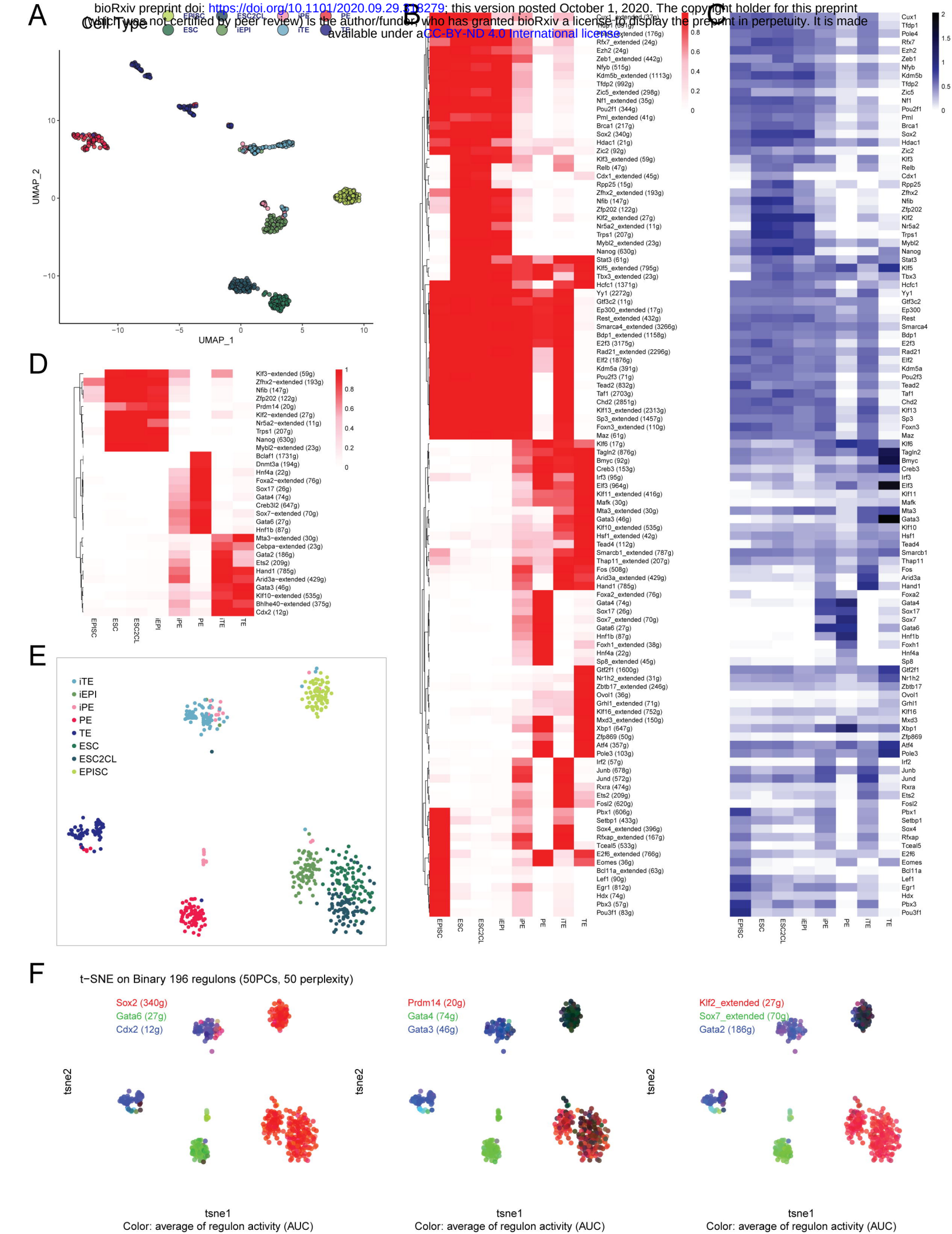
Yang, Y., JIN, Y., MARTYN, A.C., LIN, P., SONG, Y., CHEN, F., HU, L., CUI, C., LI, X., LI, Q., et al. (2013). Expression Pattern Implicates a Potential Role for Luman Recruitment Factor in the Process of Implantation in Uteri and Development of Preimplantation Embryos in Mice. J. Reprod. Dev. *59*, 245–251.

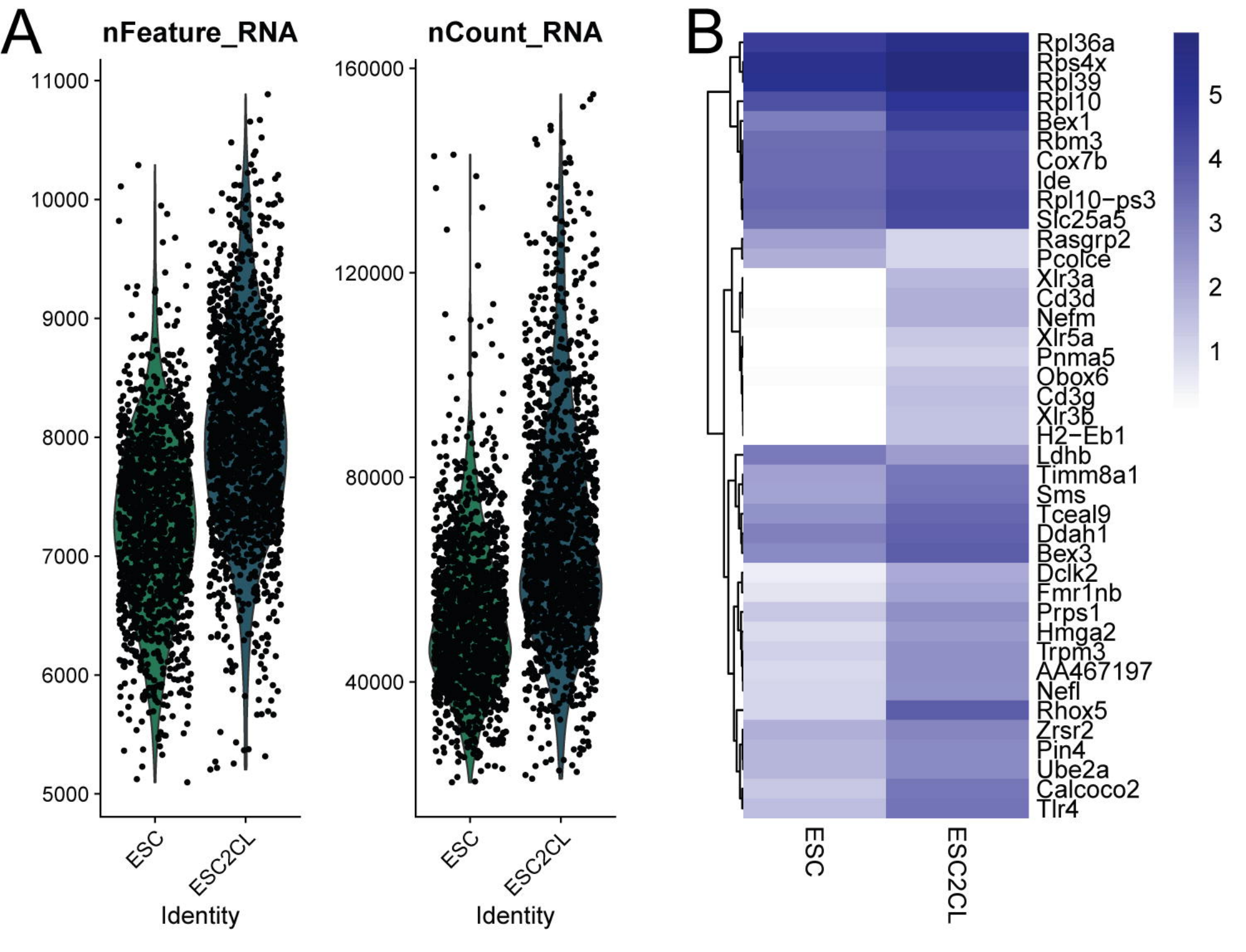
Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M., Gumucio, D.L., et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. Nature *573*, 421–425.

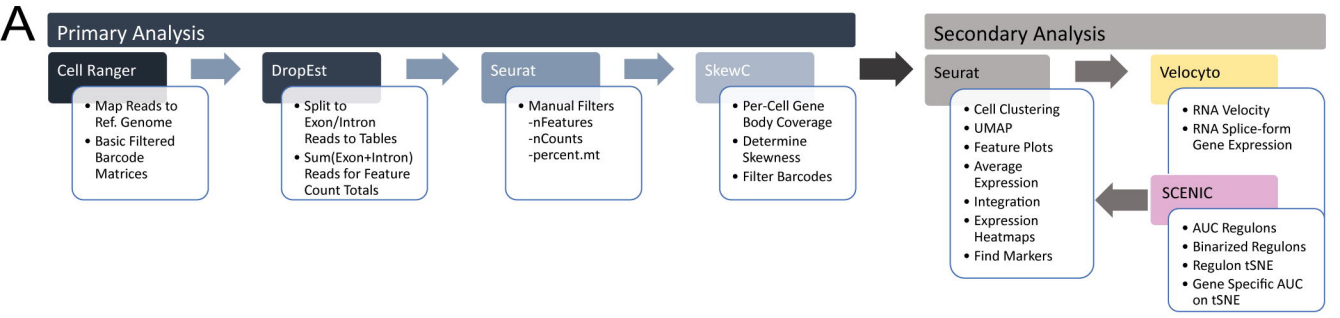

A
MERVL::RFP
B
MERVL::RFP
UMAP_2
UMAP_1
C
Cell Type
ESC
ESC2CL
UMAP_2
UMAP_1
D
Pou5f1
Klf4
Klf2
Tet2
Eomes
Id1
Prdm14
Gata3
Nr0b1
Id2
Id3
Sox2
Tdgf1
Zfp42
MERVL::RFP
Atp1b1
Nanog
Cdh1
UMAP_2
UMAP_1

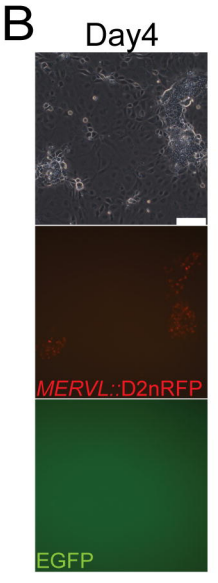

A
EPISC
B
Day4
Day5
Day7
MERVL::RFP
XGFP
C
Cell Type
Day5
Day7
EPISC
ESC
ESC2CL
UMAP_2
UMAP_1
D
MERVL::RFP
Expression Level
Identity
Day5
Day7
EPISC
ESC
ESC2CL
E
MERVL::RFP
F
XGFP
H
Pluripotent/Epiblast-like
Pou5f1
Klf4
Klf2
Prdm14
Zfp42
Nr0b1
Tdgf1
Sox2
Nanog
I
TE-like
Ets2
Tfap2c
Gata2
Gata3
Krt8
Krt18
Tead4
Cdx2
Ascl2
J
PE-like
Gata6
Gata4
Cdh2
Pdgfra
Fgfr1
Lifr
Sox17
Dab2
Lama1
K
ZGA-like Regulators
Dppa2
Dppa4
Pou2f2
Patz1
Tsc22d4
Smarca5
Smad1

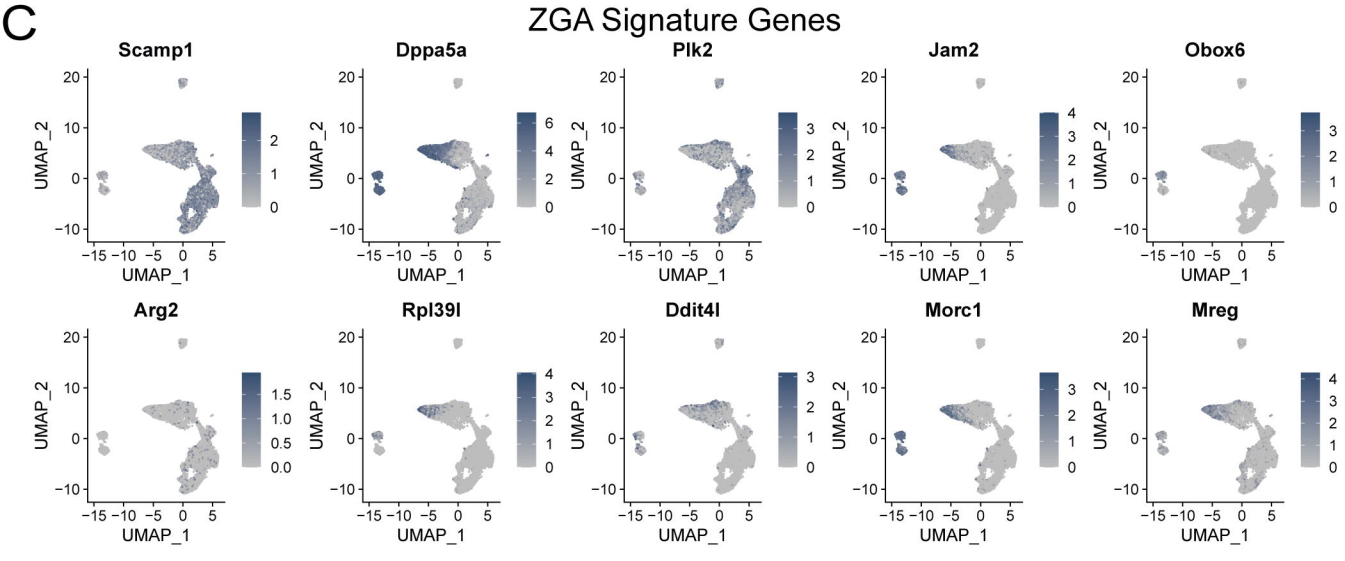


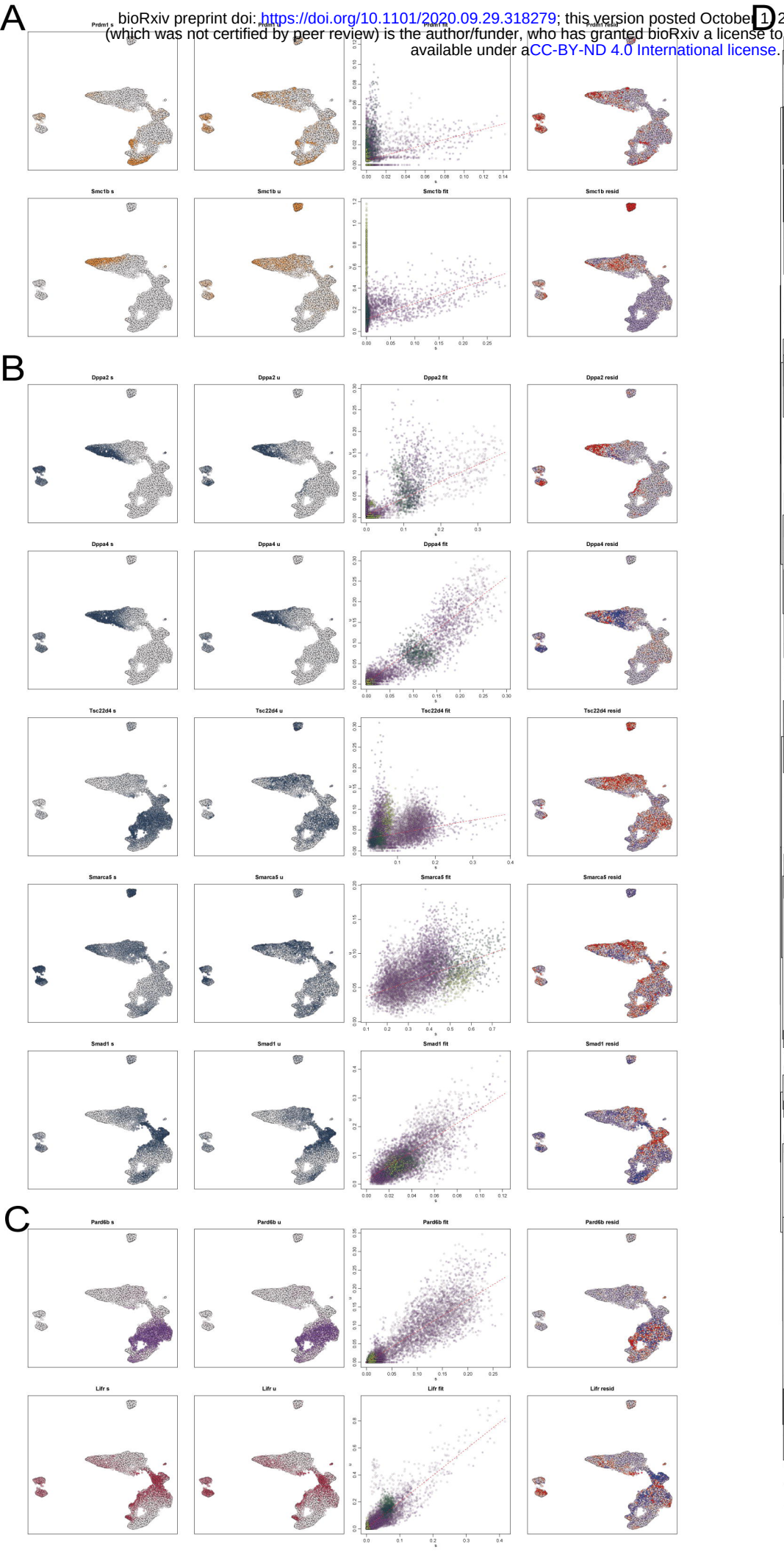


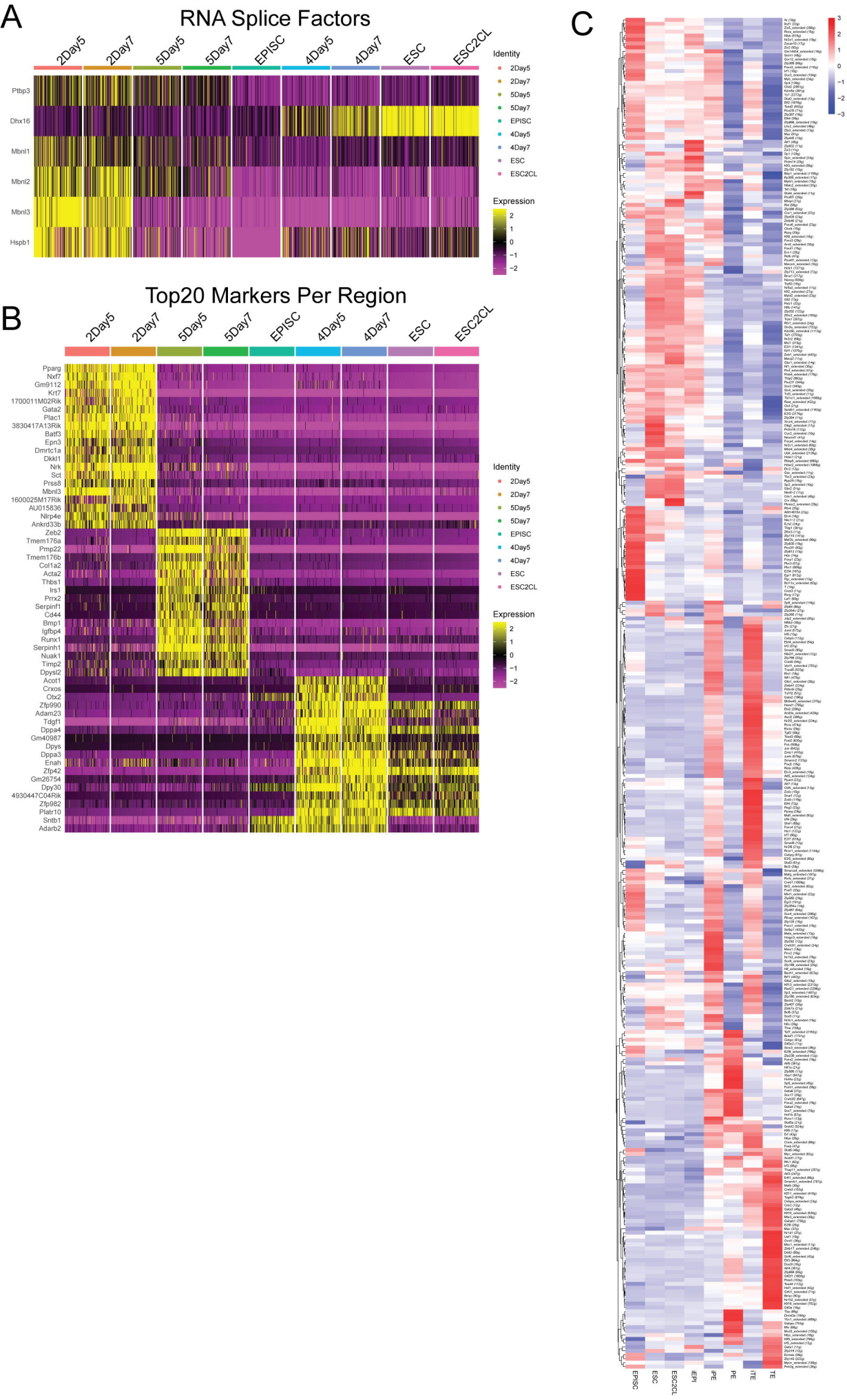

A
nFeature_RNA
11000
10000
9000
8000
7000
6000
5000
ESC
ESC2CL
Identity
nCount_RNA
160000
120000
80000
40000
ESC
ESC2CL
Identity
B
Rpl36a
Rps4x
Rpl39
Rpl10
Bex1
Rbm3
Cox7b
Ide
Rpl10−ps3
Slc25a5
Rasgrp2
Pcolce
Xlr3a
Cd3d
Nefm
Xlr5a
Pnma5
Obox6
Cd3g
Xlr3b
H2−Eb1
Ldhb
Timm8a1
Sms
Tceal9
Ddah1
Bex3
Dclk2
Fmr1nb
Prps1
Hmga2
Trpm3
AA467197
Nefl
Rhox5
Zrsr2
Pin4
Ube2a
Calcoco2
Tlr4
5
4
3
2
1
ESC
ESC2CL


A
Primary Analysis
Cell Ranger
• Map Reads to Ref. Genome
• Basic Filtered Barcode Matrices
DropEst
• Split to Exon/Intron Reads to Tables
• Sum(Exon+Intron) Reads for Feature Count Totals
Seurat
• Manual Filters
-nFeatures
-nCounts
-percent.mt
SkewC
• Per-Cell Gene Body Coverage
• Determine Skewness
• Filter Barcodes
Secondary Analysis
Seurat
• Cell Clustering
• UMAP
• Feature Plots
• Average Expression
• Integration
• Expression Heatmaps
• Find Markers
Velocyto
• RNA Velocity
• RNA Splice-form Gene Expression
SCENIC
• AUC Regulons
• Binarized Regulons
• Regulon tSNE
• Gene Specific AUC on tSNE


B
Day4
MERVL::D2nRFP
EGFP


C
ZGA Signature Genes
Scamp1
Dppa5a
Plk2
Jam2
Obox6
Arg2
Rpl39l
Ddit4l
Morc1
Mreg
UMAP_1
UMAP_2




A
RNA Splice Factors
2Day5
2Day7
5Day5
5Day7
EPISC
4Day5
4Day7
ESC
ESC2CL
Identity
Ptbp3
Dhx16
Mbnl1
Mbnl2
Mbnl3
Hspb1
Expression
B
Top20 Markers Per Region
Pparg
Nxf7
Gm9112
Krt7
1700011M02Rik
Gata2
Plac1
3830417A13Rik
Batf3
Epn3
Dmrtc1a
Dkkl1
Nrk
Sct
Prss8
Mbnl3
1600025M17Rik
AU015836
Nlrp4e
Ankrd33b
Zeb2
Tmem176a
Pmp22
Tmem176b
Col1a2
Acta2
Thbs1
Irs1
Prrx2
Serpinf1
Cd44
Bmp1
Igfbp4
Runx1
Serpinh1
Nuak1
Timp2
Dpysl2
Acot1
Crxos
Otx2
Zfp990
Adam23
Tdgf1
Dppa4
Gm40987
Dpys
Dppa3
Enah
Zfp42
Gm26754
Dpy30
4930447C04Rik
Zfp982
Platr10
Sntb1
Adarb2
C
EPISC
ESC
ESC2CL
iEPI
iPE
PE
iTE
TE