nature

https://doi.org/10.1038/s41586-023-06604-5

Accelerated Article Preview

Complete human day 14 post-implantation embryo models from naïve ES cells

Received: 11 April 2023

Accepted: 04 September 2023

Accelerated Article Preview

Cite this article as: Oldak, B. et al. Complete human day 14 post-implantation embryo models from naïve ES cells. *Nature* https://doi.org/10.1038/s41586-023-06604-5 (2023)

Bernardo Oldak, Emilie Wildschutz, Vladyslav Bondarenko, Mehmet-Yunus Comar, Cheng Zhao, Alejandro Aguilera-Castrejon, Shadi Tarazi, Sergey Viukov, Thi Xuan Ai Pham, Shahd Ashouokhi, Dmitry Lokshtanov, Francesco Roncato, Eitan Ariel, Max Rose, Nir Livnat, Tom Shani, Carine Joubran, Roni Cohen, Yoseph Addadi, Muriel Chemla, Merav Kedmi, Hadas Keren-Shaul, Vincent Pasque, Sophie Petropoulos, Fredrik Lanner, Noa Novershtern & Jacob H. Hanna

Complete human day 14 post-implantation embryo models from naïve ES cells

Bernardo Oldak[1,8], Emilie Wildschutz[1,8], Vladyslav Bondarenko[1,8], Mehmet-Yunus Comar[1], Cheng Zhao[2,3], Alejandro Aguilera-Castrejon[1], Shadi Tarazi[1], Sergey Viukov[1], Thi Xuan Ai Pham[4], Shahd Ashouokhi[1], Dmitry Lokshtanov[1], Francesco Roncato[1], Eitan Ariel[1], Max Rose[1], Nir Livnat[1], Tom Shani[1], Carine Joubran[1], Roni Cohen[1], Yoseph Addadi[5], Muriel Chemla[5], Merav Kedmi[5], Hadas Keren-Shaul[5], Vincent Pasque[4], Sophie Petropoulos[2,3,6,7], Fredrik Lanner[2,3], Noa Novershtern[1] & Jacob H. Hanna[1*]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

[2]Department of Clinical Sciences, Intervention and Technology, Ming Wai Lau Center for Reparative Medicine - Stockholm node, Karolinska Institutet, Stockholm, Sweden.

[3]Division of Obstetrics and Gynecology, Karolinska Universitetssjukhuset, Stockholm, Sweden.

[4]Department of Development and Regeneration, Leuven Stem Cell Institute, Leuven Institute for Single-cell Omics (LISCO), KU Leuven-University of Leuven, 3000 Leuven, Belgium

[5]Department of Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel

[6]Département de Médecine, Université de Montréal, Montréal Canada

[7]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Axe Immunopathologie, Montréal, Canada

[8]These authors contributed equally.

*Corresponding Author: Jacob H. Hanna (jacob.hanna@weizmann.ac.il).

Abstract

The ability to study human post-implantation development remains limited due to ethical and technical challenges associated with intrauterine development after implantation[1]. Embryo-like models with spatially organized morphogenesis of all defining embryonic and extra-embryonic tissues of the post-implantation human conceptus (i.e., embryonic disk, bilaminar disk, yolk- and chorionic sacs, surrounding trophoblasts) remain lacking[2]. Mouse naïve embryonic stem cells (ESCs) have recently been shown to give rise to embryonic and extra-embryonic stem cells capable of self-assembling into post-gastrulation mouse Structured Stem cell-based Embryo Models with spatially organized morphogenesis (SEMs)[3]. Here, we extend these findings to humans, while using only genetically unmodified human naïve ESCs (in HENSM conditions)[4]. Such human fully integrated SEMs recapitulate the organization of nearly all known lineages and compartments of post-implantation human embryos including epiblast, hypoblast, extra-embryonic mesoderm, and trophoblast surrounding the latter layers. These human complete SEMs demonstrated developmental growth dynamics that resemble key hallmarks of post-implantation stage embryogenesis up to 13-14 days post-fertilization (dpf) (Carnegie stage 6a). This includes embryonic disk and bilaminar disk formation, epiblast lumenogenesis, polarized amniogenesis, anterior-posterior symmetry breaking, PGC specification, polarized yolk sac with visceral and parietal endoderm, extra-embryonic mesoderm expansion that defines a chorionic cavity and a connecting stalk, a trophoblast surrounding compartment demonstrating syncytium and lacunae formation. This SEM platform may enable the experimental interrogation of previously inaccessible windows of human early post-implantation up to peri-gastrulation development.

Implantation of the human embryo leads to a number of changes in organization that are essential for gastrulation and future development[1]. Much of this relies on the morphogenesis of the extraembryonic tissues and the effect this has on the organization of embryonic cells. Furthermore, this is a period with high incidence of embryo loss and, for this reason, understanding the events associated with this will benefit our understanding of fertility and developmental defects[5]. However, these studies have ethical and technical challenges. Although it is possible to culture structures derived from human blastocysts, these cultures do not recapitulate the events and organization of the *in vivo* structures[6].

Integrated stem cell-derived embryo models of human post-implantation stages can provide a useful platform to understand these critical stages of development[5]. Defining elementary hallmarks of human integrated post-implantation embryo models must include: (i) continued presence of equivalents of all key cell lineages of the developing early post-implantation embryo (e.g., trophoblast-, primitive endoderm-, extra-embryonic mesoderm (ExEM)- and pluripotent epiblast-like cells) (ii) clear self-organization of fundamental embryonic compartments with adequate morphological and structural organization, as well as proper relative orientation between the latter structures (e.g., embryonic disc-, hypoblast-, bilaminar-disc-, polarized amnion-, polarized yolk sac-, chorionic cavity- and trophoblast-like compartments) (iii) evidence of developmental dynamism relating to ability to progress, in a structurally organized manner, through morphologically characterized developmental milestones of the early post-implantation human embryo, following initial aggregate formation[3].

Recently, mouse naïve ESCs were shown to possess the ability to be coaxed *ex utero* into post-gastrulation stage **S**tructured **S**tem cell-based **E**mbryo **M**odels with spatially organized morphology (previously termed as sEmbryos or stem cell-derived synthetic whole embryo models (SWEMs) are renamed herein as **SEM**s)[3]. These structures result from the co-aggregation of non-transduced naïve ESCs (which form the embryo proper) with naïve ESCs transiently expressing the transcription factors Cdx2 or Gata4, to promote their priming towards TE- and PrE-like lineages, respectively. Mouse SEMs developed directly into egg-cylinder shaped SEMs within complex extra-embryonic compartments and can dynamically advance beyond gastrulation and reach early organogenesis stages of development as late as day E8.5[3]. These findings established that mouse naïve pluripotent cells can serve as the sole source of embryonic and extra-embryonic tissues in advanced complete "organ-filled" embryo models[3], and thus their counterpart may enable the generation of integrated SEMs from other mammalian species from which naïven pluripotent stem cells (PSCs - refers to either ESCs or induced pluripotent stem cells (iPSCs)) have been stabilized, including humans[4,7]. Motivated by these observations and following improvements in naïve human PSC conditions[4], we tested whether these cells could be coaxed to form complex peri/post-

implantation embryo-like structures that are able to dynamically advance to pre/peri-gastrulation stages *ex utero*.

ESC priming towards extra-embryonic fate

In mouse, deriving SEMs that contain all embryonic and extraembryonic compartments requires optimal conditions and high-quality rapid priming towards PrE and TE lineages from naïve PSCs, which was achieved by ectopic expression of Gata4 and Cdx2, respectively[3]. Hence, we first set out to establish a similar platform to rapidly and efficiently obtain extraembryonic lineages through transient expression of these transgenes in human PSCs (**Fig. 1a**), while bearing in mind that the early post-implantation pre-gastrulation human, but not mouse, embryo contains an extra embryonic mesoderm compartment (ExEM)[8]. We generated Doxycycline (DOX) inducible human ESCs for GATA4 or GATA6, regulators of PrE and ExEM lineages in humans[9] (**Supplementary Fig. 1a-b**). We used FACS staining PDGFRa, which marks both PrE and ExEM lineages[9], for identifying optimal conditions to rapidly and efficiently induce HENSM naïve ESC priming towards PrE and/or ExEM-like lineages (**Supplementary Fig. 1c-e)**. While Gata4 induction (iGata4) in mouse naïve 2i/LIF conditions yielded dramatic upregulation of Pdgfra$^+$ cell fraction after 48h of DOX (**Extended Data Fig. 1a**), induction of GATA4 and GATA6 expression in human naive ESCs cultured in HENSM resulted <10% PDGFRa$^+$ cells after 6 days (**Fig. 1b, Extended Data Fig. 1b**). Since WNT stimulation by CHIR99021 (CH) has been shown to be a stimulant for mouse and human PrE induction[10], and given that CH is included in mouse, but not in recent human naïve condition versions[4], we considered that HENSM conditions during the induction phase might not be suitable for human cells. Hence, we screened for other conditions to facilitate the induction of PDGFRa+ cells from naive human ESCs (**Fig. 1a**). The mouse PrE-derivation conditions (termed C10F4PDGF)[11], resulted in very low yield of PDGFRa induction (**Fig.1c, Extended Data Fig. 1c**). RACL induction media (**R**PMI based medium supplemented with **A**CTIVIN A, **C**HIR99021 and **L**IF) which has been used to prime human naïve ESCs towards PrE and ExEM state[10], or NACL media (DMEM/F12/Neurobasal **N**2B27-based) that stabilizes naïve Endoderm cells generated in RACL conditions[10], also led to low levels of PDGFRa$^+$ fraction (**Fig. 1c, Extended Data Fig. 1c**).

Since ACTIVIN A inhibits human naïve ESC *in vitro* differentiation into ExEM cells[9], we omitted it from RACL media (termed herein as RCL). Remarkably, RCL media resulted in PDGFRa induction in the majority (>50%) of cells in iGATA4 and iGATA6 cells (**Fig. 1c**). However, high efficiencies of PDGFRa$^+$ cell formation was evident in RCL conditions from isogenic wild type (WT) cells without exogenous expression of GATA4/6 on irradiated mouse embryonic feeder cells (MEFs)(**Fig. 1d**), indicating

that the transient transgene expression is not required for efficient PDGFRa$^+$ induction in human naïve HENSM ESCs. Further optimization showed that 3 days of induction in RCL condition followed by 3 days incubation in basal N2B27 conditions yielded comparable results (**Fig 1d, Extended Data Fig. 1d)**. Notably, incubating naïve ESCs in N2B27 media also yielded PDGFRa$^+$ cells but at significantly 2.5 fold lower levels than RCL media (**Fig 1d, Extended Data Fig. 1c).** As we prefer to use genetically unmodified cells, we focused on using RCL conditions on WT non-transgenic cells for further characterization.

We tested for the existence of PDGFRa$^+$ PrE- and/or ExEM-like cells in RCL induction and to distinguish between them. Both immunostaining and PCR validated endogenous expression of PrE marker genes in RCL conditions, including SOX17 which marks only the PrE fraction, alongside GATA4, GATA6, and NID2 genes that are common between PrE and ExEM (**Extended Data Fig. 1e, Supplementary Fig. 2)**[9]. Markers of definitive endoderm (DE) like GSC or HHEX[10] were not meaningfully induced from naïve ESCs in RCL (**Extended Data Fig. 1e**) thus excluding DE identity. Applying RCL on human isogenic primed ESCs yielded higher definitive endoderm marker expression (GSC or HHEX) (**Supplementary Fig. 1f**)[10]. Immunostaining results for SOX17 and BST2 markers that distinguish between PrE or ExEM lineages[9], respectively, confirmed the co-emergence of PrE- and ExEM-like cells from naïve ESCs under the same RCL induction protocol on MEFs (with or without GATA6 overexpression) (**Fig. 1e, Extended Data Fig. 2a-b**). GATA4 positively marked both SOX17$^+$ PrE- and BST2$^+$ ExEM-like populations as expected (**Extended Data Fig. 2a**)[9]. BST2$^+$ ExEM-like cell identity was validated by upregulation of FOXF1 (**Extended Data Fig. 2c**) which marks ExEM cells, but not PrE cells or residual ESCs. To examine the identity of the starting day 3 RCL cells, we applied scRNA-seq and integrated our data with a reference dataset of a naive to PrE/ExEM cell differentiation[9]. Day 3 RCL cells derived from HENSM naïve ESCs aligned to previously described PrE cells, ExEM cells and some residual pluripotent stem cells (intermediate epiblast cells) (**Extended Data Fig. 1f, Supplementary Fig. 1g, Supplementary Fig. 3**). Thus, we adopted RCL pre-treatment of WT HENSM naïve PSCs for co-aggregation experiments with other lineages.

We set out to find optimal conditions for priming naïve ESCs towards TE lineage cells that can adequately aggregate with the other lineages and develop into post-implantation stage human SEMs. Cdx2 overexpression in mouse naïve ESCs has been shown to be efficient in rapidly generating TSCs that remained viable and correctly integrated upon coaggregation with mouse naïve ESCs and iGata4 cells and generated both chorionic and ectoplacental cone placental lineages[3] as we validated herein (**Extended Data Fig. 3a-b**). Therefore, we generated DOX inducible CDX2 human ESC lines with a constitutively expressed tdTomato marker to track cell viability and integration (**Supplementary Fig. 4a-c)**. Human iCDX2 cells

pre-treated with DOX in either HENSM or different validated human TE/TSC induction conditions, did not meaningfully expand within the aggregates generated with induced PrE/ExEM-like cells and ESCs, likely due to a drastically reduced viability of iCDX2 cells upon DOX treatment (**Supplementary Fig. 4d-f**). We also tested tdTomato labelled TSC lines derived from both naïve and primed ESCs (**Supplementary Fig. 5a**)[12]. In all tested aggregation conditions with either primed (pTSC) (**Extended Data Fig. 3c; Supplementary Fig. 5**) or naïve ESC derived TSC lines (nTSC) (S**upplementary Fig. 6a**), the TSCs did not generate an outer layer surrounding the aggregate, but rather formed focal clumps (**Extended Data Fig. 3c; Supplementary Fig. 5; Supplementary Fig. 6b-d)**. The inability of human TSCs to integrate in putative SEMs, as opposed to mouse TSCs, might stem from the fact that mouse TSC lines are $Cdx2^{+}$ and correspond to equivalent earlier stages of trophoblast development, than those TSCs isolated from human ESCs which are $CDX2^{-}$ [13,14].

We also tested induced GATA3 (iGATA3) ESCs. PCR analysis showed higher expression levels of CDX2 and TACSTD2 in the non-transduced group under BAP(J) conditions[14] (**Supplementary Fig. 7**). Although GATA3 overexpression induced endogenous expression of GATA2 TE marker, the cells did not uniformly express TFAP2C, CDX2, and CK7, while in the absence of transgene overexpression, we observed higher and uniform expression of TFAP2C, CDX2, and higher occurrence of CK7 positive cells under BAP(J) conditions (**Extended Data Fig. 3d**). Flow cytometry analysis for TACSTD2 (marker of early and late TE) and ENPEP (expressed only in the late TE)[14] showed the highest percentage of a double positive population in WT cells under BAP(J) whose TE-like identity was validated by scRNA-seq (**Fig. 1f; Extended Data Fig. 3e, f, g**)[14].

Notably, following the induction and aggregation with naïve ESCs and PrE/ExEM-like cells, iGATA3 cells remained viable but did not surround the aggregates (**Fig. 1g**). In contrast, TE-like cells derived from genetically unmodified naïve ESCs under the same BAP(J) protocol, uniformly surrounded the aggregates (**Fig. 1g**), which is a decisive criterion expected to be fulfilled in integrated SEM as the crosstalk of the TE lineage with the rest of the embryo and its role on proper morphogenesis continues to be an open question in human development. The ability to derive relevant extra-embryonic lineages from genetically unmodified WT human naive ESCs without the need for transgene overexpression are in line with the recent studies demonstrating that human naïve pluripotent cells can be more easily coaxed to give rise to early progenitors of PrE-, amnion-, ExEM- and TE-like cells when compared to mouse naïve ESCs, which require ectopic transcription factor overexpression[4,9,14,15]. The latter is consistent with our observation that enhancers of key TE and PrE regulators (GATA3, GATA6, GATA4) are accessible in human, but not in mouse naïve ESCs while being transcriptionally inactive in both (**Extended Data Fig.**

4a). The latter might render human naïve ESCs to be relatively more responsive to mere addition or omission of signaling cues to activate endogenous GATA3, GATA6, and GATA4 genes without the additional obligatory need for their ectopic expression, to nudge naïve ESC cell fate towards extraembryonic lineages.

To evaluate the projected contribution of the three induced populations co-aggregated at day 0, we interchangeably omitted each of the input cell fractions (HENSM, BAP(J) and RCL-induced cells) (**Extended Data Fig. 5a**). Omitting HENSM cells completely abolished $OCT4^+$ epiblast-like formation, while $CK7^+$ Tb-like cells formed and surrounded the aggregates that still contained disorganized $SOX17^+$ cells (**Extended Data Fig. 5b**), suggests that only HENSM cells can give rise to the epiblast-like compartment. Omitting day 3 BAP(J)-induced TE-like cells abolished the formation of the Tb-like compartment (**Extended Data Fig. 5b**), demonstrating that the Tb-like compartment is derived only from the induced BAP(J) cells. In addition to the FACS analysis that showed that the naïve HENSM cells give rise to ~25% $PDGFRa^+$ cells when plated in N2B27 basal conditions (vs. ~65% from RCL-induced cells), we validated with immunostaining the emergence of $SOX17^+$ PrE- and $BST2^+$/$FOXF1^+$ ExEM-like cells in N2B27 basal conditions (**Extended Data Fig. 4b**). Consistent with the outcome of FACS and immunofluorescence analyses, we could still observe proper formation of SEMs with the yolk-sac and epiblast after omitting the RCL-induced cell fraction, albeit with arguably reduced YS-like morphological quality and a trend for lower efficiency and (**Extended Data Fig. 5c**). The latter is consistent with the significantly lower yield (2.5 fold decrease) of PDGFRa+ from WT naïve ESCs when placed in N2B27 vs RCL pretreatment (**Extended Data Fig. 2c**). Attempting to generate SEMs by aggregating only HENSM naïve ESCs, did not yield any organized SEM structures but rather disorganized embryonic body (EB)-like structures with dispersed $OCT4^+$ ESCs, $GATA6^+$ PrE/EXEM-like cells, and nearly undetectable $GATA3^+$ TE-like fraction (**Extended Data Fig. 5d**). This indicates insufficiency of naïve ESCs induced in basal conditions to give rise to extra-embryonic-like cells in an optimal frequency for self-organizing into complete embryo-like structures. As we are seeking highest efficiency possible, we continued with including RCL-primed cells in our SEM aggregation regimen.

Up to Day 14 Human SEMs from HENSM ESCs

We proceeded to test the capacity to form embryo-like structures solely from naïve PSCs as a starting population, that could mimic different stages of natural human *in utero* development (**Fig. 2a)**. We calibrated the aggregation conditions, such as cell numbers needed, ratios within cell mixtures, and media compositions for different stages (**Supplementary Fig. 8,9**). The protocol starting with 120 cells per individual aggregate at the ratio 1:1:3 (nPSC: PrE/ExEM-like: TE-like) in basal N2B27 conditions

supplemented with BSA (which was found critical to reduce human aggregate stickiness) for three days (**Supplementary Fig. 10a**), resulted in optimal aggregation, validated by the presence of epiblast- and extra-embryonic-like lineages via immunostaining (**Extended Data Fig. 6a**). To support growth of the SEM and prevent TE-like cells attachment to the plate surface, which disrupts the morphology after day 3, we continued our culture using orbital shaking conditions (**Supplementary Fig. 10b-e**)[3]. The composition of the human Ex-Utero Culture Media 2 (hEUCM2) was adapted from the mouse SEMs protocol[3], while increasing FBS concentration was found optimal for human SEM structural organization (**Fig. 2b, Extended Data Fig. 6a-b, Supplementary Fig. 10-11**).

Throughout the study we focused on and interchangeably generated SEMs from two human ESC lines: WIBR3 (46XX) and WIBR1 (46XY) and analyzed multiple aspects of SEM structure that yielded equivalent SEM structure and morphogenesis results at slightly different efficiency per line (**Fig. 2c-e; Extended Data Fig. 6**). During eight days of *ex utero* culture, the aggregates grew extensively, forming a 3D spherical structure with evident tissue compartments self-organization and inner cavities formation (**Fig. 2c**, **Supplementary Videos 1-3**). Human SEMs did not only express the respective lineage markers, but also established the structures morphologically characteristic of *in utero* implanted embryos (**Fig. 2d, f**). From the beginning of the *ex utero* culture (day 3-4), SEMs became enclosed by the Tb-like compartment, marked by GATA3, CK7, and SDC1 that marks syncytiotrophoblast (STb) (Days 3-4; **Fig. 2d; Extended Data Fig. 6**), which is reminiscent of human early *in utero* development by 8 dpf or Carnegie Stage 5 (CS5) (**Fig. 2a**). At this stage, the implanting embryo starts to become surrounded by a layer of STb, which directly invades maternal endometrium and supports future histotrophic nutrition *in utero*. Notably, the co-aggregation protocol devised herein does not result in a blastocoel cavity formation nor in a condensed ICM-like structure, as opposed to blastoid models, indicating that SEMs do not go through a blastocyst-like stage (**Supplementary Fig. 12**). Furthermore, human blastoids did not develop further when cultivated under suspended SEM culture conditions from day 3 onwards **(Supplementary Fig. 13)**.

In human and non-human primates (NHP), the inner cell mass segregates into two lineages, the epiblast (Epi), formed by columnar epithelium expressing OCT4[16,17], and the hypoblast located underneath the Epi (Hyp, comprised of cuboidal cells expressing SOX17, GATA6, GATA4, PDGFRa) (CS5a, **Fig. 2a**). Importantly, in human SEMs, Epi- and PrE-like cells segregated into two distinct compartments, differentially expressing the respective lineage marker genes (OCT4 and SOX17) (**Fig. 2d**). Both Epi- and Hyp-like compartments were surrounded by the Tb-like compartment, marked by CK7 starting from day 3 (**Fig. 2d, Extended Data Fig. 6a**), and SEMs advanced morphologically by day 6 (**Extended Data Fig. 6b**). The Epi-like compartment initiated formation of the amniotic-like cavity (AC), whilst the hypoblast-

like layer formed a yolk sac (YS)-like cavity, establishing a bilaminar disk structure in between (**Fig. 2d, Extended Data Fig. 6b**) reminiscent of the 9-10 dpf human embryo (CS5b, **Fig. 2a**). The efficiency of correctly organized post-implantation human SEM at day 6 was estimated to be 1.64% for WIBR3 and 1.09% for WIBR1 of all starting aggregates at day 0 as judged by co-immunofluorescence for lineage markers and morphology criteria (**Fig. 2e, Extended Data Fig. 6c**). We noted that the human SEMs show a notable degree of asynchrony within each cell line and within individual experimental batch, with up to 2 days difference in developmental staging for SEMs found at days 6-8, leading to some more advanced and some earlier structures when evaluating SEMs at the same specific time point. Starting the human SEM protocol with human primed, rather than naïve, ESCs did not generate equivalent SEMs (**Extended Data Fig. 5e**), as also seen in mice[3].

The early post-implantation pre-gastrulating human embryo already contains ExEM[18] , contributes to the remodeling of the chorionic cavity (ChC) resulting in the formation of the connecting stalk and participates in formation of blood and placental vasculature by filling the chorionic villi[19]. ExEM tissue becomes abundant between the primary yolk sac (PYS) and the Tb by CS5a, forming a ChC underneath the PYS by CS5c (11 dpf, **Fig. 2d**). The latter was observed upon close examination of day 6-8 SEMs, revealing the presence of the cavity formed by an additional tissue layer between the YS- and the Tb- like compartments, that corresponds to ChC-like structure, consistent with what has been characterized in natural human embryos corresponding to these stages (**Fig. 2c-d**). Later, the ExEM expands allowing the remodeling of the PYS into the secondary yolk sac (SYS) and formation of a connecting stalk (Sk), the structure that crosses through the ChC and holds the bilaminar disc to the chorion, later contributing to the umbilical cord[19]. In human day 8 SEMs, we also observed 3D expansion of all the above mentioned lumina-like structures and growth of the extraembryonic-like tissues **(**0.42% efficiency out of all starting day 0 aggregates **– Extended Data Fig. 7a)**, suggesting differentiation and remodeling of the PYS- into SYS-like compartment alongside ExEM-like compartment expansion, and formation of Sk-like structure (**Fig. 2c-d**; **Fig. 3f**, **Extended Data Fig. 7; Supplementary Video 1**).

By 11-12 dpf (CS5c) in human embryos, the embryonic disk segregates into a ventral pseudostratified epiblast and a dorsal squamous amnion (Am) (**Fig. 2a**). The bilaminar epiblast will give rise to the embryo proper, whereas the amnion will comprise the protective membrane surrounding the fetus until birth. Starting from day 6 until day 8 of the *ex utero* culture, the dorsal segment of the epiblast-like compartment acquired squamous morphology resembling the amnion-like layer, and the bilaminar structure adapted a disk shape (**Fig. 2d, f; Extended Data Fig. 7b; Supplementary Videos 1-3**), resembling a key hallmark of the *in utero* human post-implantation development in preparation for gastrulation. Altogether,

our approach exploits the developmental plasticity of genetically unmodified transgene-free human naïve ESCs, demonstrating their unique capacity to self-assemble into early post-implantation human embryo models that comprise both embryonic and extraembryonic compartments.

SEM with a bilaminar-disk-like structure

We aimed to characterize the development of key lineages in SEMs in more detail. As the epiblast is derived from initially naïve ESCs, we checked whether they undergo priming after co-aggregation with other lineages in aggregation conditions. Indeed, shortly after co-aggregation, we noted the loss of expression of naïve marker genes (DNMT3 and STELLA) that were originally expressed in the starting naïve ESCs in HENSM conditions, and upregulation of the primed pluripotency marker OTX2 by day 3 (**Fig. 3a-b**). Consistent with developmental progression, by day 4, ESCs formed an evident epiblast-like tissue inside the SEM which considerably grew during subsequent development (**Fig. 3c**, **Supplementary Videos 1-4**). The Epi-like compartment in human SEMs showed cell polarization and lumenogenesis (evident by apical Podocalaxyn localization) from day 6, as judged by the apical localization of phosphorylated Erzin, Podocalaxyn (**Fig. 3e**), and aPKC (**Extended Data Fig. 8a**), as well as alignment of the Epi-like cells towards the emerging cavity (**Extended Data Fig. 8b**). The timing of lumenogenesis at day 6 also corresponded to a significant increase in the Epi-like cell number (**Fig. 3d**), in agreement with the histological descriptions of the *in utero* human embryo[20].

Early emergence of the anterior-posterior (AP) axis is prevalent in mammals when a proportion of Epi cells initiates T expression at the prospective posterior side of the embryonic disk. Therefore, we checked for the expression of T in human SEMs and identified a T^+ population of Epi-like cells in the posterior part of the SEM epiblast (**Fig. 3f, Extended Data Fig. 8c**). In parallel to the emergence of the anterior visceral endoderm (AVE)-like compartment, which constitutes the anterior signaling center for epiblast patterning, was seen by the expression of CER1 in the Epi-adjacent part of the VE from day 6 (**Extended Data Fig. 8d**). The vesicular localization of CER1 was evident in human SEMs (**Extended Data Fig. 8d**). Co-immunostaining for these markers further supported establishment of the AP axis and symmetry breaking starting from day 6 with efficiency of 1.02% (**Extended Data Fig. 8e**), where T^+ Epi-like cells could be found in the region opposite to $CER1^+$ AVE-like cells (**Fig. 3f**).

From day 6 on, the Epi-like compartment exhibited patterning of the early amniotic sac-like structure with a dorsal squamous cell population resembling the putative amnion, and ventral columnar pseudostratified epiblast-like cells (**Extended Data Fig. 8f, g**). Immunostaining for TFAP2A and ISL1 in SEMs, revealed their co-localization in multiple squamous dorsal cells, also depleted of SOX2 expression,

confirming their amnion-like identity by localization, morphology, and gene expression (**Fig. 3g; Supplementary Video 5**)[21]. Based on the cell morphology and localization, the polarized amnion can be distinguished from the rest of the Epi between 10 and 12 dpf of human *in utero* development[20]. Moreover, by day 8 in SEM, the Epi-like structure acquired an apparent disk shape (**Fig. 3h**) with an enlarged amniotic-like cavity, while the amnion formed a thinner squamous-shaped epithelium highly resembling *in utero* embryo morphology as documented in Carnegie collection at CS6a (**Fig. 3h – l, Supplementary Video 5**). Lastly, we asked whether advanced *in utero*-like development of the embryonic disk would also lead to induction of early primordial germ cells (PGCs)-like cells, as was seen in mouse SEMs derived from naïve ESCs[3]. Indeed, in some day 8 SEMs, co-immunostaining for several PGC markers identified a population of PGC-like cells positive for OCT4, SOX17, and BLIMP1 (**Fig. 3m; Extended Data Fig. 8h**).

Yolk sac- and ChC-like structures

The yolk sac starts forming from the hypoblast between CS4 and CS5 in humans. The part of the hypoblast located underneath the epiblast (forming together the characteristic bilaminar disc), belongs to the visceral endoderm (VE) and is a dynamic signaling center for epiblast patterning. It is connected to the parietal endoderm (PE), forming the inner cavity, the primary yolk sac (PYS), which becomes reorganized during development[18]. Eventually, the YS serves multiple functions for the growing embryo, supplying nutrients and maintains blood cell progenitors during the embryonic period until the placenta takes over[22]. Formation of the YS cavity-like structure was frequently seen in SEMs with all segregated lineages and became more prominent at day 6 of the *ex utero* development (**Fig. 4**). Once formed, SOX17^{+} YS-like compartment was comprised of the columnar VE-like cells in the proximity of the Epi-like layer, and the squamous PE-like cells lining the opposing side of the cavity (**Fig. 4a – c**), resembling VE and PE cell morphology in PYS of CS5c natural embryos (**Fig. 4c**). The latter was uniformly observed in SEMs with YS-like compartment formation. Both VE- and PE-like cells acquired apicobasal polarity, as judged by the apical localization of aPKC (**Extended Data Fig. 9a; Supplementary Video 6**), which agrees with hypoblast cell morphology in equivalent developmental stage rhesus monkey embryos[23]. Consistent with polarity, cell shape, and localization in natural embryos, SOX17^{+} hypoblast-like compartment always co-expressed primitive endoderm markers GATA6 and GATA4 (**Fig. 4d-e; Extended Data Fig. 9b**). Testing SEM organization was done with SOX17-tdTomato; SOX2-Citrine RUES2 reporter hESC line, that allowed detection of the SOX17^{+} YS-like and the SOX2^{+} Epi-like compartments, confirming formation of structured SEMs from a third independently generated naïve ESC line albeit with a lower (0.08%) efficiency (**Extended Data Fig. 9c**).

We characterized the above-mentioned OCT4$^-$ and SOX17$^-$ cell population beneath the YS-like compartment (**Fig. 2d**). These cells had mesenchymal rather than epithelial morphology, extending cell protrusions towards the surrounding tissues and forming an intermediate mesh-like 3D structure (**Extended Data Fig. 9d**). We then checked the expression of multiple lineage-specific transcription factors, differentially marking ExEM vs. YS during the relevant stages in marmoset embryos[8]. The ExEM-like cells expressed GATA6 and GATA4, but not SOX17 (**Fig. 4e; Extended Data Fig. 9e**), consistent with the ExEM expression profile from marmoset embryos[8] (**Extended Data Fig. 9f**) and human naïve ESC *in vitro* derived ExEM cells[9]. The observed lower GATA6 level in ExEM-like cells compared to PrE-like cells (**Fig. 4e, Extended Data Fig. 9e**), is also in agreement with the expression for those lineages in the *in utero* marmoset embryos[8] (**Extended Data Fig. 9f**). Immunostaining for additional mesenchymal markers, such as BST2[9], VIM[16] and FOXF1, which allow distinguishing ExEM from PrE, further validated the ExEM-like identity of these cells in day 6 – 8 SEMs (**Fig. 4f-g, Extended Data Fig. 9g; Supplementary Video 7**). Hence, we concluded that the inner cavity of the ExEM-like cells represents the ChC-like structure, which is formed by the remodeling of a mesh-like population of mesenchymal cells predominantly visible starting from as early as day 6 in SEMs that corresponds to the human ExEM.

Histological descriptions of human embryos suggest the remodeling of the PYS cavity by the ExEM expansion with a subsequent pinching-off and vesiculation of the PYS remnants, resulting in the formation of the secondary yolk sac (SYS)[18]. In some of the day 8 SEMs, BST2$^+$ and FOXF1$^+$ cells formed a complex filamentous meshwork with multiple cavities, contributing to the SEM complex architecture (**Fig. 4h, i; Supplementary Fig. S14**). Closer examination of these structures revealed clusters of SOX17$^+$ cells entrapped between ExEM-like cells, suggesting residual PYS-like cells after the tissue remodeling (**Fig. 4i-j; Supplementary Fig. S14**). Moreover, an ExEM-like cell population connecting the bilaminar disk to the trophectoderm could be seen in day 8 SEMs, resembling a connecting stalk structure (Sk) (**Fig. 4h, Fig. 2d, Extended Data Fig. 7b**).

Upon deeper analysis of the structure patterns obtained in lineage omission experiments (**Extended Data Fig. 5a-c**) we noted that when BAP(J) cells were omitted from the initial aggregation mixture, this abolished internal structure and cavity formation, including YS-like compartment formation and patterning and the self-organization of the bilaminar disc (**Fig. 4k**), and aberrant expansion of OCT4$^-$/SOX17$^-$/CK7$^-$ population that correspond to ExEM-like cells included in aggregates (**Extended Data Fig. 2)**. More specifically, in the absence of BAP(J) cell input, SOX17$^+$ cells fully surrounded the aggregates, as opposed to control SEMs surrounded by Tb-like compartment that had capability to form a SOX17$^+$ inner YS- and bilaminar disc-like structures (**Fig. 4k-l**). Consistently, in positive control SEM aggregation settings

conducted throughout this study, we could never observe correct YS-like compartment formation in the absence of surrounding Tb-like compartment (**Extended Data Fig. 9h, Fig. 4k-l**). These results provide evidence in human context that the trophoblast compartment functionally influences the organization of the Epi/Hb-like compartments and cavity formation, at least in the context of this specific human embryo model regimen. It will be of future interest to decipher the mechanistic basis for this, which could be signaling based and/or influenced by physical cues dictated by the confining Tb compartment.

Trophoblast-like cell maturation in SEMs

In utero, the human embryo develops surrounded by the trophoblast, which is essential for truly integrated experimental models of early post-implantation development. Immunofluorescence analysis shows that the majority of SEM aggregates are surrounded by trophoblast-like cells with a high efficiency of 48-74%, express multiple trophoblast marker genes, such as GATA3, CK7, and SDC1 (**Fig. 5a, b; Extended Data Fig. 10a, b**)[14]. Marker gene expression and cell morphology further indicated that the outer most trophoblast-like layer is formed by syncytiotrophoblast-like cells, confirming the development of post-implantation trophoblast in human SEMs (**Fig. 5a; Extended Data Fig. 10a-c**). Notably, SDC1 is not expressed on the starting TE-like cells upon BAP(J) induction prior to the aggregation, indicating that maturation of the TE-like cells occurs in the aggregates (**Extended Data Fig. 3d**).

The lacunar phase of trophoblast development begins around 14 dpf, when the fluid-filled spaces from within the trophoblast syncytium merge and partition the trophoblast into trabeculae[24], later contributing to the placental villi. In ~90% of aggregates surrounded by Tb-like compartment, we observed multiple cavities with variable sizes forming inside the syncytiotrophoblast-like layer, which were predominantly located at the SEM periphery (**Fig. 5c-d; Extended Data Fig. 10d; Supplementary Video 8**), which resembled the trophoblast lacunae on the embryo periphery *in utero* at CS5c (**Fig. 5e**). Immunostaining for human chorionic gonadotropin beta (HCGB) demonstrated abundance of the hormone protein in surrounding syncytiotrophoblast-like cells of the SEM, enriched in the intracellular vesicles[14] (**Fig. 5c; Extended Data Fig. 10e-f, i**). Its secretion was confirmed by detection of soluble HCG in media in which SEMs were cultured from day 7 to day 8 (**Extended Data Fig. 10g**).

Examination of the trophoblast syncytium-like layer in SEMs revealed a microvilli-like structures on plasma membrane of syncytiotrophoblast-like cells in all validated SEMs analyzed (**Extended Data Fig. 10h, i**), like the placental syncytium *in utero* [24]. Lastly, we checked whether the syncytium in SEMs is multinuclear, as is also typical during early placental development. Phase-images alongside co-immunostaining of the trophoblast marker SDC1 enabled to see clearer the trophoblast cell shape forming

a layer of thick outer syncytium (**Fig. 5f**). Co-immunostaining with F-ACTIN, that helps define individual cell membrane boundaries, and DAPI, validated that the trophoblast-like cells have multiple nuclei (**Fig. 5g; Extended Data Fig. 10f**).

scRNA-seq analysis validates human SEMs

We performed a single cell transcriptomic analysis by Chromium 10X scRNA-seq on selected SEMs (**Supplementary Fig. 15a-c**). Uniform Manifold Approximation and Projection (UMAP) analysis identified a total of 13 separate cell clusters (**Fig. 6a**). Cluster annotation was performed based on expression (+) or lack-of expression (-) of previously defined lineage-specific markers that allowed us to annotate all 12 identified cell clusters **(Fig. 6 b,c)**: Epiblast-like (Epi – in total four clusters: $OCT4^+$, $SOX2^+$), Yolk-Sac/Hypoblast-like (YS/Hb – in total three clusters: $SOX17^+$, $APOA1^+$, $LINC00261^+$ in addition to $GATA6^+$, $GATA4^+$, $PDGFRa^+$), Extra embryonic mesoderm-like (ExEM – in total four clusters: $FOXF1^+$, VIM^+, $BST2^+$, in addition to $GATA6^+$, $PDGFRa^+$), amnion-like (Am – one cluster: $ISL1^+$, $GABRP^+$, $VTCN1^+$), and Syncytiotrophoblast-like (STb - one cluster: $SDC1^+$, $GATA3^+$, CPM^+) (**Fig. 6a,b**).

From the four annotated Epi-like clusters (all being $OCT4^+SOX2^+$) (**Extended Data Fig. 11a**), we subclassified two of them. The first one, which we termed Posterior epiblast-like cluster (cluster 4), was marked by upregulation of T, co-expression of MIXL1, EOMES, MESP1 and WNT8A **(Extended Data Fig. 11a)**, which are all markers of epithelial to mesenchymal transition (EMT) process, known to accompany upregulation of T during peri-gastrulation in mammalian species. The second Epi-like (cluster 7) we termed as "committed epiblast" as it was marked by upregulation of ZIC2, ZEB2 and VIM lineage commitment marker expression, and absence of NANOG while maintaining OCT4 and SOX2 pluripotency markers **(Extended Data Fig. 11a)**. Pseudotime analysis on epiblast-like cells (**Extended Data Fig. 11b-d)**, showed a progression of the transcriptional profile, starting with the unpatterned epiblast-like cells, progressing through two trajectories towards either committed epiblast- or posterior epiblast- like cells, consistent with developmental progression in NHPs[8].

PGC-like cells (co-expressing $POU5F1/OCT4^+PRDM1^+SOX17^+$ human definitive PGC marker combination[25]) (n=27) could be also identified in SEMs within epiblast clusters, most of them (n=19) in cluster 4 **(Fig. 6d)**. Rare cells co-expressing $CD34^+TAL1^+ERG^+$ cells could also be detected corresponding to early blood progenitors **(Extended Data Fig. 12a-b)**. Although we had technical problems in maintaining high trophoblast viability and recovery following different SEM enzymatic dissociations tested, we still observed a well separated cluster expressing lineage markers of Syncytiotrophoblast-like cells (STb- cluster 12) **(Extended Data Fig. 12c)**. While amnion cells can share certain markers with Tb

(like GATA3 and TFAP2A)[16,26], they formed a separate cluster from STb, and expressed amnion markers like ISL1, GABRP and VTCN1, that are specific to amnion but not to CTb or STb (**Extended Data Fig. 11e**). We note that BMP4 and FURIN are expressed in a fraction of human SEM derived amnion-like cells (**Extended Data Fig. 11f**). While Seurat cluster analysis did not yield a separate cluster containing CTb-like cells (likely due to their low abundance and low capture rate for Tb after SEM enzymatic dissociation), when mapping cytotrophoblast specific markers like PAGE4 and S100P[14] we could identify a subcluster (labeled as 10*) within the amnion-like cluster 10, which was also validated and accurately annotated in integrated human embryonic reference[26], that expressed other specific CTb specific markers like CCKBR, SIGLEC6, OVOL1 and IFI6 (that are neither expressed in amnion or in STb) (**Extended Data Fig. 12d-e**). RT-PCR analysis on day 6 SEMs with and without BAP(J) cell input validated detection of both STb (TP63, TEAD3 and OVOL1)[14] and CTb (CGa, CGb and SDC1) specific markers only when BAP(J) primed cells were coaggregated (**Extended Data Fig.12f**).

Among the three different yolk-sac-like cell (YS) clusters that commonly expressed GATA4, GATA6, PDGFRA and APOA1 (clusters 3,5 and 9), we found that recently identified markers of secondary yolk-sac (SYS) in marmoset[8], TTR, APOB and GSTA1, are expressed only in cluster 9, supporting its subclassification as SYS-like compartment **(Extended Data Fig. 13a, Fig. 6a)**. STC1, LHX1 are absent in SYS-like cluster 9, but not in primary yolk-sac-like clusters 3 and 5, also consistent with findings in marmoset[8]. Lack of uniform HHEX and GSC among most cells in all YS-like compartment clusters, is consistent with primitive, rather than definitive, endoderm identity for these three YS-like related clusters **(Extended Data Fig. 13a)**[10]. The co-expression of DKK1 and LHX1 detected via scRNA-seq alongside CER1 expression among some of the $SOX17^{+}$ cluster 3 YS/Hb-like cells, supports the validity of AV-like cell identity in SEMs (**Extended Data Fig. 13a)**. Pseudotime analysis over yolk-sac-like compartment clusters, showed a progression of the transcriptional profile, starting with yolk-sac and progressing toward secondary yolk-sac-like state (**Extended Data Fig. 13b-d)**. Furthermore, PrE-like cells obtained at day 3 following RCL induction mapped to YS, but not to SYS cluster, further indicating that maturation towards SYS-like identity occurred during SEM *ex utero* culture (**Extended Data Fig. 13e)**. The four identified ExEM-like clusters in SEMs, based on FOXF1 and VIM expression (**Fig. 6b**), aligned in a single cluster together with previously described reference human ExEM cells[9] (**Extended Data Fig. 13f-g)**.

We next conducted a comparison analysis of the transcriptional profile of cells in the human SEMs dataset to an integrated human embryonic reference[26] consisting of 6 human embryonic data sets spanning early zygotes[27], in vitro cultured human blastocysts[28–30], 3D *in vitro* cultured human blastocysts until pre-gastrulation stages[31], and a Carnegie Stage 7 (CS7) 16-19dpf human gastrula[32] (**Fig. 6e**). UMAP projection

confirmed the annotated identity of the SEM clusters and validated their resemblance to the transcriptome and cell type composition of early post-implanted human embryos (**Fig. 6e; Supplementary Fig. S15d-e, Supplementary Fig. 16**). Remarkably, some human SEM derived cells projected on primitive streak cells annotated in the integrated human embryonic reference (**Fig. 6e**). Projection of SEM cells onto the human embryonic reference further highlighted that the transcriptional profile of cells differed from human pre-implantation embryos as expected, and all cells corresponded to a post-implanted lineage (**Fig. 6e**).

Seurat cluster analysis did not resolve a separate distinct CTb-like cell cluster, although detailed marker analysis suggested that a subset of the amnion cluster 10 is CTb-like cells (**Fig. 6a; Extended Data Fig. 12d-f)**. The integrated human embryonic reference resolved the presence of CTb-like cells among the amnion-like cells within the Seurat cluster 10 **(Fig. 6a,e; Supplementary Fig. 15d-e)**. Examining an extended panel of amnion and trophoblast markers[31] further supported this result, establishing final lineage annotation of amnion, CTb and STb cells (**Extended Data Fig. 14; Supplementary Table 1-2**). This analysis validated the detection of post-implantation Epi-, YS-, ExEM-, STb- and CTb-like cells, as well as emergence of other transcriptomic states, such as amnion-, Hb/YS- and posterior Epi-like cells and compartments (**Fig. 6e; Supplementary Fig. 15d-e, Extended Data Fig. 14**). Overall, despite the known limitations of cross-dataset, cross-platform comparisons, the single cell transcriptomic analysis presented above supports the conclusion that human SEMs recapitulate lineage differentiation of the early human post-implantation embryo.

Discussion

Studying early human post-implantation development is crucial for understanding human embryogenesis as well as developmental birth defects and early pregnancy loss. Moreover, optimizing protocols for 2D *in vitro* human PSC direct differentiation into mature cell types, would greatly benefit from understanding the key mechanical, transcriptional, and signaling pathways active during early embryogenesis, in order to improve PSC differentiation quality and efficiency. Such research endeavors would require large numbers of donated human embryo derived materials from post-implantation stages but justified ethical barriers and the scarcity of such samples, make conducting human embryo-based research ethically and technically impossible. Given the capacity to generate from human naïve PSCs both embryonic cells and extraembryonic primed cells, that can co-assemble into structured and dynamic human stem-cell derived models capable of mimicking the key developmental milestones occurring during the early post-implantation stages is becoming a necessary element to establish advanced stage fully integrated embryo models[1].

Here, we readapted a recently described approach in mice and generated self-organizing human post-implantation SEMs exclusively from naive ESCs[3], and without the need to genetically modify or overexpress exogenous lineage factors for priming the naïve ESCs towards the three different extra-embryonic lineages prevalent at these developmental stages, contrary to what is currently still inevitable in mouse SEM derivation protocols[3]. The latter further underscores the self-organization capability of naïve PSCs to generate both embryonic and all extra-embryonic compartments[4], including the ExEM. The human SEMs generated *ex utero* herein mimic the 3D architecture and key developmental landmarks of *in utero* developed natural human embryos from 7-8dpf to 13-14dpf (Carnegie stages 5a-6a). We observed proper spatial allocation of cell lineages into defined embryonic and extra-embryonic compartments in the complete absence of fertilization or interaction with maternal tissues and without the need of providing external targeted signaling pathway induction during the self-organization of the aggregated cells. Moreover, we emphasize that our goal is to make an embryo model with recognizable recapitulation of the key milestones and the critical embryonic structures, which may be helpful for research purposes even if it not fully identical the natural human embryo. At the structural level, our human SEM highly resembles, but not identical, to the *in utero* situation.

While this study was under revision, two papers reported stem cell-derived aggregates with claims to mimic stages of human post-implantation development[33,34]. However, their generated embryonic-body (EB)-like aggregates did not contain the most elementary and defining hallmarks of integrated embryo models. Such EB-like aggregates (i) showed absence of essential key cell lineages of the developing embryo (e.g., trophoblast lineage, visceral and parietal primitive endoderm) (ii) lacked of key hallmarks of structural compartments with correct morphological organization (e.g., embryonic disc-, hypoblast-, bilaminar-disc-, yolk sac-, chorionic cavity- and surrounding trophoblast-like compartment) (iii) lacked dynamism relating to ability to progress structurally to next stages in development following initial aggregate formation[3]. Thus, following their inability to meet any of the elementary criteria and hallmarks laid out in the introduction, these aggregates do not qualify as models for embryos and cannot be developmentally assigned a Carnegie stage, since such developmental day staging is based on structural organization and embryo-like morphological criteria.

The current low efficiency and developmental stage variability observed during the formation of our human SEMs is a limiting factor that needs to be overcome to facilitate the use of such platforms for certain experimental set-ups. Nevertheless, the emergence of well-defined complete structures suggests that this will likely be possible. It will be of interest to further explore whether omission of RCL induced cells and relying instead on the ability of HENSM naïve ESCs ability to spontaneously give rise to PrE and

ExEM lineages in N2B27 basal conditions, can be tweaked to yield equivalent or enhanced SEM outcome to the coaggregation regiments utilized herein. Given that human embryo implantation invades the endometrium entirely[15], it will also be important to experiment whether further and better trophoblast-like compartment development and maturation (for both CTb- and STb-like cells) would occur if human SEMs would be embedded within different extracellular micro-environments. Finally, testing of whether human SEMs described here can develop further towards completing gastrulation and advancing through organogenesis as we recently achieved with mouse SEMs exclusively made from naïve ESCs[3], may be of critical experimental importance and will offer insights into previously inaccessible windows of early human development.

Acknowledgments: J.H.H lab was funded by Pascal and Ilana Mantoux; Flight Attendant Medical Research Institute (FAMRI); MBZUAI-WIS Program grant, Israel Science Foundation (ISF) - Regular Research grant 1220/20, Minerva Stiftung, Israel Cancer Research Fund (ICRF) – Research Professorship grant, Kimmel Stem Cell Research Center Grants (2019-2023) at the Weizmann Institute, United states – Israel Binational Science Foundation (BSF) research grant 2017094, a sponsored research program by RenewalBio Ltd., Dr. Barry Sherman Institute for Medicinal Chemistry and Helen and Martin Kimmel Institute for Stem Cell Research. This work in the Hanna lab was also Funded/Co-funded by the European Union: ERC-COG-2022 #101089297 – ExUteroEmbryogenesis (that funded mouse SEM related experiments only), ERC-CoG-2016 #726497 – CELLNAIVETY. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. F.L. lab is supported by the Ming Wai Lau Center for Reparative Medicine, Ragnar Söderbergs Stiftelse, Wallenberg Academy Fellow, Center for Innovative Medicine and Karolinska Institutet SFO Stem Cells and Regenerative Medicine. S.P. lab is supported by the Swedish Research Council, Swedish Society for Medical Research. S.P. holds the Canada Research Chair in Functional Genomics of Reproduction and Development (950-233204). Research in the Pasque laboratory was supported by the Research Foundation–Flanders [FWO ; Odysseus Grant G0F7716N; G0C9320N and G0B4420N.]; KU Leuven Research Fund [BOFZAP grant StG/15/021BF. and C1 grant C14/21/119]; FWO Ph.D. Fellowship to T.X.A.P. [11N3122N]; Pandarome project 40007487 [G0I7822N] (funded by the FWO and F.R.S.-FNRS) under the Excellence of Science (EOS) program. V.B is supported by the Weizmann Postdoctoral Excellence Fellowship of the Feinberg Graduate School of Science. We thank Y. Stelzer for assistance on using his previously published pipeline for scRNA-seq analysis of the placental compartment in mouse SEMs (**Extended Data Fig. 3a-b**). We acknowledge the Virtual Human Embryo (VHE) for permission to use VHE images (https://www.ehd.org/virtual-human-embryo/). J.H.H thanks Prof. Y Groner for his scientific guidance and discussions. We thank Weizmann Institute Board and Management for providing critical financial and infrastructural support to the Hanna lab.

Author contributions: B.O and E.W established the SEM aggregation conditions and protocols for ex utero culture, designed and conducted most of the wet lab work and contributed on manuscript elaboration. B.O established the human stem cell conditions, and inductions. E.W and V.B conducted most embryo immunostaining and confocal imaging. V.B made light sheet microscopy analysis and helped writing the manuscript. B.O. generated cell lines with assistance from S.V and optimized the final SEM protocol. A.A.C conducted some aggregation experiments and roller culture adaptation for the system, immunostaining and microscopy and sample preparation for 10X scRNA-seq experiments. M.Y.C helped

on human stem cell culture expansion and SEM protocol optimizations and reproducibility. C.Z conducted the sc-RNAseq comparative analysis to previous human datasets, under the supervision of F.L and S.P. T.X.A.P conducted integrative scRNA-seq analysis of TE and ExEM-like cells with existing 2D reference datasets under the supervision of V.P. S.T contributed on optimization of lineage inductions. R.C generated MEF and other critical reagents for stem cell maintenance. S.A and D.L conducted immunostainings and RT-PCR. F.R, C.J, M.R assisted in immunostainings. N.L assisted on lentivirus production and flow cytometry experiments. E.A supervised flow cytometry and sorting experiments. T.S helped in bioinformatic analysis. S.V generated plasmids. Y.A. assisted in light-sheet microscope operation and data analysis. A.A.C, M.K., M.C. and H.K.S performed RNA library preparation and sequencing. S.A provided input on optimizing light sheet microscopy experimentation and analysis. N.N. – conducted and supervised bioinformatics analyses. B.O., E.W and M.Y.C independently reproduced human SEM generation in the Hanna lab. J.H.H. conceived the idea for the project, supervised data analysis and manuscript writing.

Declaration of interests: J.H.H together with B.O, E.W, V.B, A.A.C, N.N, S.V, S.T, C.J, T.S and F.R submitted (through Yeda - Weizmann Institute of Science) patent applications relevant to the findings and technologies reported and media compositions utilized herein: (i) "ISOLATED NAIVE PLURIPOTENT STEM CELLS AND METHODS OF GENERATING SAME" filed on 23-Apr-2013 (EP 61/814,920); (ii) "MEDIA FOR CULTURING NAIVE PLURIPOTENT STEM CELLS" filed on 30-Jul-2014 (EP 62/030,792); (iii) "CULTURE MEDIA FOR PLURIPOTENT STEM CELLS" filed on 23-Jan-2020 (EP 62/795,626); (iv) "METHODS AND DEVICES FOR EX-UTERO MOUSE EMBRYONIC DEVELOPMENT" (for natural and synthetic embryos) filed on 16-March-2021 (EP 281,561); (v) "METHODS OF GENERATING A SYNTHETIC EMBRYO" (for mouse and human SEMs/synthetic embryo models) filed on 06-March-2022 (EP 63/317,036). J.H.H is a co-founder and chief scientific adviser of Renewal Bio Ltd. that has licensed technologies described herein and co-funded some parts of this project. RSeT^{TM} defined human naïve-like pluripotency growth media was licensed and commercialized by Stem Cell Technologies Inc. based on some of the indicated patents/patent applications above (RSeT^{TM} is the commercialized version of NHSM growth media originally generated and patented by J.H.H and N.N). The remaining authors report no conflict of interest.

Figure Legends

Figure 1. Optimizing human naive ESC differentiation towards extra-embryonic lineages competent for early post-implantation SEM generation. a, scheme of the tested induction (iGATA4, iGATA6, iCDX2, iGATA3) and media conditions for generating the three different extraembryonic lineages constituting the post-implantation human embryo (right) from HENSM naïve PSCs (nPSC). Epiblast (cyan), hypoblast (yellow), ExEM (grey), and trophoblast (magenta) compartments. **b**, FACS plots of PDGFRa versus SSC for PrE/ExEM-like cell induction using iGATA4 with DOX in HENSM after 6 days (right) and the control condition of naïve cells (WT cells without iGATA4, left). **c,** FACS plots of PDGFRa versus SSC for PrE/ExEM-like cell induction using iGATA4 and iGATA6 with DOX for 6 days in different media conditions as indicated (C10F4PDGF, RACL, RCL). **d,** FACS plots of PDGFRa for PrE/ExEM-like cells, using WT naïve ESCs induced for 6 days in different media conditions (N2B27, RCL, and RCL for 3 days followed by 3 days of N2B27). **e**, immunofluorescence images of WT nESCs induced for 6 days in RCL media for SOX17 (yellow), BST2 (red), and nuclei (DAPI, white). Outline indicates mutually exclusive gene expression pattern of SOX17 and BST2. **f**, FACS plots of ENPEP against TACSTD2 for TE-like lineage induction of HENSM naïve ESCs in BAP(J) regimen for 3 days. **g**, immunofluorescence images of day 6 SEM aggregates stained for OCT4 (cyan), SOX17 (yellow), and SDC1 (magenta). RACL **R**PMI based medium with **C**HIR99021, ACTIVIN, and **L**IF; RCL, **R**PMI based medium with **C**HIR99021 and **L**IF without ACTIVIN; HENSM, **H**uman **E**nhanced **N**aïve **S**tem cell **M**edia; BAP(J), DMEM/F12 based medium with ALK4/5/7 inhibitor **A**83-01, ERKi/MEKi PD0325901, and **B**MP4 for 24h and substituted with **J**AK inhibitor for the next 48h. All scale bars, 100 μm.

Figure 2. Self-assembly of human post-implantation SEM exclusively from non-transgenic naïve ESCs. a (left to right), Carnegie Stages (CS) images reused from[20] (courtesy of the Virtual Human Embryo) and schemes of early post-implantation human embryos at CS5a (7-8 days post-fertilization, dpf), CS5b (9-10 dpf), CS5c (11-12 dpf), and CS6a (13-14 dpf). **b**, scheme of the protocol (see Methods). Hypoblast/ExEM- (yellow/grey), epiblast- (cyan), and trophoblast-like (magenta) lineage priming for 3 days from naïve PSCs in HENSM is followed by aggregation (day 0) in N2B27. From day 3, SEMs are cultured in non-adherent 6-well plates on an orbital shaker in human *Ex-Utero* Culture Medium 2 (hEUCM2). **c**, representative brightfield images of day 0–8 SEMs showing growth and formation of the embryonic structures, defined by the lineage-specific immunofluorescence (see **Extended Data Fig. 6b**). **d (from left to right),** representative immunofluorescence images of SEMs from days 4, 6, and 8, showing OCT4 (cyan), SOX17 (yellow), CK7 (magenta), and nuclei (DAPI, white). **e**, quantification of the protocol efficiency for WIBR3 (left) and WIBR1 (right) ESC lines according to the morphological criteria (see

Methods). For WIBR3 N = 3 across 232, 344, and 344 aggregates; for WIBR1 N = 3 across 866, 1222, and 960 aggregates. Bars show mean values, whiskers mark s.d. **f (top)**, 3D reconstruction of the day 8 SEM shown in d (right) with segmented epiblast-like and hypoblast/YS-like compartments. **f (middle)**, segmentation of the epiblast- and hypoblast/YS-like compartments shown in 0 and 90^0 degrees of rotation. **f (bottom)**, image section of the day 8 SEM shown in d (right). Epi, epiblast-like; AC, amniotic cavity-like; Am, amnion-like; TE, trophectoderm-like; Tb, trophoblast-like; PrE, primitive endoderm-like; Hb, hypoblast-like; PYS, primary yolk sac-like; VE, visceral endoderm-like; PE, parietal endoderm-like; SYS, secondary yolk sac-like; ExEM, extraembryonic mesoderm-like; ChC, chorionic cavity-like; Sk, stalk-like; L, lacunae-like. All scale bars are 50 µm; except c. (days 0-2) 200 µm, f. (bottom): 30 µm.

Figure 3. Human SEMs undergo epiblast morphogenesis and form a bilaminar disk-like structure. a – b, representative immunofluorescence images of ESCs in HENSM **(a)** and day 3 SEMs in N2B27 **(b)**, expressing OCT4 (cyan), DNMT3L (green, top), and STELLA (green, middle), but not OTX2 (red). **c**, images of day 4 – 8 SEMs showing epiblast- (OCT4, cyan), Hypoblast- (SOX17, yellow), Hypoblast/ExEM- (GATA6, yellow), and trophoblast-like (CK7, SDC1 or GATA3, magenta) compartments. Right, zoom. **d**, Epi-like cell numbers in successfully developed SEMs from day 4 (N = 1, n = 5), day 6 (N = 3, n = 12), and day 8 (N = 3, n = 6). Whiskers extend 1.5x of the interquartile range (box) around median line. p-values: 0.0003, 0.0043, 0.1024; two-sided Mann-Whitney U-test. **e**, images of day 6 SEMs showing phospho-Ezrin/Radixin/Moesin (ph-ERM, red; top), Podocalaxyn (PODXL, red; bottom), and F-ACTIN (green)**. f**, image of the day 8 SEM with Anterior-Posterior (A-P) axis with T/BRA (red) in epiblast-like compartment (OCT4, cyan) opposite to CER1 (green). Right, zoom. **g**, image of day 8 SEM with $TFAP2A^+$ (magenta)/$ISL1^+$ (green)/$SOX2^-$ (cyan) amnion-like cells (arrows on the right). **h**, 3D segmentation of the amnion- (TFAP2A, magenta) and the embryonic disk-like (SOX2, cyan) compartments in day 8 SEM; XY and XZ views. **i**, image of amnion-like structure in day 7 SEM with squamous OCT4+ cells (cyan). **j**, **k**, Carnegie Stage (CS) CS5c and CS6a reused histological sections from [20]. Amnion (Am, arrow). **k**, image of squamous $TFAP2A^+$ (magenta) and $SOX2^-$ (cyan) amnion-like cells in day 8 SEMs. **m**, images of day 8 SEM (Z slices 22, 26). Right, zoom into the $OCT4^+$ (cyan)/$SOX17^+$ (yellow)/$BLIMP1^+$ (red) PGC-like cells (arrows). Inset, brightfield image, SEM perimeter is outlined. Single and double asterisks mark proamniotic- and amniotic-like cavities, respectively. Nuclei, (DAPI, white). All scale bars are 50 µm; except e. (right, zoom): 25 µm, h.: 20 µm, m. (right, zoom): 10 µm.

Figure 4. Human SEMs recapitulate YS-like lumenogenesis and SEM scaffolding by ExEM-like cells. **a**, representative immunofluorescence image of day 6 SEM; yolk sac (YS)-like, parietal-endoderm (PE)-like and visceral endoderm (VE)-like (SOX17, yellow) compartments, F-ACTIN (white), nuclei (DAPI,

blue). Right, zoom into VE- and PE-like cells. **b**, cell aspect ratio in VE- (n=14) and PE-like cells (n=12) of the SEM **(a)**. Whiskers extend 1.5x of the interquartile range (box) around a median line. Student's t-test p-value=0.0000000864. **c**, Carnegie Stage (CS) CS5c histological sections reused from[20], showing primary yolk sac (PYS)-, VE- and PE- like compartments (right). **d**, images of day 6 SEMs; Epi- (OCT4, cyan), Hypoblast- (SOX17, yellow), Hypoblast/ExEM- (GATA6, yellow), and trophoblast-like (CK7, GATA3, magenta) compartments. Right, 2x zoom; ExEM-like cells. **e**, image of day 6 SEM showing chorionic cavity (ChC)-like structure within $GATA6^+$ (red)/$SOX17^-$ (yellow) ExEM-like tissue (outlined). **f**, image of day 6 SEM expressing $BST2^+$ (red) underneath $SOX17^+$ YS-like structure (yellow). Right, zoom. **g**, image of day 8 SEM expressing VIM (red) underneath SOX17+ YS-like cells (yellow). Bottom, zoom. **h (inset)**, schematic of the ExEM cells (grey) and secondary yolk sac (SYS, yellow) in 14 dpf human embryo. **h**, images of day 8 SEM showing $FOXF1^+$ (green)/$GATA6^+$ (red) ExEM-like cells. **i**, image of day 8 SEM with a cavitated BST2+ ExEM-like cells (white); SOX17 (yellow), red arrows point at PYS remnants-like cells. **j**, CS6 histological section reused from[18], showing filamentous ExEM and SYS; 100x. **k (top)**, scheme of the regular aggregation experiment (control) and without Tb-like cell (no BAP(J)). **k (bottom)**, images of control SEM and no BAP(J) aggregate from day 6; SOX17 (yellow), CK7 (magenta). **l,** quantification of day 6 aggregates in control (N =2, n = 344, 1321) and no BAP(J) (N =2, n = 472, 193) conditions; bars show mean values from 2 biological replicates, and each dot indicates an average value of a biological replicate. OCT4 (cyan); **c – k,** nuclei (DAPI, white). All scale bars are 50 μm; except Zoom-ins of a., c., f: 10 μm.

Figure 5. Trophoblast-like compartment integration and maturation in human SEMs. a (top), representative immunofluorescence images of day 6 SEMs showing epiblast- (OCT4, cyan), hypoblast- (SOX17 or GATA6, yellow), and trophoblast-like (SDC1, CK7, or GATA3, magenta) compartments. **a (bottom)**, single-channel images of the trophoblast-like compartment, surrounding the SEMs. **b**, average percentage of aggregates surrounded by trophoblast-like compartment at day 6, as judged by the expression of SDC1, CK7, or GATA3. SDC1 (N=3 across 533, 232, and 94 aggregates), CK7 (N=5 across 302, 153, 344, 344, and 85 aggregates), and GATA3 (N=3 across 295, 170, and 62 aggregates). Error bars indicate s.d. **c (left)**, maximum intensity projection image of day 6 SEMs showing HCGB expression in the outer cells; F-ACTIN (red). **c (right)**, image of the same SEM showing lacunae-like structures (marked with asterisks) inside the outer syncytiotrophoblast-like layer; nuclei (DAPI, white). **d**, 3D projection of the lacunar-like structures (outlined) in the trophoblast-like layer of the day 6 SEM. Immunofluorescence for SDC1 (magenta), HCGB (green), nuclei (DAPI, white). **e**, histological section and 3D reconstruction (top right) reused from the Carnegie collection of a human embryo[20] at Carnegie Stage (CS) 5c, showing lacunae in the syncytiotrophoblast (asterisks). **f**, representative brightfield and immunofluorescence images of two

different Z planes (number 3 and 11, top and bottom panels, respectively) of day 6 SEMs showing epiblast- (OCT4, cyan), hypoblast- (SOX17, yellow), and trophoblast-like (SDC1, magenta) compartments. **f (top)**, lacunae-like structures are outlined and marked with asterisks. Right, zoom into the lacunae-like structure (top). **f (bottom right)**, zoom into the outer syncytiotrophoblast-like layer; brackets mark thickness of syncytium-like tissue. **g**, image of day 6 SEM showing CK7 (magenta), F-ACTIN (red), and nuclei (DAPI, white). Bottom, zoom into the multinucleated syncytiotrophoblast-like cell; arrows point at multiple nuclei inside the single cell. All scale bars are 50 µm; except d.: 20 µm, f. (zoom, top): 10 µm.

Figure 6. scRNA-seq analysis validates key cell type identity comprising the human SEM.
a, UMAP of single cells from human SEMs colored by the cell clusters, identified based on known marker genes and gene signatures (see Methods). Cluster sizes: 1963 (cluster 0), 1483 (1), 1431 (2), 1344 (3), 1265 (4), 957 (5), 905 (6), 898 (7), 662 (8), 448 (9), 441 (10), 265 (11), 128 (12). Extra-embryonic mesoderm-like (ExEM), yolk sac-like (YS), hypoblast-like (Hb), secondary yolk sac-like (SYS), amnion-like (Am), syncitiotrophoblast-like (STb). **b**, normalized expression of key marker genes projected on the UMAP. Cell type clusters are highlighted in red. **c**, dot plot illustrating the expression of key marker genes across the 12 cell type clusters. Cluster 10* indicates a sub-group of cells (n=31) reannotated as CTb (cytotrophoblast-like cells) **(Extended Data Fig 12d, 14)**. Color intensity indicates average expression. Dot size indicates the percentage of the cells in the cluster which express the marker. **d (top)**, heatmap showing normalized expression of PGC markers (SOX17, PRDM1, NANOS3, NANOG, POU5F1 (OCT4)). 9 cells were $POU5F1^{+}/PRDM1^{+}/SOX17^{+}$ and $NANOS3^{+}$. **d (bottom)**, $POU5F1^{+}/PRDM1^{+}/SOX17^{+}$ PGC-like cells (n=27) were identified in Epi-like clusters, with majority in cluster 4 (n=19, green). **e (left)**, UMAP projection of integrated human embryonic reference[26] consisting of 6 human embryonic data sets spanning early zygotes, *in vitro* cultured human blastocysts[28–30], 3D *in vitro* cultured human blastocysts until pre-gastrulation stages [31], and a Carnegie Stage 7 (CS7) 16-19 dpf human gastrula[32]. The color of each data point corresponds to the cell annotations from the respective publication. **e (right)**, grey data points represent embryonic reference cells, as in the left panel. Colored triangles represent the projection positions of the neighborhood nodes from SEM cells onto the human embryonic reference. SEM-CTb*: neighborhood nodes representing cluster 10 cells (Amnion-like cells) projected on the reference CTb.

References

1. Rossant, J. & Tam, P. P. L. Opportunities and Challenges with Stem Cell-Based Embryo Models. *Stem Cell Reports* **0**, (2021).
2. Oldak, B., Aguilera-Castrejon, A. & Hanna, J. H. Recent insights into mammalian natural and synthetic ex utero embryogenesis. *Curr Opin Genet Dev* **77**, 101988 (2022).
3. Tarazi, S. *et al.* Post-gastrulation synthetic embryos generated ex utero from mouse naive ESCs. *Cell* **185**, 3290-3306.e25 (2022).
4. Bayerl J *et al.* Principles of Signaling Pathway Modulation for Enhancing Human Naïve Pluripotency Induction. *Cell Stem Cell* **28**, 1549-1565 (2021).
5. Rivron, N. C., Arias, A. M., Pera, M. F., Moris, N. & M'hamdi, H. I. An ethical framework for human embryology with embryo models. *Cell* **186**, 3548–3557 (2023).
6. Deglincerti, A. *et al.* Self-organization of the in vitro attached human embryo. *Nature* **533** 251-254. (2016) doi:10.1038/nature17948.
7. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
8. Bergmann, S. *et al.* Spatial profiling of early primate gastrulation in utero. *Nature* **609**, 136–143 (2022).
9. Pham, T. X. A. *et al.* Modeling human extraembryonic mesoderm cells using naive pluripotent stem cells. *Cell Stem Cell* **29**, 1346-1365.e10 (2022).
10. Linneberg-Agerholm, M. *et al.* Naïve human pluripotent stem cells respond to Wnt, Nodal and LIF signalling to produce expandable naïve extra-embryonic endoderm. *Development* **146**, (2019).
11. Ohinata, Y. *et al.* Establishment of mouse stem cells that can recapitulate the developmental potential of primitive endoderm. *Science (1979)* **375**, 574–578 (2022).
12. Viukov, S. *et al.* Human primed and naïve PSCs are both able to differentiate into trophoblast stem cells. *Stem Cell Reports* **17**, 2484–2500 (2022).
13. Okae, H. *et al.* Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22**, 50-63.e6 (2018).
14. Io, S. *et al.* Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell* **28**, 1023-1039.e13 (2021).
15. Kagawa, H. *et al.* Human blastoids model blastocyst development and implantation. *Nature* **601**, 600-605 (2022).
16. Yang, R. *et al.* Amnion signals are essential for mesoderm formation in primates. *Nature Communications 2021 12:1* **12**, 1–14 (2021).
17. Cui, G. *et al.* Spatial molecular anatomy of germ layers in the gastrulating cynomolgus monkey embryo. *Cell Rep* **40**, (2022).
18. Luckett, W. P. *Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos*. *American Journal of Anatomy* vol. 152 (1978).
19. Knöfler, M. *et al.* Human placenta and trophoblast development: key molecular mechanisms and model systems. *Cellular and Molecular Life Sciences* **76**, 3479–3496 (2019).
20. O'Rahilly, R. & Müller, F. Developmental stages in human embryos: revised and new measurements. *Cells Tissues Organs* **192**, 73–84 (2010).

21. Zheng, Y. *et al.* Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573,** 421-425 (2019).
22. Ross, C. & Boroviak, T. E. Origin and function of the yolk sac in primate embryogenesis. *Nat Commun* 1–14 (2020) doi:10.1038/s41467-020-17575-w.
23. Enders, A. C., Schlafke, S. & Hendrickx, A. G. Differentiation of the Embryonic Disc , Amnion , and Yolk Sac in the Rhesus Monkey. Am J Anatomy **177**, 161–185 (1986).
24. Turco, M. Y. & Moffett, A. Development of the human placenta. *Development* **146** 1–14 (2019) doi:10.1242/dev.163428.
25. Irie, N. *et al.* SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253–268 (2015).
26. Zhao, C. *et al.* Reprogrammed blastoids contain amnion-like cells but not trophectoderm. *bioRxiv* 2021.05.07.442980 (2021) doi:10.1101/2021.05.07.442980.
27. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20**, 1131–1139 (2013).
28. Yanagida, A. *et al.* Naive stem cell blastocyst model captures human embryo lineage segregation. *Cell Stem Cell* **28,** 1016-1022 (2021). doi:10.1016/j.stem.2021.04.031.
29. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012-1026 (2016).
30. Meistermann, D. *et al.* Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. *Cell Stem Cell* **28**, 1625-1640.e6 (2021).
31. Xiang, L. *et al.* A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature 2019* **577***:7791* **577**, 537–542 (2019).
32. Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature 2021 600:7888* **600**, 285–289 (2021).
33. Pedroza, M. *et al.* Self-patterning of human stem cells into post-implantation lineages. *Nature* (2023) doi:10.1038/s41586-023-06354-4.
34. Weatherbee, B. A. T. *et al.* A model of the post-implantation human embryo derived from pluripotent stem cells. *Nature* (2023) doi:10.1038/s41586-023-06368-y.

Methods

Ethics

All experiments reported herein involving human ESCs/iPSCs were conducted following obtaining approval by the Weizmann Institutional Review Board (IRB) (approval #1868-2) to generate human embryo(id) models (with and without transgenes expression to form extra-embryonic cells) from human ESC and iPSC lines[35]. Consent forms for all the previously derived human ESC lines used in this study (WIBR1, WIBR2, WIBR3 and RUES2 human ESCs) allow their use in human embryo model development and research. The newly derived human iPSC lines used herein (JH22 and JH33 iPSCs) were consented for use in human embryo model development according to Weizmann Institutional Review Board (IRB) approval #1871-2 and #1868-2. All the experiments reported herein follow the latest ISSCR guidelines released in 2021[36]. This study does not involve derivation of new human ESC lines, does not use any newly obtained samples from fetal abortions, and does not use any newly donated human blastocysts. Further, this study does not involve in utero transfer of any human SEMs into any other species, consistent with ISSCR guidelines and Israeli legislation. Finally, all the human SEMs described herein do not correspond to developmental stages beyond 14 dpf. Not all the features of a 14 day human embryo are present by morphology and immunostaining at the experimental endpoint.

Data reporting

No statistical methods were used to predetermine sample size. Samples were randomly allocated when placed in the different growth conditions. Other experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment since there was no relevant scientific reason to do so. Throughout the manuscript all data points and samples represent biological replicates (N indicated in figure legends or in graphs; n number of samples per biological replicate is indicated where relevant), except for RT-PCR where per individual representative experimental graph, each sample was run as a technical triplicate per each gene, number of biological replicates of each RT-PCR panel from which the representative experiment was taken and shown is indicated per panel.

Statistics and Reproducibility

The following figure panels showing microscopy images are representative of independent biological replicates as follows (N) as detailed below: Fig. 1: (e) N=3, (g) N=11 ; Fig. 2: (d) N=11 ; Fig. 3: (a-b) N=3, (c) N=11, (e-k) N=2, (m) N=3 ; Fig. 4: (a) N=12, (d) N=11, (e-i) N=5, (k) N=3 ; Fig. 5: (a,f) N=11, (c-d,g) N=5 ; Extended Data Fig. 2: (a-c) N=4 ; Extended Data Fig. 3: (c) N=3, (d) N=4 ; Extended Data Fig. 4: (b) N=4 ; Extended Data Fig. 5: (b-Control) N=11, (b-No HENSM, e) N=3, (b-No BAP(J))

N=4, (b-No RCL) N=3, (d) N=3 , (e) N=3 ; Extended Data Fig. 6: (a) N=2, (b,c) N=11 ; Extended Data Fig. 7: (b) N=5 ; Extended Data Fig. 8: (a, c-e, g) N=3, (f) N=3, (h) N=3 ; Extended Data Fig. 9: (a-b, g-h) N=3, (c) N=3, (d-e) N=5 ; Extended Data Fig. 10: (a) N=11, (b-c) N=7, (e, f) N=3, (i) N=2 ; Supplementary Fig. S1 (b) N=5 ; Supplementary Fig. 2: (a-b) N=4 ; Supplementary Fig. 4: (b) N=6, (c-f) N=3 ; Supplementary Fig. 5: (a) N=3, (c) N=3 ; Supplementary Fig. 6: (a, c-d) N=3 ; Supplementary Fig. 7: (b) N=3 ; Supplementary Fig. S8: (b-c) N=5 ; Supplementary Fig. 9: (b-c) N=3 ; Supplementary Fig. 10: (a) N=3, (b) N=4, (d) N=2, (e) N=3 ; Supplementary Fig. 11: (b) N=11, (c) N=3 ; Supplementary Fig. 12: (b,d) N=3, (c,e-f) N=11, (g-h, j) N=4 ; Supplementary Fig. 13: (b) N=3, (c) N=3 ; Supplementary Fig. 14: N=3.

Pluripotent stem cell lines

The following already established human ESC lines were used: WIBR3 (WT female), and WIBR1 (WT male) human ES lines[35]. WIBR2 female hESC line so far has failed to give us validated complete day 8 SEMs. RUES2 hESC line carrying fluorescent reporters for endoderm, mesoderm and ectoderm differentiation (Kind gift from. A. Brivanlou) was also used to validate endogenous marker gene expression in SEMs via the reporters. Newly derived JH22 and JH33 iPSC lines (taken from a healthy adult male middle eastern donor and obtained following donor informed consent to make genetically unmodified iPS lines from donated peripheral blood (as approved by the Weizmann Institutional Review Board (IRB), approval# 1871-2)), were used for lineage priming efficiency experiments where indicated. Lis49 hESC was used for naïve vs. primed scRNA-seq comparison alongside WIBR3 hESC line. Previously established mouse Tet-ON iGata4 KH2-WT ESC line was used for comparing induction efficiency of $Pdgfra^{+}$ cells[3]. Tet-ON iGata4 KH2-WT ESC, Tet-ON iGata4 KH2-iCdx2 ESC and WT BVSC mouse ES lines were used for mouse SEM experiments. All cell lines were routinely checked for Mycoplasma contaminations every month (Lonza–MycoAlert), and all samples analyzed in this study were not contaminated.

Human HENSM naïve and RSeT naïve-like PSC *in vitro* culture conditions

Golden stocks of human PSCs were cultured on a feeder layer of irradiated mouse embryonic fibroblasts (MEFs - from E13.5 ICR-DR4 mouse embryos) and maintained in conventional human FGF/KSR primed conditions and frozen for future experimentation. FGF/KSR conditions on MEF substrate: 400ml of DMEM-F12 (Invitrogen 10829) supplemented with 20% Knockout Serum Replacement (KSR, Invitrogen 10828-028), 1mM GlutaMAX (Gibco 35050061), 1% nonessential amino acids (BI 01-340-1B), 1% Sodium-pyruvate (BI 03-042-1B), 1% Penicillin-Streptomycin (BI 03-031-1B) and 8 ng/mL bFGF (Peprotech 100-18B-1MG).

From those primed human PSC lines were routinely expanded in RSeT media (Stem Cell Technologies Inc. Cat. #17148311) (RSeT Medium (2-component) – is a commercialized version of NHSM

media described in Gafni et al.[7], and was assembled according to manufacturer instructions, including dissolving 250µl Matrigel in 500ml media as instructed - https://www.stemcell.com/products/rset-medium-2-component.html), and expanded on Matrigel or Cultrex coated plates and passaged with TrypleE every 3-5 days. Those were kept for up to 40 passages in RSeT media (with and without freezing/thawing in this state) and used to quickly and homogenously convert into naïve state in HENSM conditions (Detailed step-by-step protocol for key experiments in this study is published on Protocol Exchange[37]

To reprogram the latter naïve-like/formative (RSeT) PSCs to a naïve HENSM state, human ESCs were transferred and expanded as in Bayerl et al [4], in serum-free HENSM on plates coated with 1% Matrigel (corning 356231) (detailed HENSM protocol is also found on Hanna lab website protocol section: https://www.weizmann.ac.il/molgen/hanna/) (or less preferably for SEM protocol on MEF/gelatin-coated plates). HENSM version used herein consists of: 470 ml of 1:1 mix of Neurobasal (Invitrogen 21103-049) and DMEM/F12 (Invitrogen 21331), 5 ml Pen-strep (Biological Industries 03-033-1B), 5 ml GlutaMAX (Invitrogen 35050061), 5 ml NEAA (Biological Industries 01-340-1B), 5 mL Sodium Pyruvate (Biological Industries 03-042-1B), 10ml B27 supplement (Gibco 17504-044 (LOT 17504-044) or in house prepared or by Lifegene NCS21 - D5482), 50 µg/ml L-ascorbic acid 2-phosphate (Sigma - A8960), 1 mL of Geltrex (Invitrogen A1413202/A1413302) (0.2% final concentration), 5ml N2 supplement (Gibco 17502048 or in house prepared), 10ng/ml LIF (Peprotech 300-05 or in house prepared), 2µM WNTi/TNKi = XAV939 (Sigma X3004), 2µM PKCi Gö6983 (Axon 2466), 1-1.2µM MEKi/ERKi PD0325901 (Axon 1408), 1.2 µM SRCi CGP77675 (Axon 2097 (Batch 9 or 10) or SigmaAldrich SML0314 (Batch 10)), 5 ng/ml Activin A (Peprotech 120-14E), 1.2 µM ROCKi Y27632 (Axon Medchem 1683), and 0.8 µM BIRB796 P38i (Axon Medchem 1358). At least 2-3 passages in HENSM conditions were applied before cells were used for experiments. Human naïve HENSM PSC lines were used for up to 10 passages since transfer into HENSM conditions. For maintenance of ESCs/iPSCs in naïve HENSM conditions, cells were passaged every 3-5 days using TryplE (Gibco 12604054). Different batches of B27 gives some difference in naïve pluripotency markers expression, which can influence SEM derivation efficiency. We highlight that since large numbers of cells were needed to obtain a proper induction yield per experimental batch, we used HENSM that include low dose 5ng/ml ACTIVIN A which is a booster for human naïve cell proliferation rate (as indicated in Bayerl et al [4]). We could not reach very high numbers of cells in HENSM conditions without ACTIVIN to meaningfully run experiments and reach equivalent conclusions. Naïve HENSM, naïve-like (RSeT) and primed hESCs were expanded and primed towards different lineages in a 5% CO2 incubator at 5% O2 at 37C. If HENSM cells are not homogenous and differentiated cells are apparent, check that HENSM was assembeled according to protocol and no components were mistakenly omitted. Also consider using a different B27 vendor or batch, and make sure that the medium is not more than 10 days old. Some PSC lines may require small modifications in the HENSM protocol (e.g., slight increase in ACTIVIN dose used

or MEKi/ERKi used), therefore please refer to Hanna lab HENSM protocol guidelines posted on the Jacob Hanna lab website (https://www.weizmann.ac.il/molgen/hanna/) for further standardization if needed.

Generation of iGATA4, iGATA6, iCDX2 and iGATA3 human ESCs clones

To generate Tet-ON inducible lines, we employed a PiggyBac plasmid expressing cDNA insert and transposase vector of choice under the control of a doxycycline-inducible promoter (a kind gift from Volker Busskamp, Addgene plasmid #104454). The donor vector carries M2RtTa and a site for cDNA insert of transcription factor of interest. We used this vector to generate 4 different DOX inducible lines in WIBR3 WT human female ESCs: human iCDX2 or human iGATA3 (to promote human ESC differentiation towards trophectoderm) and human iGATA4 or human iGATA6 (to promote human primitive endoderm/ extra embryonic mesoderm priming from human ESCs). Puromycin selection was applied for approximately 6-8 days. Resistant clones were picked and cultured for downstream characterization. Insertion was validated by immunostaining after DOX (2µg/ml) induction of the gene of interest. Transgene expression was verified to be specifically detected only after DOX addition in the corresponding lines. Detailed generation, characterization and validation of these lines can be found on (Extended Data Figure 1, 3, 7). Generation of fluorescent labeled WIBR3 iCDX2 line was made after transduction with lentivirus constitutively expressing tdTomato protein. For lentivirus generation, HEK293T cells were used as the most conventional and commonly used line for lentiviral packaging and generation. HEK293T cells were plated on 10 cm dishes filled with 10 ml DMEM 10% FBS and Pen/Strep, at a density of 5.5 million cells per plate. On the next day, cells were transfected with Second-generation lentiviral vectors (Addgene 8455 and 8455), using X-tremeGENE 9 transfection reagent, along with 16µg of the target plasmid for the transduced fluorescent protein (tdTomato). The supernatant containing the virus was collected 48hr following transfection, filtered using 0.45µm filter and concentrated by ultracentrifugation. Human ESCs were plated in mTESR medium on Matrigel coated 6-well plates at low density, next day they were transduced with lentivirus in the presence of protamine sulfate (10 µg/ml) for 6 hours, afterwards medium was exchanged. After 2 days, the infected human ESCs were expanded for 1 passage and the positive population was sorted using FACS and further expanded for experimentation.

Derivation of a stable human TSC line from human ESCs

TSC lines were produced from Naïve (HENSM) (nTSC) and primed conditions (pTSC) according to Okae et al and Viukov et al, respectively[12,13]. Briefly, human naïve WIBR3 ESCs were expanded at least 3 passages in naïve or primed conditions and then transferred into TSC media (TSCm) on 1% Matrigel (corning 356231)-coated plates. After 3 passages, stable TSC lines could be established and could be passaged up to 70 times or more. Cells were expanded in a 5% CO_2 incubator at 5% O_2. For maintenance,

human TSC were passaged with TrypLE (Gibco 12604054) when reached 70%-80% confluency. Immunostaining was performed to confirm human TSC identity: human TSC cells were negative for CDX2, and positive for Cytokeratin7 (CK7) GATA3 and TFAP2C, consistent with previous reports. Only confirmed lines were used for SEM experiments. Human TSC media (TSCm) used herein was previously described in[13] with slight modifications: 470 ml DMEM/F12 (Invitrogen 21331), 5 ml Commercial N2 supplement (Invitrogen 17502048), 10 ml B27 supplement (Invitrogen 17504-044), 5 mL Sodium Pyruvate (Biological Industries 03-042-1B), 5 mL Penicillin/Streptomycin (Biological Industries 03-033-1B) 5ml (Biological Industries 03-033-1B), 5 mL GlutaMAX (Invitrogen 35050061), 5 mL NEAA (Biological Industries 01-340-1B), 50 µg/ml L-ascorbic acid 2-phosphate (Sigma A8960), 50ng/ml Human EGF (Peprotech AF-100-15), 0.75-1µM TGFRi A83-01 (Axon 1421), 2µM GSK3i CHIR99021 (Axon Medchem 1386), and 5µM ROCKi Y27632 (Axon 1683).

Priming of human naïve HENSM ESCs towards trophectoderm (TE)-like cells for SEM generation

Human TE-like cells were obtained from human naïve ESCs expanded in human HENSM conditions for at least 3 passages, 24 hours before the BAP(J) induction/priming initiation, HENSM naïve PSCs were seeded in HENSM onto 1% Matrigel (corning 356231) coated plates supplemented with 10µM ROCKi Y27632 (Axon 1683). Next day of the seeding, HENSM was removed and the 72h BAP(J) protocol was started. The 3-day BAP(J) media treatment for human TE-like cell priming/induction was previously described and is adapted from Io et al[38] as follows: BAP(J) media was used for 72 hours total. This medium consisted on 2µM TGFRi A83-01 and 2µM MEKi/ERKi PD0325901 base, which was complemented with 10ng/ml Human recombinant BMP4 (Peprotech) only for the first 24h, and then BMP4 was substituted with for 1µM JAK inhibitor I (Calbiochem 420099) on day 2 and 3. The base medium consisted on: 470 ml of 1:1 mix of Neurobasal (Invitrogen 21103-049) and DMEM/F12 (Invitrogen 21331), 5 ml penicillin-streptomycin (Biological Industries 03-033-1B), 5 ml GlutaMAX (Invitrogen 35050061), 5 ml NEAA (Biological Industries 01-340-1B), 5 ml Sodium Pyruvate (Biological Industries 03-042-1B), 10 ml B27 supplement (Gibco 17504-044), 5 ml N2 supplement (Invitrogen 17502048 or in house prepared), 2µM TGFRi A83-01 (Axon Medchem A83-01), 2µM MEKi/ERKi PD0325901 (Axon Medchem 1408). All the process was incubated in a 37C incubator with 5% O_2 and 5% CO2. The end of 72h BAP(J) induction/priming regimen becomes day 0 of the SEM co-aggregation protocol which is the day in which BAP(J) treated cells are harvested and co-aggregated. Importantly, please note that cell confluency during the TE-like cell induction was shown to be essential for high quality and highly efficient reproducible results and needs to be calculated for each cell line and condition tested. In our case for WIBR3 hESC line growth on Matrigel, 1x10^6 HENSM PSCs per 10c"m Matrigel coated plate showed optimal results. For WIBR1 hESC line grown in HENSM on Matrigel coated plates, seeding 2x10^6 HENSM cells on a 10c”m

Matrigel coated dish and then initiating BAP(J) regimen, gave optimal results as checked by FACS on day 3 of BAP(J) regimen. We recommend conducting cell confluency plating curve for BAP(J) induction to be checked on day 3 since induction initiation using ENPEP/TACSTD2 expression in FACS, that can give an idea on adequate HENSM PSC numbers to be seeded for different line calibration of the optimal TE-like cell induction/priming expected outcome, that then leads to successful SEM generation upon moving to the coaggregation stage of the protocol.

To validate the identity of the starting BAP(J) cells derived after 3 days from HENSM ESCs, we applied scRNA-seq and integrated the data with a reference naive TE and naive cytotrophoblast differentiation protocol involving BAP(J) during the first 3 days from t2iLGö cells[14]. BAP(J) cells induced herein for 3 days, aligned predominantly to previously described day 2 and day 3 naive TE but not day 10 naive cytotrophoblast (**Extended Data Fig. 3g**). Thus, BAP(J) treatment induces naive TE-like cells from WT hESCs expanded in our devised HENSM conditions. Please note that the cells induced in BAP(J) regimen do not maintain their TE-like identity if maintained in BAP(J) for more than 3-4 days or if moved to N2B27 media after day 3 (which was used in the first 3 days of SEM formation protocol). This indicates that the switch from BAP(J) to N2B27 is not the reason for reduced efficiency.

Primitive Endoderm (PrE)- and Extra-Embryonic Mesoderm (ExEM)-like cell priming from human naïve HENSM PSCs

Pre/ExEM-like cells were primed/induced from human naïve ESCs expanded in HENSM conditions for at least 3 passages as described above. For priming into Pre/ExEM-like cells, HENSM cells were plated onto gelatin-MEF coated plates in HENSM media with 10µM ROCKi Y27632, the day before the 72h RCL induction initiation. On the next day after cell seeding, HENSM medium was changed to RCL for 72h, and RCL media was exchanged every 24h. RCL is composed of: 480 ml RPMI media (GIBCO 21875-03), 10 ml B27 minus insulin supplement (Invitrogen A18956-01), 1mM GlutaMAX (Invitrogen 35050061), 1 % penicillin-streptomycin (Invitrogen), 3µM CHIR (Axon Medchem 1386) and 10ng/ml LIF (Peprotech 300-05). RCL medium contains the same composition as RACL but without adding recombinant Activin. WT human naïve ESCs or iGATA4/iGATA6 cell lines were employed for induction in the presence or absence of DOX as indicated. All the process was incubated in a 37ºC incubator with 5% O_2 and 5% CO_2. The end of 72h RCL induction/priming regimen becomes day 0 of the SEM co-aggregation protocol which is the day in which RCL treated cells are harvested and co-aggregated Please note in **Fig. 1d** that shows 6 days of RCL induction or 3 days of RCL and 3 days of basal N2B27 induction yielded similar outcome of $PDGFRa^+$ fraction, so in this case the switch of media is not the reason for reduced efficiency.

We note that cell confluency during the RCL induction was found to be essential for high efficiency and reproducible induction results and need to be calculated for each HENSM naïve PSC line tested. In our case for WIBR3 cell line grown in HENSM on Matrigel in preparation for RCL induction, plating $8x10^5$ HENSM naïve ESCs per one 10c"m MEF coated petri dish showed best results regarding RCL induction efficiency outcome at day 6, while in the case of WIBR1 cell line, plating $2x10^6$ HENSM PSCs per one 10c"m MEF coated petri dish showed optimal RCL induction results on day 6. Per each HENSM naïve line calibration, we recommend conducting cell number plating curves, followed by using PDGFRa expression by FACS on day 6 after RCL induction initiation, can give expermintalists a good idea on optimal HENSM PSC cell numbers to be plated for each line to obtain optimal induction efficiency of Pre/ExEM-like cells that then leads to successful SEM generation upon moving to the coaggregation stage of the protocol.

Generation of human structured stem cell-derived embryo models (SEMs)

To generate SEMs from human naïve pluripotent cells (WIBR1, WIBR3 and RUES2 hESC lines), three starting cell mixtures were co-aggregated using AggreWell 400 24-well plate (STEMCELL Technologies 34415):

1) Naïve ESC/PSC (nESC/nPSC) WT cells cultured in HENSM medium in a 5% CO_2 incubator at 5% O_2 at 37ºC.
2) For the primitive endoderm-like and extra-embryonic mesoderm-like compartments (PrE-like/ExEM-like), naïve WT cells were plated on irradiated MEF (mouse embryonic fibroblast conditions)/Gelatin coated plates in HENSM supplemented with ROCKi 10µM (Axon Medchem 1683). The next day, cells were washed with PBS twice (without harvesting), and HENSM was replaced by RCL medium. RCL was kept for 72h with 24h medium exchanges in a CO_2 incubator at 5% O_2 at 37ºC.
3) For the trophectoderm-like lineage, naïve WT cells were plated on feeder free conditions (Matrigel) in HENSM supplemented with ROCKi 10µM. The next day, cells were washed with PBS twice (without harvesting), and HENSM was replaced with BAP medium for 24 hours, following by replacement with AP(J) for another 48 hours in a 5% CO_2 incubator at 5% O_2 at 37ºC (termed BAP(J) protocol).

Co-aggregation was defined as time point 0 of the protocol. 12-24h before aggregation all donor cells were supplemented with ROCKi 10µM (Axon Medchem 1683). At the day of aggregation (day 0), AggreWell 400 24-well plate preparation was done according to manufacturer instructions. Briefly, 500 µl of anti-adherence rinsing solution (STEMCELL Technologies 07010) was added to each well, the plate was centrifuged at 2,000g for 5 minutes and incubated 30 min at room temperature. Subsequently, rinsing solution was removed and the plate was washed with PBS. Each well was filled with 500 µL of aggregation

medium and kept at 37º C for medium equilibration. Aggregation medium (BSA supplemented N2B27 media) consisted in 500ml 1:1 mix of Neurobasal (Invitrogen 21103-049) and DMEM/F12 (Invitrogen 21331), 5 ml penicillin-streptomycin (Biological Industries 03-033-1B), 5 ml GlutaMAX (Invitrogen 35050061), 5 ml NEAA (Biological Industries 01-340-1B), 5 ml Sodium Pyruvate (Biological Industries 03-042-1B), 10 ml B27 supplement (Invitrogen 17504-044), 5 ml N2 supplement (Invitrogen 17502048), 1ml β-mercaptoethanol 50mM (Gibco 31350-010), 2.25ml of BSA solution 35% (Sigma A7979).

The three cell populations were collected with TrypLE (Thermo Fisher 12604054) (3 minutes for the HENSM and RCL-induced cell populations, and 5 minutes for the BAP(J) primed cells) at 37º C, Afterwards TrypLE was removed with vacuum and the cells were incubated for two minutes at room temperature and cells were subsequently collected with PBS. Cells were centrifuged at 1300 rpm for 3-5 minutes and resuspended in aggregation medium. Next, RCL-induced cells were plated on gelatinized tissue culture plates on MEF medium consisting on 500ml DMEM (Gibco 41965-039) 20% FBS (Sigma, F7524-500ml), 5 ml penicillin-streptomycin (Biological Industries 03-033-1B), 5 ml GlutaMAX (Invitrogen 35050061), 5 ml NEAA (Biological Industries 01-340-1B), 5 ml Sodium Pyruvate (Biological Industries 03-042-1B), for MEF depletion for 30 minutes at 37º C. At the end of MEF depletion, the supernatant was collected and passed through a 70uM cell strainer, and all three cell types were centrifuged separately and resuspended and passed through a 70μM cell strainer in N2B27 medium. The three cell fractions were counted and combined as follows in an Aggrewell 400 plate (in each well there are 1200 microwells): Total cell number per single individual microwell/aggregates is 120 cells, total number of cells per a single well of a 24 well plate is 144,000 cells (based on the calculation of 120 cells * 1,200 microwells). Ratio of 1:1:3 (HENSM: RCL: BAP(J)) or (Epi-like: PrE/ExEM-like: TE-like) = 28,800 Epi (HENSM) cells , 28,800 Pre/ExEM-like (RCL) cells, and 86,400 TE-like (BAP(J)) cells per each well of a 24-well Aggrewell 400 plate. To start an aggregation, HENSM ESCs/iPSCs (PSCs) on day 3-4 after passaging, RCL day 3 and BAP(j) day 3 cells are needed. BAP(J) are required in bigger amounts since aggregation ratios are 1:1:3 (ES:RCL:BAP(J)), so usually 2-3 X 10c"m plates of BAP(J) primed cells are prepared per 1 X 10c"m plate of HENSM naïve PSCs and 1 X 10c"m plate of RCL primed fraction. Detailed step-by-step protocol for making human SEMs is published on Protocol Exchange[37] and includes a Supplementary Excel sheet macro-template for cell number and ratio calculation. Complete cells mixture after cell counting were prepared as 2x concentration (288,000 cells/ml with 20μM ROCKi. 500μl of cell-mix suspension was gently added drop wise to each well of the AggreWell plate (final yield per each well = 1ml final volume with 10μM final ROCKi concentration and 144,000 cells). The plate was centrifuged at 100g for 3 minutes and incubated at 37 ºC in hypoxic incubator conditions (5% O_2 and 5% CO_2).

Next day (day 1), pre-warm 1 ml of aggregation medium per well for 30-60 min at 37ºC water bath. Remove AggreWell plate from the incubator and observe under the microscope and ensure cells have

started forming aggregates inside the microwells. Remove 800-900 µL of medium gently from each well and carefully and gently add 1ml of pre-warmed aggregation medium to each well and place back the plate in the hypoxia incubator. The same volume of medium is replaced at day 2, and plates returned to 37ºC hypoxia 5% O2 5% CO2 incubator again.

At aggregation day 3, aggregates were gently transferred to 6-well cell suspension non-adherent tissue culture plates (Greiner, 657185) filled with 3 ml of pre-equilibrated hEUCM2 (20% FBS) per well and placed on an orbital shaker rotating at 60 rpm (Thermo Scientific 88881102 + 88881123) located inside a 5% CO2 incubator in 20% O2. The latter delicate step should be done as follows: on Day 3 (72h after aggregation) SEMs are transferred to non-adherent six-well plates: 1-Prepare 3ml of hEUCM2 20% FBS per well of non-adherent six-well plate. 2.- Prewarm in a Normoxia 37ºC 20% O2 5% CO2 incubator for 30-60 min on the 6-well plate. 3- Take out the aggrewell plate with the aggregates, and carefully remove most (nearlly all) of the aggregation medium without disturbing the aggregates and replace with 1 ml of hEUCM 20% taken form the 6-well (the plan is to transfer aggregates from two wells of a 24-well plate to one well of a 6-well plate), and distribute 2ml out of the 3ml hEUCM2 of the 6-well, into two 24-wells (1ml per each well). 4- Using a 3ml sterile Pasteur pipet, with up and down slow movements, harvest the aggregates. 5- Collect back to the 6-well (total volume of media per each well of the non-adherent 6-well plate should be 3ml after finishing these transfers) and incubate the 6-well plate in a 20% O2 5% CO2 37ºC normoxic incubator on top of an orbital shaker at 60 RPM placed within the incubator. Please note that it is important not to significantly deviate from total 3 ml volume per each well of the 6-well plate to avoid clumping of the aggregates.

On day 4, 2 ml of medium were gently removed per well and were replaced with 2 ml of pre-heated (37ºC in water bath) hEUCM2 with 30% FBS. Same procedure was repeated on day 5, refreshing with 2 ml of hEUCM2 with 30% FBS and put back to the same shaker/incubator setting. After 6 days, 2 ml of medium were gently removed per each well of six-well plate, and were replaced with 2 ml of pre-heated hEUCM2 with 50% FBS and placed back on the shaker in the Normoxic incubator conditions. Same procedure was repeated at day 7, and placed back on the shaker in the Normoxic incubator conditions. Cultures were ended at day 8 post-aggregation.

hEUCM2[3], was adapted from EUCM2, and is formulated as follows for human SEM: Advanced DMEM/F12 (GIBCO 21331-020), 1 mM GlutaMAX (GIBCO 35050061), 1% penicillin streptomycin (Biological Industries – Sartorius 03-031-1B), 1x of ITS-X supplement (Thermo Fisher Scientific 51500-056), 8 nM B-estradiol (Sigma-Aldrich, E8875), 200 ng/ml progesterone (Sigma-Aldrich, P0130), 25 µM N-acetyl-L-cysteine (Sigma-Aldrich, A7250), 20-50% FBS (Sigma Aldrich F7524 – heat inactivated and filtered) as indicated in **Fig. 2b**., and optionally extra added 1 mg/mL D(+)-Glucose Monohydrate (J.T. Baker - 0113) (e.g., add 500mg per 500mL media). Culture media was pre-heated for at least an hour at

37ºC water bath. FBS batches are tested and qualified for SEM assay, by expanding V6.5 ES carrying an OCT4-GFP reporter and expanded for 3 passages on Gelatin coated plates and checked for >95% GFP+ signal by FACS analysis. We emphasize that the success of our human SEM protocol described above is not a single FBS batch specific phenomenon, as three out of three tested FBS batches used in the lab from two different vendors over the last two years of this ongoing project yielded consistent results with some differences in efficiency between serum batches ((i) Sigma FBS F7524 LOT#0001654682; (ii) Biological Industries European grade FBS 04-0071A LOT#2004013and (iii) Sigma FBS F7524 LOT#1664377) yielded similar SEMs. To study the development of our human SEMs in further detail and with adequate reference, we mostly relied on the data from the Virtual Human Embryo Project based on the Embryo Carnegie Collection as the most detailed and relevant source to date[20,39].

We critically emphasize that before starting to conduct co-aggregation experiments for human SEM generation as described herein, it is critical that first to test that the RCL and BAP(J) induction protocols are yielding high efficiency induction as can be tested by FACS on day 3 for BAP(J) and day 6 for RCL cells. If RCL induction from naïve HENSM PSCs that were expanded in hypoxia conditions on Matrigel, is not showing a majority of PDGFRa+ cells (>65% PDGFRa+) at day 6, please be reminded that the RCL induction efficiency and quality is mainly dependent on the optimal initial cell seeding confluency prior to induction initiation which can vary between human naïve PSC lines. It is most recommended to make FACS analysis for seeding curve starting with different initial cell numbers plated for induction and conduct FACS analysis on day 6. Also make sure to use fresh RCL medium (less than a week old) and use B27 without insulin. If BAP(J) induction is not giving >85% ENPEP/TACSTD2 double positive TE-like cells at day 3, be reminded that BAP(J) induction quality and efficiency is also mainly dependent on the cell confluency which can vary between human PSC lines. It is most recommended to make FACS analysis for seeding curve starting with different initial cell numbers plated for induction. Revise the cell confluency following TACSTD2/ENPEP FACS analysis on day 3 of the BAP(J) induction regimen. Make sure not to keep BMP4 for more than the 24h (only during the 1st day of the 3 day BAP(J) priming regimen). Once the two latter priming/induction protocols are yielding adequate efficiency as indicated above, then one can proceed to conducting co-aggregation experiments.

In our case, for WIBR3 cell line grown in HENSM on Matrigel coated plates for RCL induction, plating 8x10^5 cells were seeded onto each single 10c”m Gelatin/MEF coated petri dish showed best results regarding RCL induction efficiency outcome at day 6, while in the case of WIBR1 cell line, plating 2x10^6 HENSM PSCs grown on Matrigel coated plates, onto a single Gelatin/MEF coated 10c”m plate, showed optimal RCL induction results measured by FACS on day 6. For BAP(J) priming protocol from HENSM naïve ESCs, from WIBR3 hESC HENSM naïve cells grown on Matrigel, 1x10^6 HENSM PSCs were plated per single 10c"m Matrigel coated plate showed optimal results as measured by FACS on day 3

of BAP(J) regimen. For WIBR1 hESC line, seeding 2x10^6 HENSM PSCs per 10c"m Matrigel coated petri dish yielded optimal BAP(J) results as measured by FACS on day 3.

For initial training and calibration in conducting the human SEM protocol, we recommend stopping and evaluate initial experiments on day 6 and look within the generated aggregates for human SEMs (with amniotic-like cavity, bilaminar disc-like structure, hypoblast-like layer, yolk sac-like compartment and trophectoderm-like outer layer). Then, once confident of success over multiple biological replicates, proceed to later stages up to day 8 of the protocol.

Human blastoid generation and PALLY/PALY conditions

Human blastoids were generated according to Kagawa et al.[15] with few modifications. WIBR3 hESCs were grown in HENSM for at least 3 passages in feeder free conditions (1% Matrigel or Cultrex coated plates) and were used for generating human blastoids. After 3 days of growth of naïve ESC were harvested and counted and 55 cells were seeded per microwell (total of 66,000 cells were seeded per 24-well in 1 ml of medium) AggreWell 400 24-well (Stemcell Technologies cat 34415) in N2B27-BSA aggregation media supplemented with 10μM ROCKi Y27632. The next day medium was changed to PALLY consisting of N2B27 base with 1 μM MEKi/ERKi PD0325901 (Axon Medchem 1408), 1μM TGFRi A83-01 (Axon Medchem A83-01), 1μM LPA (Tocris, 3854), human LIF 10ng/ml and 10μM ROCKi Y2763. This medium was repeated on day 2, but on day 3 medium was changed for LY (1μM LPA (Tocris, 3854) and 10μM ROCKi Y27632) for another 48h. Afterwards human blastoids were manually selected and collected for further analysis. The entire procedure was conducted in 37C, 5% O2 and 5% CO2 conditions. PALY media has the same composition of PALLY but without including LIF.

Mouse SEM generation

Mouse SEM aggregations were made according to Tarazi et al.[3] Mouse animal experiments pertained only to mouse SEM and were performed according to the Animal Protection Guidelines of Weizmann Institute of Science and approved by the following Weizmann Institute IACUC (#01390120-1, 01330120-2, 33520117-2). Mus Musculus (mouse) ICR strain derived embryo samples were used as reference controls for mouse SEM related experiments. 4-10 week old male and female ICR mice were used for timed matings for natural embryo dissection that were used as controls. Mouse aggregation medium was also tested for human SEMs but was found inappropriate. Mouse aggregation medium (mouse AM) consisted of 1x DMEM (GIBCO-41965) supplemented with 20% FBS (Sigma), 1 mM GlutaMAX (GIBCO, 35050061), 1% penicillin streptomycin (Biological Industries – Sartorius 03-031-1B), 1% Sodium Pyruvate (Biological Industries – Sartorius 03-042-1B), 1% non-essential amino acids (Biological Industries – Sartorius 01-340-1B) and 0.1 mM β-mercaptoethanol (Thermo 31350010).

Morphological evaluation of human early embryonic development and efficiency calculations

Assessment of appropriate human development was performed by careful analysis of available *in-utero* histological embryo collections (predominantly Carnegie collection), taking in account different tissue morphology and structure organization through different stages of development. Furthermore, available work on primate development was employed as a reference for anatomical structure[8,40,41], specific markers of each of the compartments was inferred from previous in vitro human development works[6,31], in-vitro differentiation protocols and primate existing databases[8]. Most of human histological descriptions and figures used for this paper are mentioned in the virtual human embryo website (https://www.ehd.org/virtual-human-embryo/)[20] and available human embryology papers[18] and textbooks (Langman, Larsen, Carlson).

All human *in utero* data and figures included in this study were made only from Carnegie collections after obtaining the appropriate copyright approvals. Only SEMs presenting all the previously defined features were considered as properly developed. Percentage of human SEMs generation is calculated based on the number of properly developed structures observed per random fields of view at a specific time point on independent experiments, while relying on immunofluorescence to corroborate the different lineage self-organization. Efficiency quantification was performed for SEMs according to the three criteria below at the indicated time points:

I. Surrounding by Tb-like layer, defined by the expression of CK7 or SDC1 in the perimeter of the SEM.

II. Presence of Epi- and Hb-like compartments, defined by the expression of OCT4 and SOX17, respectively, in groups of cells inside the SEM.

III. Epi-like and Hb-like compartment forming a bilaminar disc-like structure along with the presence of amniotic- and yolk sac-like cavities, defined by the absence of the nuclear immunostaining signal in the central area of the Epi-like and YS-like tissues.

Number of biological replicates (N) and number of samples per biological replicate sampled (n) are indicated in figure legends and/or in figure panels where relevant),

Quantification and Statistical Analysis

Statistical analyses of real time PCRs were performed in QuantStudio software v1.3 and visualized in GraphPad Prism 7. Visualization and statistical analyses of the cell numbers and SEM efficiencies were performed with Python v3.8.5 software using scipyv1.8.0 and seaborn v0.11.0 libraries. Boxplot graphs indicate medians with interquartile ranges, the whiskers mark distribution range. The barplots show average values plus s.d. The dots mark individual numerical values used for visualization of the data distributions and analyses. Significant difference between two samples was evaluated by the two-sided Mann-Whitney

test for non-normally distributed data or two-tailed student t-test as indicated per panel. $p<0.05$ threshold was considered as statistically significant.

Whole-mount immunostaining

Human SEMs were collected using sterile plastic Pasteur-pipettes (Alex-red (Israel) (www.alexred.co.il)), PE-3ml / Size155mm, catalogue number SO P12201) and fixed in 4% PFA EM grade (Electron microscopy sciences, 15,710) in PBS at room temperature (RT), for 1h in glass spot plates (Corning, 722085). Then, SEMs were washed in PBS 3 times for 5 minutes, and permeabilized in PBS with 0.5% Triton X-100 (Sigma, 9002931) /0.1 M glycine (Sigma G7126) for 30 minutes. Blocking was performed in blocking solution (PBS/0.01% TWEEN20 (Sigma, 9005-64-5)/10% normal donkey serum (Jackson ImmunoResearch, 017000121) /0.1%bovine serum albumin (Sigma, A7906)) for 1h at RT, and incubated overnight at RT with primary antibodies, diluted in blocking solution.

Afterwards, SEMs were rinsed 3 times for 5 min each in PBS/0.1% Triton X-100, and incubated with Alexa Fluor (488, 568 and/or 647)-conjugated donkey secondary antibodies (Jackson ImmunoResearch) diluted in blocking solution (1:200) for 2h. The samples were counterstained with DAPI for nucleus (1 mg/mL in PBS) for 10 min and washed with PBS for 5 min 3 times. When membranal staining was required rhodamine phalloidin (Invitrogen, R415) or Wheat Germ Agglutinin (WGA) (Invitrogen, W21404) was added in a 1:200 dilution with the secondary antibodies.

The antibodies and dilutions employed for immunofluorescence were the following: Mouse monoclonal anti-Oct3/4 (clone C-10) (Santa Cruz Cat# SC-5279) 1:100; Rabbit polyclonal anti-Oct3/4 (clone H-134) (Santa Cruz Cat# SC-9081) 1:100; Goat polyclonal anti-Sox17 (R&D Cat# AF1924) 1:100; Rabbit monoclonal anti-Cytokeratin 7 (Abcam Cat# ab181598) 1:200; Rabbit monoclonal anti-Cytokeratin 7 (Abcam Cat# ab68459) 1:200; Goat polyclonal anti-Gata3 (R&D Cat# AF2605) 1:100; Rabbit monoclonal anti-Syndecan1 (Abcam Cat# ab128936) 1:400; Mouse monoclonal anti-Cdx2 (Biogenex Cat# MU392A-UC) 1:200; Rabbit monoclonal anti-Phospho-Ezrin (Cell Signaling Cat# 3726) 1:400; Rabbit monoclonal anti-Brachyury(D2Z3J) (Cell Signaling Cat# 81694) 1:100; Goat polyclonal anti-Cer1 (R&D Cat# AF1075) 1:100; Rabbit monoclonal Nanog (Abcam Cat# ab109250) 1:100; Mouse monoclonal anti-PKC zeta Antibody (H-1) (Santa Cruz Cat# SC-17781) 1:200; Mouse monoclonal anti-Podocalaxyn [clone 222328] (R&D Cat# MAB1658) 1:200; Rabbit polyclonal anti-Gata4 (Abcam Cat# ab84593) 1:100; Mouse monoclonal anti-Vimentin (Abcam Cat# ab8978) 1:100; Rabbit monoclonal anti-BST2/Tetherin antibody [EPR20202-150] (Abcam Cat# ab243230) 1:100; Rabbit monoclonal anti-hCG beta [5H4-E2] (Abcam Cat# ab9582) 1:200; Rabbit monoclonal anti-Gata6 (clone D61E4) (Cell Signaling Cat# 5951) 1:100; Rabbit monoclonal anti-Islet1 [EP4182] (Abcam Cat# ab109517) 1:100; Mouse monoclonal anti- Anti-TFAP2a (AP-2α) (3B5) (Santa Cruz Cat# SC-12726) 1:100; Goat polyclonal anti-Sox2 (R&D Cat#

AF2018) 1:200; Rabbit polyclonal anti-Dnmt3l (Imgene Cat# IMG-6804A) 1:100; Goat polyclonal anti-Otx2 (R&D Cat# AF1979) 1:200; Mouse molonclonal anti-Stella (D-5 clone) (Santa Cruz Cat# SC-376862) 1:100; Rabbit monoclonal anti-Blimp1/PDRI-BF1 [Clone C14A4] (Cell Signaling Cat# 9115) 1:100; Goat polyclonal anti-FoxF1 (R&D Cat# AF4798) 1:100.

Immunofluorescence

Cells were fixed in 4% paraformaldehyde in PBS at RT for 10 minutes. Samples were then washed 3 times in PBS, permeabilized in PBST (PBS with 0.1% Triton X-100) for 10 min, blocked in PBS/0.05% Tween/5% fetal bovine serum/1% bovine serum albumin for 1h and incubated with primary antibodies diluted in blocking solution at 4°C overnight. Subsequently, cells were washed in PBS/0.05% Tween (three times, 5 min each) and incubated with Alexa Fluor (488, 568 and/or 647)-conjugated secondary antibodies (Jackson ImmunoResearch) diluted in blocking solution (1:200). Samples were counterstained with 1 µg/ml DAPI for 10 min at RT, washed with PBS three times (5 min each) and mounted with Shandon Immuno-Mount (Thermo Scientific) or PBS.

The antibodies and dilutions employed for cell immunofluorescence were the following: Rabbit monoclonal anti-BST2/Tetherin antibody [EPR20202-150] (Abcam Cat# ab243230) 1:100; Goat polyclonal anti-Sox17 (R&D Cat# AF1924) 1:100; Rabbit polyclonal anti-Gata4 (Abcam Cat# ab84593) 1:100; Goat polyclonal anti-FoxF1 (R&D Cat# AF4798) 1:100; Goat polyclonal anti-Gata3 (R&D Cat# AF2605) 1:100; Mouse monoclonal anti-Cdx2 (Biogenex Cat# MU392A-UC) 1:200; Goat monoclonal Tfap2c (R&D Cat# AF5059) 1:200; Rabbit monoclonal anti-Syndecan1 (Abcam Cat# ab128936) 1:400; Rabbit monoclonal anti-hCG beta [5H4-E2] (Abcam Cat# ab9582) 1:200; Rabbit monoclonal anti-Cytokeratin 7 (Abcam Cat# ab68459) 1:200; Mouse monoclonal anti-Vimentin (Abcam Cat# ab8978) 1:100; Mouse monoclonal anti-Oct3/4 (clone C-10) (Santa Cruz Cat# SC-5279) 1:100; Rabbit polyclonal anti-Oct3/4 (clone H-134) (Santa Cruz Cat# SC-9081) 1:100; Goat polyclonal Gata6 (R&D Cat# AF1700) 1:200; Rabbit monoclonal Gata6 (Cell signaling Cat# 5951) 1:200; Goat polyclonal Nidogen2 (R&D Cat# AF3385) 1:100; Rabbit monoclonal anti-Gata2 [EPR2822] (Abcam Cat# ab109241) 1:200.

Flow cytometry

Flow cytometry analysis were done on a BD FACS-Aria III. Cells were harvested with TrypLE and washed once with PBS afterwards they incubated for half an hour with conjugated primary antibodies (5ul) on 100ul PBS/0.5% BSA. The primary antibodies used are as follows: Mouse monoclonal TROP2-488 labeled (R&D Cat# FAB650G); Mouse monoclonal TROP2-PE labeled (R&D Cat# FAB60P); Mouse monoclonal CD249 (ENPEP)-BV421 labeled (BD Cat# 744872); Rat monoclonal anti mouse CD140a (PDFGR-a)-PE/Cy7 labeled (BioLegend Cat# 135912); Mouse monoclonal anti human CD140a (PDFGR-

a)-PE/Cy7 labeled (BioLegend Cat# 323508); Mouse monoclonal anti human CD140a (PDFGRa)-APC labeled (BioLegend Cat# 323512). FSC and SSC singlets were gated, and only single cells were considering for all analyses. An unstained control was employed to determine the negative/positive populations for all antibodies, ensuring that 100% of the unstained population was allocated on the negative area of the histogram/dot plot**.** Flow cytometry data was analyzed using FlowJo v10.7. **Supplementary Fig. 17** demonstrates FACS gating strategies used in this study.

RNA extraction & RT-PCR analysis

Total RNA was isolated using RNeasy mini kit (Qiagen) following manufacturer instructions. 1 µg of total RNA was reverse transcribed using a High-Capacity Reverse Transcription Kit (Applied Biosystems). RT-PCR was performed in triplicate technical wells for each of the sample included per each gene, using SYBR Green PCR Master Mix (Qiagen) and run on Viia7 platform (Applied Biosystems). Values were normalized to Actin and/or Gapdh and/or HPRT and/or RPL3 across all experiments, data presented as relative expression compared reference sample using ΔΔCT method, samples were visualized using Prism version 7 plotting mean value with s.d.. RT- PCR primer list is shown in **Supplementary Table 3**.

Confocal microscopy

The immunofluorescence images were acquired using Zeiss LSM 700, as well as Zeiss LSM 800 inverted confocal microscopes, both equipped with 405 nm, 488 nm, 555 nm and 635 nm solid state lasers, using a Plan-Apochromat 20× air objective (numerical aperture 0.8) or an EC Plan Neofluar 10× air objective (numerical aperture 0.3). Images of the trophoblast cell surface were acquired with C-Apochromat 40× water objective (numerical aperture 1.2) using LSM 800. For a detailed description of the imaging parameters, see **Supplementary Table 4.** Confocal 3D images and maximum intensity projections were processed using Fiji version 1.52p or 1.53t [42], Zen 2 blue edition software 2011 or ZEN Z3.5 (Zeiss), and Adobe Illustrator 2023 CC.

Light-sheet microscopy

The immunofluorescence images were acquired using Zeiss Z7 light-sheet microscope, equipped with 405 nm, 488 nm, 561 nm, and 638 nm lasers, using a single water 20x Plan-Apochromat (numerical aperture 1.0) detection objective (Zeiss) and two air 10x Plan-Apochromat (numerical aperture 0.2) illumination objectives (Zeiss). Prior to imaging, the sample was mounted in a glass capillary filled with 1% low-melting temperature agarose. Upon solidification, the agarose was pulled out of the capillary with a custom plunger to hang into the imaging chamber filled with PBS.

The single sample was imaged at a time from several angles using Multiview acquisition. Light-sheet volumes along the Z-axis were acquired in a dual scanning mode, using a pivot scan. Light-sheet thickness was set to 3.77 µm, and laser power in the 1 – 50% range was applied. Frame size, 1920x1920 px, exposure time, 50 msec. The light-sheets for left- and right-side illuminations were adjusted independently inside the sample volume for each channel based on the signal intensity in the focal plane of the detection lens. See also **Supplementary Table 4.**

Light-sheet image processing was performed in ZEN 3.5 software. Dual side images were fused based on the maximum intensity signal. Multiview fusion was performed using interactive registration of the brightest channel (typically DAPI) or each channel independently in front and side views. The 'Blending' parameter was typically set to 50 and the intensities of the fused images were averaged ("Method: Mean Fusion"). The image deconvolution was applied for single-view images prior to their fusion in the "Fast Iterative" or "Constrained Iterative" settings.

Confocal Imaging of human SEMs/aggregates

To assess the quality of the experiment and select putative SEMs for further imaging, an overview of the majority of the SEMs is gathered with a tiled scanning implemented in Zeiss LSM 700 and LSM 800 inverted confocal microscopes. SEM(s) were mounted in a 35 mm glass bottom dish (Mattek, P35G-1.5-14-C) covered with PBS. To generate the overview images, a 10× EC Plan Neofluar air objective (0.3 NA) or 20x Plan-Apochromat air objective (1.0 NA) were employed using the Zen software black edition (ZEISS).

Confocal imaging of multiple SEMs/aggregates: 1- After mounting the sample, the image acquisition parameters were established for each of the assessed wavelengths (405, 488, 568, 647), according to the used secondary antibodies. 2- Tiled images were obtained to sample a significant portion or most of the aggregates. 3- Since aggregates are dispersed in different directions and angles, Z-stacks were used to better understand their structure. 4- The efficiency of the experiment can be calculated after gathering the overview image (see below). In addition, the images of individual SEMs/aggregates can be taken by movement of the objective to the recorded positions within a tiled scanning. 5- For light-sheet microscopy, the SEMs were examined and picked with a mouth pipette (aspirator tube; Sigma, A5177), connected to a thinned glass capillary pulled from a glass microliter capillary (Blaubrand intraMark 708744) with an inner diameter above the diameter of the SEM. Under the binocular view, the SEM of interest is taken from the imaging plate while trying to avoid disturbing other aggregates and without generating air bubbles. The picked SEM is placed in a new drop of PBS on a petri dish for further analysis.

Confocal imaging of individual SEM/aggregate: 1- The previously collected overview images of the experiment are used to select the SEMs based on the microscope stage position. Importantly, to maintain the same stage position, the SEMs should not be moved or disturbed during the entire imaging process. Hence, for every single imaging session, the new overview image is required. 2- Redirect the stage to one of the chosen positions, and switch to a higher magnification objective and recalibrate acquisition parameters for each channel. 3- Use the digital zoom and rotation to center the SEM as desired. Importantly, ensure the entire SEM is fit into the imaging frame when examining morphology.4- If 3D volumes are aimed, set the beginning and end frames to cover the SEM and select the slice spacing according to the desired sampling along the Z-axis. 5- Optionally, use the laser intensity adjustment (implemented in Zeiss LSM) to reduce decay of the signal along the Z-axis. 6- Acquire the image.

Quantification of the experiment efficiency: Quantification of the efficiency of the experiment is performed based on the overview tile scanned images (see "Confocal imaging of multiple SEMs/aggregates"). The cell-type markers for immunostaining should be selected prior to the experiment depending on the defined efficiency criteria. For assessing the efficiency and quality of the experiments, we recommend using markers for the main three lineages: epiblast-like (OCT4, SOX2), yolk-sac-like (SOX17), ExEM-like (VIM1 and FOXF1) and trophoblast-like (GATA3, CK7 or SDC1). Afterwards, the following steps were taken: 1- Using FIJI software, open the previously acquired overview image. 2- Assign adequate colors to each channel. 3- Using the Cell Counter tool, manually count the total number of aggregates that entirely fit into the image. 4- Subsequently, select the aggregates that meet the desired criteria using another Cell Counter tool. Calculate the efficiency of the experiment by dividing the number of adequate SEMs by the total number of structures. Multiply by 100 to get percentage values.

Picking human SEMs under the confocal microscope:

To select stained SEMs for further analysis and imaging, picking them with the help of the confocal microscope while imaging is our current best solution, since morphology of human SEMs is difficult to discriminate by brightfield and long training is required: 1- Using the confocal microscope localize the chosen SEM (the previously gathered overview can help to guide the stage on the right direction). 2- Using the live view, make sure to have selected the right SEM. 3- Using the binocular view and a mouth pipette with a previously pulled glass capillary, the structure is taken from the plate while trying to avoid disturbing other structures as less as possible and without generating bubbles. 4- The picked SEM is placed in a new drop of PBS on a petri dish for further analysis.

Electronically controlled *ex utero* roller culture platform

Human SEMs can be kept in the *ex utero* electronically controlled roller culture platform after day 6 which provides continuous flow of oxygenating gas[43]. The system consists of an electronic gas modulating

unit (Designed by Jacob H. Hanna and assembled and sold by Arad Technologies, Ashdod, Israel – Hanna lab model #1) adapted to the roller culture unit from B.T.C. Engineering, – Cullum Starr Precision Engineering Ltd - UK), as previously described[3,43]. On day 7, all human SEMs from one well of the 6-well plate were picked and transferred to glass culture bottles (50-100 aggregates per bottle) containing 4 mL of fresh hEUCM2 50% FBS. The bottles were placed on the rolling culture system, rotating at 30 revolutions per minute at 37°C, and continuously gassed with an atmosphere of 21% O_2, 5% CO_2 at 6.5-8 pounds per square inch (psi), yielding gas input into the humidifier bottle of (1L/min). Bottles were kept inside glass culture bottles rotating on a spinning wheel allocated inside a "precision" incubator system (BTC01 model with gas bubbler kit - by B.T.C. Engineering, – Cullum Starr Precision Engineering Ltd - UK) (BTC 04). Gas flows into the gas mixing box at 0.5psi, and from the gas mix box through the inlet into the humidifier water bottle (designed by Jacob H. Hanna and manufactured as a modifier for the incubator part by Arad Technologies. Ltd. and this upgrade is absolutely essential for the incubator set-up) at 6.5-8 psi yielding has input flow to the humidifier of 1L/min, and then to the inside of the bottles in the rotating drum. The bubble rate was adjusted using the valve on the lid of the water humidifier bottle to the first point where continuous bubbling is observed, which generally corresponds to 0.06-0.1 psi output after the humidifier water bottle, yielding ~60-100mL/min gas flow output from the humidifier bottle. The rate of bubbles created inside an outlet-test tube filled with water is used to generate gas output at 0.06-0.1 psi and gas flow ~60-100mL/min into the wheel compartment. A black cloth or large diaper was used cover the incubator to provide protection against phototoxicity. 2 ml of hEUCM2 with 50% FBS was used in the roller culture for human SEMs at these stages.

Chromium 10X single cell RNA sequencing

To further validate and examine the milieu of cell types present in the human SEMs generated herein in a more unbiased manner, we performed a single cell transcriptomic analysis by Chromium 10X scRNA-seq. Human SEMs grown ex utero were manually selected based on morphological criteria that matched representative structures shown in **Fig. 2c** between day 4-8 and harvested for single cell RNA sequencing (**Supplementary Table 1**) at day 4, 6 and 8 (from two biological replicates run in parallel to reduce intrinsic variability per each time point), using the Chromium Next GEM Single Cell 3' platform (V3.1). All human SEMs analyzed by scRNA-seq were generated by co-aggregating WT WIBR3 naïve ESC lines grown in HENSM with RCL-induced or BAP(J)-induced WT cells. At day 4, ~80 human SEMs were pooled into one lane of the 10X chromium chip, while at day 6 and 8, a pool of ~50 SEMs were sequenced per lane. All SEMs samples were processed including extraembryonic compartments without any dissection. SEMs were dissociated with Trypsin-EDTA solution C (0.05%) for 10 minutes (Biological Industries; 030501B). Trypsin was neutralized using media with 10% FBS, and cells were washed and

resuspended in 1x PBS with 400 µg/ml BSA. Cell suspension was filtered with a 100µm cell strainer to remove cell clumps. Cell viability of at least 90% was determined by trypan blue staining for all samples. Cells were diluted at a final concentration of 1000 cells/µL in 1x PBS with 400µg/ml BSA. scRNA-seq libraries were generated using the 10x Genomics Chromium v3.1 Dual Index system (5000 cell target cell recovery) and sequenced using Illumina NovaSeq 6000 platform according to the manufacturer's instructions.

10X Single cell RNA-seq analysis for SEM, RCL and BAP(J) samples

10x Genomics data analysis was performed using Cell Ranger 7.1.0 software (10x Genomics) for pre-processing of raw sequencing data, and Seurat 4.3.0 for downstream analysis. To filter out low-expressing single cells, possible doublets produced during the 10x sample processing, or single cells with extensive mitochondrial expression, we filtered out cells with under 1000 expressing genes, over 8,000 expressing genes and over 15% mitochondrial gene expression. We analyzed ~4,000-8000 cells from pooled high-quality SEMs. After quality control and strict filtering, a total of 12,190 single cells were used for subsequent analyses (**Supplementary Fig. 15a-c**). Seurat integrated analysis and anchoring of all individual samples was performed and then normalized by log-normalization using a scale-factor of 10,000. The top 2,000 variable genes were identified by the variance stabilizing transformation method, and subsequently scaled and centered. Principal components analysis was performed for dimensional examination using the 'elbow' method. The first 10 dimensions showed the majority of data variability. Therefore, UMAP dimensional reduction was performed on the first 10 dimensions in all samples. Clusters were detected using Seurat Find Clusters function, with resolution parameter =0.5. Dot-plot describing expression and prevalence of specific genes was generated using Seurat DotPlot() function. Projection of selected genes on SEM UMAP was generated with Seurat FeaturePlot() function. Heatmaps were generated with Seurat DoHeatmap() function, or with R pheatmap package (v1.0.12). Analysis Code is available in GitHub https://github.com/hannalab/Human_SEM_scAnalysis. We note that while ExEM-like cell cluster 8 cells express CER1, DKK1 and LHX1 AVE markers, they lack other key VE/AVE markers like SOX17 and APOA1 and thus were not annotated as AVE-like, but rather as ExEM-like cells as they predominantly express mesenchymal signature (**Extended Data Fig. 13a)**. The latter is consistent with CER1, DKK1 and LHX1 being co-expressed also in extra-embryonic mesodermal cells as detected in primates[8] and gastrulating human embryo datasets [32]. The fact that PGC-like cells did not create their own cluster or subcluster likely results from their relative scarcity within SEMs, as was observed in mouse SEMs[3].

Integration of scRNA-seq of the RCL starting population with published scRNA-seq day 6 primitive endoderm RACL based conversion[9] **(Extended Data Fig. 1f)** was performed using the Seurat v3 integration standard workflow[44,45]. Prior to integration, datasets were normalized, and the top 2000 most

variable genes were selected. Integration anchors were identified using the FindIntegrationAnchors function with default arguments, incorporating all available parameters and data across features. An integration based Uniform Manifold Approximation and Projection (UMAP) was constructed using the runUMAP function with dimensions ranging from 1 to 10 for visualization. Similar analyses were performed to integrate scRNA-seq of the BAP(J) starting population with naive trophectoderm and naive cytotrophoblasts[14] **(Extended Data Fig. 3g)** and scRNA-seq of SEM ExEM-like cell populations with the previously published time course dataset of naive to TSC/ExEM dataset [9] **(Extended Data Fig. 13f)**.

HENSM Naïve and primed 10X Single-cell Multiomics Library prep, sequencing and computational analysis

HENSM scRNA-seq **(Supplementary Fig. 3)** was processed alongside scATAC-seq measured from the same cells (multimodal strategy), in a protocol described below: LIS49 and WIBR3 hESCs were cultured for 3 passages in HENSM and primed conditions on Matrigel coated plates. Cell nuclei were isolated using the protocol provided by 10X genomics (cat CG000365), aiming for ~6000 nuclei. 10X Genomic library preparation was performed via the protocol provided by 10X genomics (cat CG000338) and sequenced by 2 units of NovaSeq SP sequencing system (100 cycles). For computational analysis, sequencing output were demultiplexed using Cell-Ranger-arc V2.0 (10X Genomics) mkfastq command, and counts were estimated with Cell-Ranger count software. ATAC and RNA data pre-processed separately and filtered with Seurat 4.0 R package. For ATAC data, cells with very high (>10000) or very low (<1000) number of overlapping peaks were filtered out, as well as cells with high (>3) nucleosome signal score, or low TSS enrichment score (<2). For RNA data, cells with low number of features (bottom 10%), high number of UMIs (top 10%) or high mitochondrial percentage (top 10%) were filtered out. Doublet cells were also assessed by RNA data using scDblFinder V1.6. Cells that passed both RNA and ATAC filtering parameters were further examined together. Seurat integrated analysis and anchoring of RNA data was performed and then normalized by log-normalization using a scale-factor of 10,000. The top 5,000 variable genes were identified by the variance stabilizing transformation method, and subsequently scaled and centered. UMAP was performed on the first 30 PCs. ATAC data was normalized by term frequency inverse document frequency (TF-IDF) and partial singular value decomposition (SVD) was performed using RunSVD by Signac v1.6.0. Batch effect correction was performed using RunHarmony by harmony package (https://portals.broadinstitute.org/harmony/index.html), while the cell line type was excluded from the correction and reduction was based on latent semantic indexing (LSI) produce by SVD. UMAP was performed on the harmony v3 corrected 2:50 dimensions. Construction of the weighted nearest neighbor was perform using FindMultiModalNeighbors by Seurat based on RNA PC1:30 and ATAC harmony 2:50.

Pseudotime Analysis

Pseudotime analysis on scRNA-seq data, was done using monocle3 R package v1.3.1 over the relevant Seurat objects. Heatmap of top differentially expressed genes was presented with the order determined by the pseudotime, using pheatmap R package v1.0.12.

Projection on Human Embryo Reference Compendium

The human embryo reference[26] was built by integrating previously published datasets consisting of 6 human embryonic data sets spanning early in vitro cultured human blastocysts[27], 3D- in vitro cultured human blastocysts until pre-gastrulation stages[28–30], and a Carnegie Stage 7 (CS7) 16-19 dpf human gastrula [32], using fastMNN from batchelor (v 1.6.2)[46] as recently described[26]. The raw counts for cells of human SEMs were aggregated within neighborhood nodes as calculated by Milo[47] resulting in 945 representative neighborhoods, followed by projecting the summed counts matrix onto the assembled human embryo reference using MNN (mutual nearest neighbor correction)[46], followed by stabilized UMAP projection using the umap_transform function from R package uwot (v 0.1.14)[48] (**Fig. 6e**). MILO neighborhoods for projects was chosen (rather than scMAP tool to map single-cells onto existing datasets) given the noise, and after verifying that the neighborhoods are approximately homogenous in their cell-type composition. Indeed, we have checked the assigned cluster information for representative neighborhoods and related neighborhoods. As shown in **Supplementary Fig. 16a**, 97.5% and 83.7% of the related neighborhoods have the same cell identity with representative neighborhoods based on lineage information and Seurat cluster information, respectively. Thus, validating that the neighborhoods are highly homogenous in their cell-type composition. Alluvial plot comparing the cell-type annotations of representative node SEM cells to the predicted identities obtained from scMAP[31,49] (**Supplementary Fig. 16b)**. The prediction results for SEM cells (after aggregation using neighborhood methods) were shown here. The majority of SEM-Epiblast-, SEM-STb- and SEM-YS/Hb-like cells were identified. However, the SEM-Amnion-like and SEM-ExEM-like cells were determined as unassigned which is likely due to the low number of embryonic reference cells for these cell types and the relatively low sequencing depth in Tyser et al.[32]

Mouse Extra-embryonic annotation analysis

Rhox5 positive cells (>1 counts) were chosen for the analysis of extra-embryonic tissue. The cells were annotated based on marker genes as previously conducted in[50], such that if at least 4 markers (3 in the case of SpA-TGC and SpT-Gly) were expressed (>0 counts), the cell was annotated in that category. 26% of the annotated cells, were annotated by multiple categories. The markers are as following: Chorion (Irx4, Esx1, Id1, Id3, Phlda2, Klhl13), Chorion progenitors (Sox3, Dusp6, Nat8l, Bmp4, Sox2, Esrrb, Eomes),

Intermediate Chorion (Ascl2, Fgfr2, Cited1, Gjb3, Ndrg1, Irx2, Irx3), uncommitted EPC (Chsy1, Gjb3, Krt19, Lgals1, Cald1, Ctsl), SpA-TGC (Ctla2a, Pecam1, Ramp3, Igfbp7, Nos3), SpT-Gly (Dlx3, Car2, Ncam1, Pcdh12, Tpbpa), TGC-progenitors (Adm, Fosl1, Hand1, Trpm5, Maged2, Prl5a1), p-TGC (Star, Serpinb9d, Hsd3b6, Rhox6, Cts7) as in[50]. R ggplot was used to generated scatter plot, along with geom_smooth(method="lm").

Mouse and Human IGV Analysis

Bulk ATAC-seq and RNA-seq profiles in the proximity of selected genes (GATA3, GATA4, GATA6) are presented using Broad IGV genome browser v2.16.2. Mouse datasets were taken from published datasets[51–53] and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181053 and so were human datasets[12 4,54]. Mouse enhancers were taken from[52], human enhancers were taken from GeneHancer[55]. Open regions that overlap with promoter or exon were excluded from the analysis. Potential enhancers in these regions were manually curated.

Data availability

All newly generated scRNA-seq and 10x Chromium Single Cell Multiome ATAC + Gene Expression data are deposited under GEO: GSE239932. GSE numbers and references are indicated for all other previously published and publicly available scRNA-seq and ATAC-seq data are indicated. Any other data is available upon request. All other information required to reanalyze the data reported in this work is available upon request from the corresponding author. Source data are provided with this paper.

Code availability

The custom code generated in this study is provided at GitHub: https://github.com/hannalab/Human_SEM_scAnalysis. The custom code was not essential to the main conclusions of this study.

Additional References

35. Lengner, C. J. *et al.* Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* **141**, 872–883 (2010).
36. Hyun, I. *et al.* ISSCR guidelines for the transfer of human pluripotent stem cells and their direct derivatives into animal hosts. *Stem Cell Reports* **16**, 1409–1415 (2021).
37. Oldak B, Wildschutz E. & Hanna, J. H.. Generation of a Complete Integrated Human Post-Implantation Structured Embryo Model (SEM) Derived Ex Utero Solely from Genetically Unmodified Naïve (HENSM) Pluripotent Stem Cells. Protocol Exchange (2023). DOI: 10.21203/rs.3.pex-2268/v1

38. Io, S. *et al.* Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell* **28**, 1023-1039.e13 (2021).
39. De Bakker, B. S. *et al.* An interactive three-dimensional digital atlas and quantitative database of human development. *Science (1979)* **354**, (2016).
40. Ma, H. *et al.* In vitro culture of cynomolgus monkey embryos beyond early gastrulation. *Science* **366**, (2019).
41. Niu, Y. *et al.* Dissecting primate early post-implantation development using long-term in vitro embryo culture. *Science (1979)* **366**, (2019).
42. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* vol. 9 676–682 Preprint at https://doi.org/10.1038/nmeth.2019 (2012).
43. Aguilera-Castrejon, A. *et al.* Ex utero mouse embryogenesis from pre-gastrulation to late organogenesis. *Nature* **593**, 119-124 (2021).
44. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
45. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411–420 (2018).
46. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology* **36**, 421–427 (2018).
47. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol* **40**, 245–253 (2021).
48. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
49. Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: projection of single-cell RNA-seq data across data sets. *Nature Methods 2018 15:5* **15**, 359–362 (2018).
50. Lau, K. Y. C. *et al.* Mouse embryo model derived exclusively from embryonic stem cells undergoes neurulation and heart development. *Cell Stem Cell* **29**, 1445-1458.e8 (2022).
51. Sheban, D. *et al.* SUMOylation of linker histone H1 drives chromatin condensation and restriction of embryonic cell fate identity. *Mol Cell* **82**, 106-122.e9 (2022).
52. Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).
53. Hada, M. *et al.* Highly rigid H3.1/H3.2-H3K9me3 domains set a barrier for cell fate reprogramming in trophoblast stem cells. *Genes Dev* **36**, 84–102 (2022).
54. Wu, J. *et al.* Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).
55. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, 1–17 (2017).
56. Lau, K. Y. C. *et al.* Mouse embryo model derived exclusively from embryonic stem cells undergoes neurulation and heart development. *Cell Stem Cell* **29**, 1445-1458.e8 (2022).

Extended Data Figure Legends

Extended Data Figure 1. Optimization of extra embryonic lineage induction using transient overexpression of GATA4 and GATA6. a, representative Flow Cytometry (FACS) plots of Pdgfra-PE/Cy7 marking primitive endoderm (PrE)-like cells priming from mouse embryonic stem cells (ESC) using iGata4 with DOX for 48 hours **(right)** versus the control condition without DOX **(left)**. **b**, representative FACS plots of PDGFRA-APC for putative PrE/ExEM-like priming from human naïve ESCs (nESCs) using iGATA6 with DOX in different media (N2B27 and HENSM) as indicated. **c**, quantification of the PDGFRa+ population by FACS analysis among the PrE/ExEM-like cell optimization conditions presented in this study (number of biological replicates is indicated for each condition). Average values and s.d. error bars are shown. Two-sided Student t-test *p* values are indicated where relevant. >40% PDGFRa+ set as a threshold for follow-up, followed by extended characterization and validation. **d**, representative FACS plots of PDGFRa-APC for putative PrE/ExEM-like priming/induction from human nESCs after three days in RCL medium (conventional 2D conditions) followed by three days of RCL **(left)** or basal N2B27 **(right)** in aggregation setting. **e**, representative RT-qPCR gene expression (normalized by GAPDH and ACTIN) of the endodermal marker genes, in RCL (yellow), RACL (blue), NACL (dark blue), and basal N2B27 (grey) media conditions versus naïve PCSs used as a control (white). PrE/ExEM and definitive endoderm (DE)-specific genes are separately underlined. Bar plot based panel showing the average value of each sample (which represents average value of 3 technical replicates), error bars indicate s.d.. A single representative experiment out of N=3 biological replicates performed is shown. **f**, integration of day 3 RCL starting cell population with Pham et al. 2021 ExEM conversion dataset[10]. UMAPs of RCL starting cell population integrated with the published reference dataset of day 6 RACL conversion protocol from naïve ESCs containing PrE cells, ExEM, and intermediate epiblast. Selected cell type annotations are shown.

Extended Data Figure 2. Optimization of PrE/ExEM-like cells priming protocol from human HENSM naïve ESCs. a, representative immunofluorescence images of WT WIBR3 (W3) nESCs induced in RCL media for 6 days, showing expression of GATA4 (cyan), SOX17 (yellow), and BST2 (red); nuclei (DAPI, white). GATA4 marks both SOX17+ PrE-like and BST2+ ExEM-like populations (see also **Fig. 1e**). This panel is an extended version of **Fig. 1e** but showing staining patterns of more markers. Scale bar, 100 μm. **b**, representative immunofluorescence images of iGATA6 nESCs induced in RCL media for 6 days (with or without DOX), showing expression of SOX17 (yellow) and BST2 (red). In both set-ups, $BST2^+$ (ExEM-like) and $SOX17^+$ (PrE-like) cell populations have a mutually exclusive expression pattern. Scale bar, 200 μm. **c**, representative immunofluorescence images of WT nESCs in RCL media for 6 days. FOXF1 (yellow), BST2 (red), GATA4 (cyan), nuclei (DAPI, white). $BST2^+$ cells are also

GATA4^{+}/FOXF1^{+}, excluding the possibility that they represent residual pluripotent cells in the RCL induced cultures from HENSM ESCs. Scale bar, 100 μm. Bottom, 4x zoom, scale bar, 25 μm.

Extended Data Figure 3. Testing SEM aggregation conditions with human conventional trophoblast stem cell lines and evaluating trophoblast induction with transient overexpression of GATA3 in human naïve ESCs. a, proportion of the indicated cell types among extra-embryonic cells of mouse natural embryos grown *in utero* or *ex utero* and in day 8 mouse SEMs generated from iCdx2 mouse naïve ESCs (and not embryo derived mouse TSC lines)[3]. Three pooled samples are presented: *in utero* natural embryos (n = 2401 cells), *ex utero* natural embryos (n = 1382), and mouse iCdx2 day 8 SEMs (n = 6249). The cell types: Chorion, Intermediate-Chorion, Chorion Progenitors, Uncommitted Ectoplacental-Cone Cells (EPC), Trophoblast Giant Cells (TGC) progenitors, parietal trophoblast giant cells (pTGC), spiral artery associated trophoblast giant cells (SpA-TGC), and junctional zone spongiotrophoblast cells (SpT-Gly) based on previously published similar analysis and annotations[56]. **b**, frequencies of the cell types presented in **(a)**, showing a significant reduction of TGC-progenitors and pTGCs in *ex utero* embryos (natural and SEM), compared to *in utero* embryos. Blue line represents the linear function f(x)=x; the Shaded area represents 95% confidence interval. This analysis confirmed that mouse naïve ESCs derived TSC lineage following Cdx2 overexpression under optimized conditions[3] can contribute to both the chorionic and ectoplacental cone-like lineages in mouse SEMs generated exclusively from mouse naïve ESCs. **c (top)**, scheme of the aggregation protocol for epiblast (naïve hESCs in HENSM media) and PrE/ExEM-like cells with the validated human TSC line derived from human primed ESCs (termed pTSC) and expressing tdTomato. **c (bottom)**, representative brightfield images and live fluorescence of tdTomato (red) in aggregates with labeled TSCs; scale bar, 200 μm. Right, zoom into the several SEMs with tdTomato signal; scale bar, 50 μm. **d,** representative immunofluorescence images showing different patterns of expression of trophoblast marker genes in the wild type (WT) nESCs incubated in BAP(J) media for three days **(top)** versus iGATA3 cells, induced by DOX in BAP(J) media **(bottom)**. GATA3 (magenta), TFAP2C (magenta), GATA2, CDX2, SDC1, CK7, and HCGB (all in green), nuclei (DAPI, blue). Scale bars, 100 μm. **e**, representative FACS plots of ENPEP versus TACSTD2 for trophoblast (Tb)-like cell priming using iGATA3 induction in different media (AP(J) with DOX, BAP(J) with DOX), and using WT nESC priming to trophectoderm in AP(J) and BAP(J) regimens. Percentage of double positive population is indicated. **f**, quantification of the TACSTD2-positive population (%) across conditions in **(e),** n = 3 biological replicates; average value and s.d. error bars are indicated per condition. Two-sided Student's t-test *p* values are indicated where relevant. **g,** integration of BAP(J) starting population with previously published in vitro differentiation derived (Io et al. 2021) naive trophectoderm conversion data. UMAP of Day 3 BAP(J) starting cell population derived from HENSM were integrated with the previously published reference

dataset[14] of naive trophectoderm (nTE) and naive cytotrophoblast (nCT) following BAP(J) regimen. Selected cell type annotations are shown.

Extended Data Figure 4. Enhancer accessibility of extra-embryonic lineage master regulators GATA3, GATA4 and GATA6 in human but not mouse naïve ESCs. a, ATAC-seq and RNA-seq of GATA3, GATA4 and GATA6 genes in human vs. mouse, as measured in naïve ESCs and in multiple differentiated cell types (as indicated). HENSM conditions were used for human naïve ESCs[4]. Known enhancers are marked in grey bars at the bottom. Putative enhancers are marked in red or green: (1) green indicates potential enhancers that are already open in naïve stem cells, (2) red indicates enhancers that are closed in naïve pluripotent stem cells, but open in at least one differentiated cell state. Putative enhancers in the approximate regions of the genes were manually selected. Open regions that overlap with promoter or exon were excluded from the analysis. RNA-seq of the indicated genes in naïve ESCs are shown. **b,** immunofluorescence images of HENSM ESCs cultured for 3 days in N2B27 basal medium on Matrigel-coated palates. OCT4 marks pluripotent cells (red); VIM, FOXF1, and BST2 mark the ExEM-like cells (all in cyan); SOX17 and GATA6 mark PrE- and PrE/ExEm-like cells, respectively (yellow). GATA3 marks trophoblast-like cells (magenta), nuclei (DAPI, white). Scale bar, 200 μm.

Extended Data Figure 5. Evaluating SEM aggregation by omitting one of the lineages. a, scheme of the aggregation experiment (see Methods) where one of the lineages was omitted to evaluate its contribution to the resulting SEM morphology. **b (from top to bottom)**, representative immunofluorescence images of day 6 SEM aggregates from the three lineages (control, 120 cells), no HENSM (1:3 RCL: BAP(J), 96 cells), no BAP(J) (1:1 HENSM: RCL, 48 cells), and no RCL cells (2:3 HENSM: BAP(J), 120 cells). **b (from left to right),** representative immunofluorescence images of day 6 SEM aggregates showing OCT4 (cyan), SOX17 (yellow), and CK7 (magenta); nuclei (DAPI, white). Right, zooms into the SEMs from different aggregation conditions. Scale bars, 50 μm. **c**, quantification of the SEM efficiency from the control aggregation and the aggregation without RCL cells (Pre/ExEM-like cells), as judged by the formation of bilaminar Epi/Hb-like structure surrounded by the trophoblast-like layer. N = 3 (across 344, 157, and 1244 aggregates) and N = 3 (across 1063, 205, 929 aggregates) for control and no RCL conditions, respectively. Bars show mean values, whiskers mark s.d. p-value = 0.29; two-sided unpaired student t-test. **d**, representative immunofluorescence images of day 6 aggregates made solely from naïve ESC from HENSM media showing OCT4 (cyan), low SOX17 (yellow), and GATA3 (magenta); nuclei (DAPI, white). Scale bar, 200 μm. **e (left)**, scheme of the experiment testing the capacity of cells, differentiated from human WIBR3 primed ESCs, to form equivalent SEMs to those obtained from isogenic naïve ESCs expanded in HENSM conditions. **e (right)**, representative immunofluorescence images of day 6 SEMs showing OCT4

(cyan), SOX17 (yellow), CK7 (magenta), and nuclei (DAPI, white). When starting with isogenic WIBR3 primed ESCs, the resulting aggregates did not present organization and maturation of the key embryonic and extra-embryonic compartments as seen when starting with HENSM naïve ESCs (e.g., **Fig. 2**). Scale bar, 200 μm.

Extended Data Figure 6. Characterization of epiblast-, hypoblast-, and trophoblast-like lineages in SEMs. a (left to right), representative merged brightfield (BF) and immunofluorescence images of day 3 SEMs showing three lineages, epiblast-like (OCT4, cyan), hypoblast-like (SOX17, yellow), and trophoblast-like lineage (CK7, magenta) merged with a nuclei channel (DAPI, white). **b**, representative merged brightfield and immunofluorescence images of multiple day 6 SEMs, showing epiblast- (OCT4, cyan), hypoblast- or hypoblast/ExEM- (SOX17 or GATA6, respectively, in yellow), and trophoblast-like (SDC1 and GATA3, magenta) compartments merged with a nuclei channel (DAPI, white). **c (left)**, representative immunofluorescence images of day 6 SEMs showing epiblast- (OCT4, cyan), hypoblast- (SOX17, yellow), and trophoblast-like (CK7, magenta) compartments; nuclei (DAPI, white). Left panel represents a wide field image with adequate SEMs (outlined in green) and mis-developed structures (outlined in red); scale bar, 200 μm. Right panels zoom into the examples. **c (from left to right**), representative immunofluorescence image of the SEM developed according to the three success rate criteria (outlined in green, see **Methods**). Representative mis-developed day 6 SEMs (outlined in red) which do not fall within the success rate criteria: not surrounded by trophoblast (Tb)-like cells (I), without epiblast (Epi)- and hypoblast (Hb)- like cells (II), without the amniotic cavity (AC)-like structure and the yolk sac cavity (YS)-like structure and bilaminar Epi/Hb-like structure (III). Scale bars, 50 μm.

Extended Data Figure 7. Analyzing 3D structure of the human SEM at day 8. a, quantification of the SEM derivation efficiency at day 8 for WIBR3 line according to the criteria: I, surrounding by Tb-like layer; II, presence of Epi- and Hb-like compartments; III, presence of Bilaminar Epi/Hb disk-like structure, AC-like structure, and YS-like or Secondary(S) YS-like structures. N = 3 across 1251, 377, and 700 SEM aggregates. Bars show mean values, whiskers mark s.d. **b**, individual Z-planes of the 3D immunofluorescence image of a day 8 human SEM that meets criteria III specified in **a**. Epiblast- (OCT4, cyan), hypoblast- (SOX17, yellow), and trophoblast-like (CK7, magenta) compartments; nuclei (DAPI, white). AC, amniotic cavity-like; Am, amnion-like; ExEM, extraembryonic mesoderm-like; SYS, secondary yolk sac-like; ChC, chorionic cavity-like; STb, syncytiotrophoblast-like; Sk, stalk-like. Z-step, 20 μm; scale bar, 50 μm. See also **Supplementary Video 1**.

Extended Data Figure 8. Characterization of the epiblast-like structure in human SEM. a,

representative immunofluorescence image of day 6 SEM showing aPKC (green), OCT4 (cyan), F-ACTIN (red), nuclei (DAPI, white). Scale bar, 50 µm; 12.5 µm (zoom, right). **b (left)**, immunofluorescence image of day 6 SEM showing OCT4 (cyan), F-ACTIN (red), and phERM (green); alignment of epi-like cells in a single 2D plane is marked with dashed lines; asterisk, pro-amniotic-like cavity. Scale bar, 25 µm. **b (right),** quantification of the angle between the epiblast-like cell axis and the pro-amniotic-like cavity; the plot shows the radial histogram of the angle values and indicates predominant alignment of epi-like cells towards the center of the emerging cavity (n = 32). **c**, representative immunofluorescence image of day 6 SEM showing T/BRA expression (red); OCT4 (cyan), nuclei (DAPI, white). Right, zoom into the posterior epiblast-like structure, arrows mark individual T/BRA-positive cells. Scale bar, 50 µm, 12.5 µm (4x zoom in, right). **d**, representative immunofluorescence image of day 6 SEM showing CER1 (green) localization inside the intracellular vesicles; OCT4 (cyan), nuclei (DAPI, white). Right, zoom into hypoblast-like compartment, arrows mark apical side of the visceral endoderm-like cells with CER1 vesicles. Scale bar, 50 µm, 12.5 µm (4x zoom, right). **e (left)**, immunofluorescence image of day 6 aggregates showing T (red), OCT4 (cyan), CER1 (green), and nuclei (DAPI, white); scale bar, 500 µm. **e (right)**, zoom into the SEM with expression of T in epiblast-like compartment and CER1 in hypoblast-like compartment at the opposite sides (defined as AP-axis, see XZ cross section below) observed in 1.02% of total number of starting aggregates on day 0. N = 2 across 1251 and 700 aggregates. Scale bar, 50 µm. **f**, representative immunofluorescence image of day 6 SEM showing F-ACTIN (white) and nuclei (DAPI, blue). Right, zoom into epiblast-like structure with squamous cells (sEpi) in the top and cylindrical/columnar cells (cEpi) in the bottom parts of the epiblast-like compartment. Scale bar, 50 µm; zoom, 25 µm. **g**, representative immunofluorescence image of day 6 SEM showing F-ACTIN (white) and OCT4 (cyan). Right, zoom into epiblast-like compartment with squamous cells in the top and cylindrical cells in the bottom parts of the epiblast-like compartment. Scale bar, 50 µm; zoom, 25 µm. **h**, representative brightfield and immunofluorescence images of day 8 SEM showing NANOG (cyan), SOX17 (yellow), and nuclei (DAPI, white). Right, zoom into PGC-like cells in day 8 human SEMs, co-expressing NANOG and SOX17 (cells marked with arrows). Scale bars, 50 µm.

Extended Data Figure 9. Characterization of hypoblast-lake layer and extraembryonic mesoderm-like cells in human SEMs. a, representative immunofluorescence image of day 6 SEM showing apical polarity of the visceral and parietal hypoblast-like layers (SOX17, yellow); aPKC (heat gradient), F-ACTIN (white). Scale bar, 50 µm; zoom, 10 µm. See also **Supplementary Video 6**. **b**, representative immunofluorescence image of day 6 SEM showing OCT4 (cyan), GATA4 (red), SOX17 (yellow), and nuclei (DAPI). Right, zoom on ExEM-like cells expressing GATA4, but not SOX17. YS, yolk sac-like. Scale bar, 50 µm; zoom, 10 µm. **c (left),** zoom into the SEM example with adequate morphology and

structure (see Methods) aggregated from RUES2 reporter hESC line showing epiblast-like (SOX2-Citrine, yellow), hypoblast-like (SOX17-tdTomato, red), and trophoblast-like (CK7, magenta) compartments with 0.08% efficiency (N = 3 across 409, 295, and 528 aggregates); nuclei (DAPI, white); scale bar, 50 µm. **c (right)**, representative large field of view immunofluorescence image of day 6 SEM aggregates, the zoom on the left is outlined; scale bar, 500 µm. **d**, representative immunofluorescence image of day 6 SEM showing mesenchymal-like cells underneath the yolk sac-like structure; OCT4 (cyan), F-ACTIN (red), and nuclei (DAPI, white). Right, zoom into the region underneath the yolk sac-like structure, shown in different Z-planes (number 50, 60, and 70). Arrows point at the cells between yolk sac-like and the trophoblast-like compartments. Scale bar, 50 µm; zoom, 25 µm. See also **Supplementary Video 4. e**, representative immunofluorescence images of day 6 SEMs showing chorionic-like cavity surrounded by ExEM-like cells (outlined), negative for SOX17 (yellow), but expressing BST2 (top, red) and GATA6 (bottom, red); OCT4 (cyan), nuclei (DAPI, white). Scale bar, 50 µm. **f,** heatmap of GATA6 and SOX17 gene expression across extraembryonic tissues in marmoset, corresponding to the Carnegie stages (CS) 5-7), extracted from previously published gene expression dataset[8]. Am, amnion; Tb, trophoblast; VE, visceral endoderm; SYS, secondary yolk sac; ExEM, extraembryonic mesoderm. **g (top),** representative immunofluorescence image of Z slice from day 8 SEM and the zoom into the ExEM-like region showing VIM (red) expression and nuclei (DAPI, white). **g (bottom),** merged brightfield and maximum intensity projection showing VIM (red) expression. Scale bar, 50 µm; zoom, 10 µm. **h (top)**, representative immunofluorescence image of day 6 SEM from aggregation with trophoblast-like cells (control). The arrows in the zoom below point at the apical surface of the visceral hypoblast-like layer (SOX17, yellow); aPKC (green), F-ACTIN (red). Scale bar, 50 µm; zoom, 10 µm. **h (bottom)**, representative immunofluorescence image of day 6 aggregate from experiment where trophoblast-like cell fraction was omitted (no BAP(J)) condition (related to **Fig. 4k**). The arrows in the zoom below point at the apical surface of the PrE/hypoblast-like cells (or layer) surrounding the aggregate (SOX17, yellow); aPKC (green), F-ACTIN (red). Scale bar, 50 µm; zoom, 10 µm.

Extended Data Figure 10. Characterization of the trophoblast-like compartment in human SEM. a, representative immunofluorescence images of multiple day 6 SEMs showing epiblast-like structure (OCT4, cyan), hypoblast-like structure (SOX17 and GATA6, yellow), and trophoblast-like layer (SDC1 and GATA3, magenta) surrounding the SEMs; nuclei (DAPI, white), that were used for calculation presented in **(Fig. 5b)**. Scale bar, 200 µm. **b**, maximum intensity projection (Max. Proj.) of the immunofluorescence image of day 8 SEM showing trophoblast-like cells (GATA3, magenta) and nuclei, (DAPI, white). Scale bar, 50 µm. **c**, representative brightfield image overlayed with the immunofluorescence of day 8 SEM showing the outer trophoblast-like compartment (CK7, magenta) surrounding the entire SEM. Right, zoom

into the multinucleated trophoblast-like layer. Scale bars, 50 μm. **d**, quantification of the percentage of lacunae-like positive aggregates in all Tb-like positive aggregates (N = 3 independent experimental replicates, n = 241). Bars show mean values, whiskers mark s.d. **e**, representative immunofluorescence image of day 6 SEM showing HCGB (green) and CK7 (magenta). Right, zoom into HCGB+ syncytiotrophoblast-like cells; nuclei (DAPI, white), F-ACTIN (white, right zoom). HCGB expression was detected in 98.08% of all of aggregates with Tb-like compartment (n = 261). Arrows point at the outer syncytiotrophoblast-like cell surface. Scale bars, 50 μm. **f**, representative immunofluorescence image of a human SEM showing multinucleated HCGB-positive syncytiotrophoblast-like cells; HCGB (green), F-ACTIN (red), nuclei (DAPI, white). Arrows point at multiple nuclei inside the same single cell. Scale bars, 50 μm. **g**, commercial ELISA pregnancy test run on spent medium of the day 8 SEMs (Day 8, right) compared to unspent medium as a negative control (CTR, left) which detects the secretion of HCGB from the syncytiotrophoblast-like compartment of day 7-8 human. **h**, scheme of the 11-12 dpf human embryo with the zoom into the surface of the syncytiotrophoblast (magenta). **i**, representative immunofluorescence images of the syncytiotrophoblast-like cell surface in day 8 SEM showing SDC1 (magenta), HCGB (green), and the lipid membrane (WGA, white). Yellow and white arrows point at the HCGB+ vesicles and microvilli-like (Mv) plasma membrane protrusions, respectively. Mv-like protrusions were observed in 10 out of 10 aggregates surrounded by Tb-like compartment that expresses SDC1 (selected randomly for high-resolution imaging). Scale bar, 10 μm.

Extended Data Figure 11. Identification and validation of specific cell sub-types in human SEMs.
a, UMAP of the four annotated epiblast-like clusters (1, 4, 7 & 11) alongside normalized expression of key marker genes. From the four Epi-like clusters we subclassified two. The first we termed “Posterior epiblast-like cluster” (#4), which was characterized by upregulation of the EMT markers such as TBXT (T/Brachyury), MIXL1, EOMES, MESP1 and WNT8a. The second we termed a “committed epiblast-like” cluster (#7), which was marked by ZIC2, ZEB2, and VIM lineage commitment markers and the absence of NANOG, while maintaining OCT4 and SOX2. **b**, reclustering of the 4 epiblast-like clusters (1, 4, 7, 11) resulted in 5 finer clusters (A-E). **c,** pseudotime analysis over epiblast-like cells starting with epiblast-like (E+A) and progressing through two trajectories towards either committed epiblast-like cells (B) or posterior epiblast-like cells (C). **d,** gene expression profile of top 30 markers of committed epiblast-like sub compartment **(left)** or posterior epiblast-like sub compartment **(right)**. Cells are ordered by cell pseudotime score, showing the gradual increase in expression over pseudotime. **e**, normalized expression of key amnion marker genes projected on human SEM-related UMAP. Amnion-like cluster (#10) is highlighted in red. **f**, expression of BMP4 and FURIN in Amnion (Am)-like and STb-like clusters.

Extended Data Figure 12. Identification and validation of blood- and trophoblast-like cell sub-types in human SEMs.

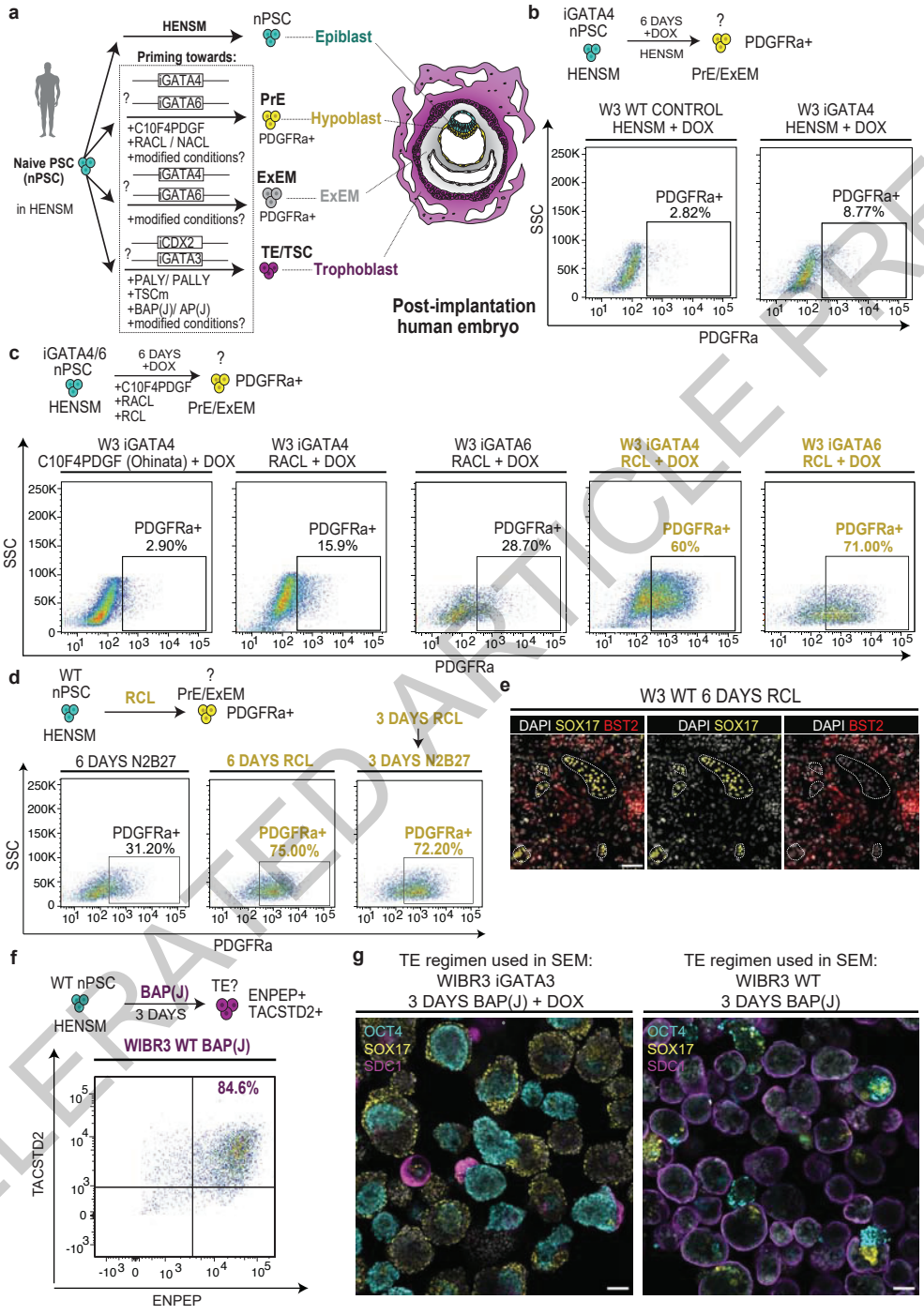
a, normalized expression of selected blood markers (TAL1, ERG, CD34), across cells with positive (>0) expression of CD34 or TAL1. In the dashed area 6 cells which are positive to all 3 markers can be observed. **b,** normalized expression of key blood marker genes (TAL1, ERG, CD34) projected on subset of SEM UMAP clusters. **c**, normalized expression of key trophoblast marker genes projected on SEM UMAP. Syncytiotrophoblast (STb)-like cluster (#12) is highlighted in red. **d**, normalized expression of key CTb marker genes projected on SEM UMAP clusters 10 (Amnion-like) and 12 (Syncytiotrophoblast-like). Cytotrophoblast (CTb)-like cells are highlighted in red. **e**, expression of top 50 differentially expressed genes (two-sided Wilcoxon test p-value < 0.05, logFC > 1.06 set as threshold), upregulated in Cytotrophoblast (CTb)-like cells compared to Syncytiotrophoblast (STb)-like cells. Annotation of CTb-like cells (n=31) and STb-like cells (n=100) was done based on the projection of SEM UMAP on the embryonic reference map (**Fig. 6e**). **f**, representative RT-qPCR gene expression (normalized by HPRT and RPL3) measured in SEM cells generated with or without BAP(J), and HENSM as control. Cell type markers were measured as following: general Tb markers (GATA2, TFAP2C), CTb-specific markers (TEAD3, TP63, OVOL1)[14], STb-specific markers (CGa, CGb, SDC1) and pluripotent marker OCT4 (used as a negative control). Bar plot based panel showing the average value of each sample (which represents average value of 3 technical replicates), error bars indicate s.d.. A single representative experiment out of N=3 biological replicates performed is shown.

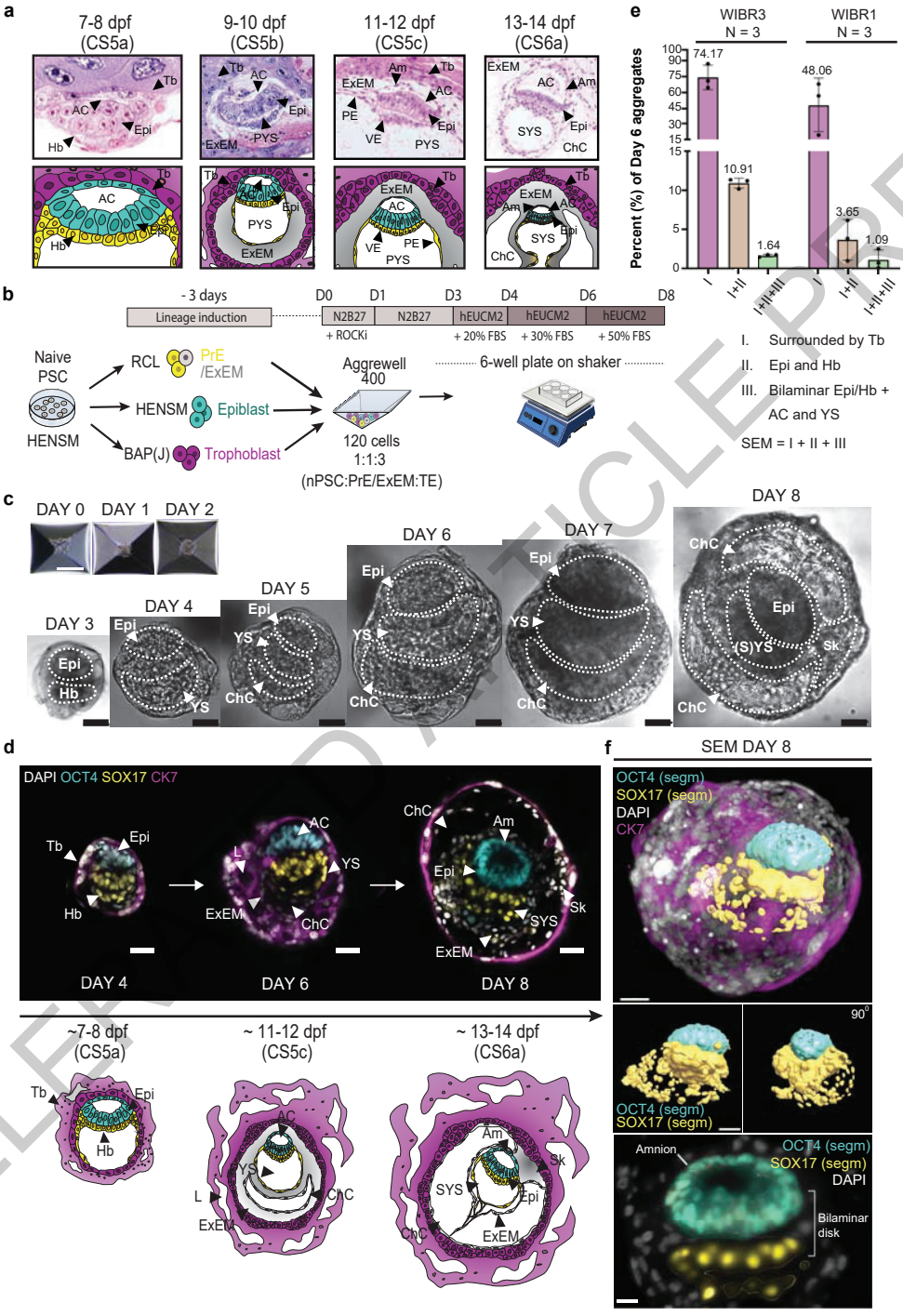
Extended Data Figure 13. Identification of Primary- and Secondary Yolk-sac-like cells, and Extra-Embryonic Mesoderm-like cells.

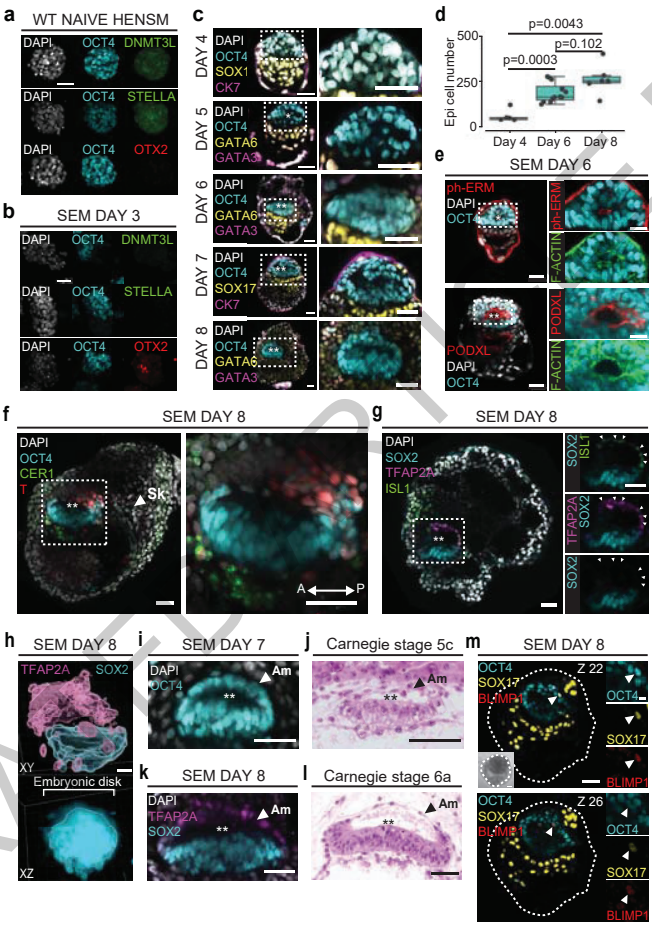
a, normalized expression of key yolk sac (YS) marker genes projected on the SEM UMAP. The co-expression of DKK1 and LHX1 alongside CER1 among SOX17+ YS-like cell population (arrows) in 9 marks AVE-like cells. **b**, reclustering of the two YS-like clusters (3 and 9) resulted in 3 finer clusters (A-C). **c**, pseudotime analysis over YS-like clusters showing progression of the transcriptional profile, reflected by pseudotime score, starting with YS-like structure and ending with SYS-like compartment. **d**, gene expression profile of the top 60 differentially expressed genes (between cluster #3 (YS-like) and cluster #9 (SYS-like), ordered by cell pseudotime score. **e (left)**, projection of day 3 RCL induced cell population (red triangle) on SEM UMAP, showing three cell populations in day 3 RCL fraction: ExEM-like, PrE/Hypoblast-like cells and residual primed ESCs. **e (right)**, the corresponding annotated SEM UMAP. **f**, integration of SEM ExEM-like cells with previously described *in vitro* ExEM[10] conversion time-course. UMAP of SEM ExEM-like populations (green hues) integrated with the previously published time course dataset of naive to PrE/ExEM conversion[9] (red hues). **f,** projection of early and late ExEM population

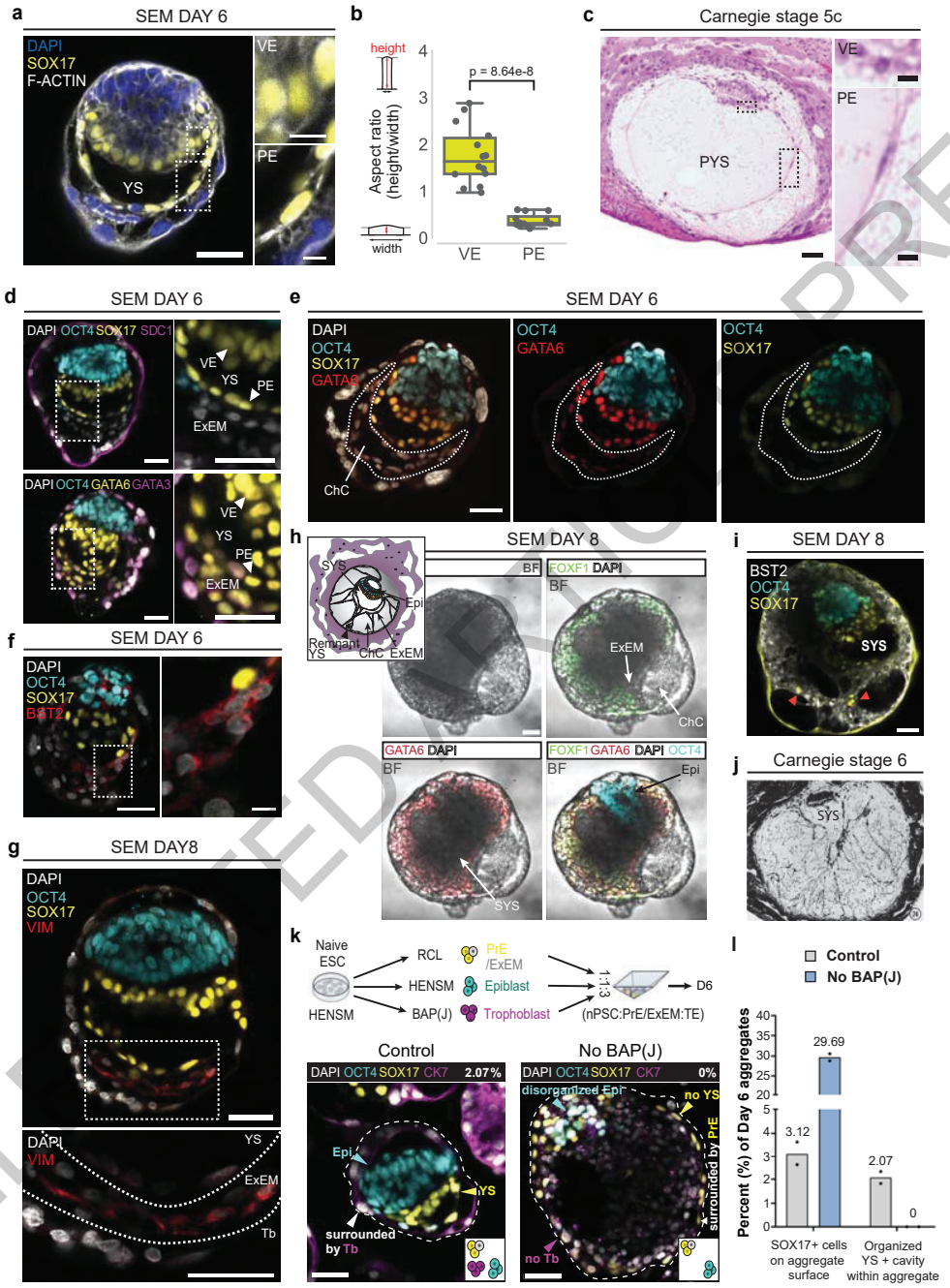
derived and sequenced in Pham et al. 2022[9] on the reference embryo meta-analysis. **f,** projection of the previously described ExEM populations[9] on the Reference human embryonic annotation map (described in **Fig 6e)**.

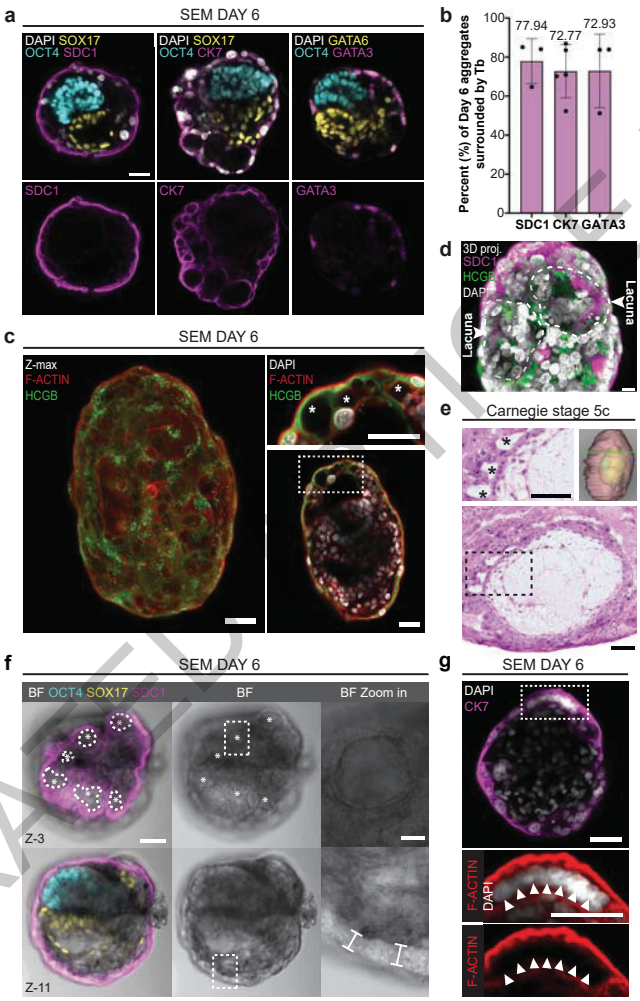
Extended Data Figure 14. Annotation and characterization of STb- and CTb-like cell populations in human SEMs. a, expression of significant differentially expressed genes between Amnion and Tb taken from Zhao et al.[26], in SEM amnion-like cells and Tb-like cells (performed using the FindMarkers function from R Seurat package and the 'roc' test). Seurat annotation and updated annotation from embryonic reference projection were indicated above. **b**, expression of top 50 significant differentially expressed genes between STb and CTb from[31] , in SEM CTb-like and STb-like cells (using the 'roc' test of FindMarkers function).

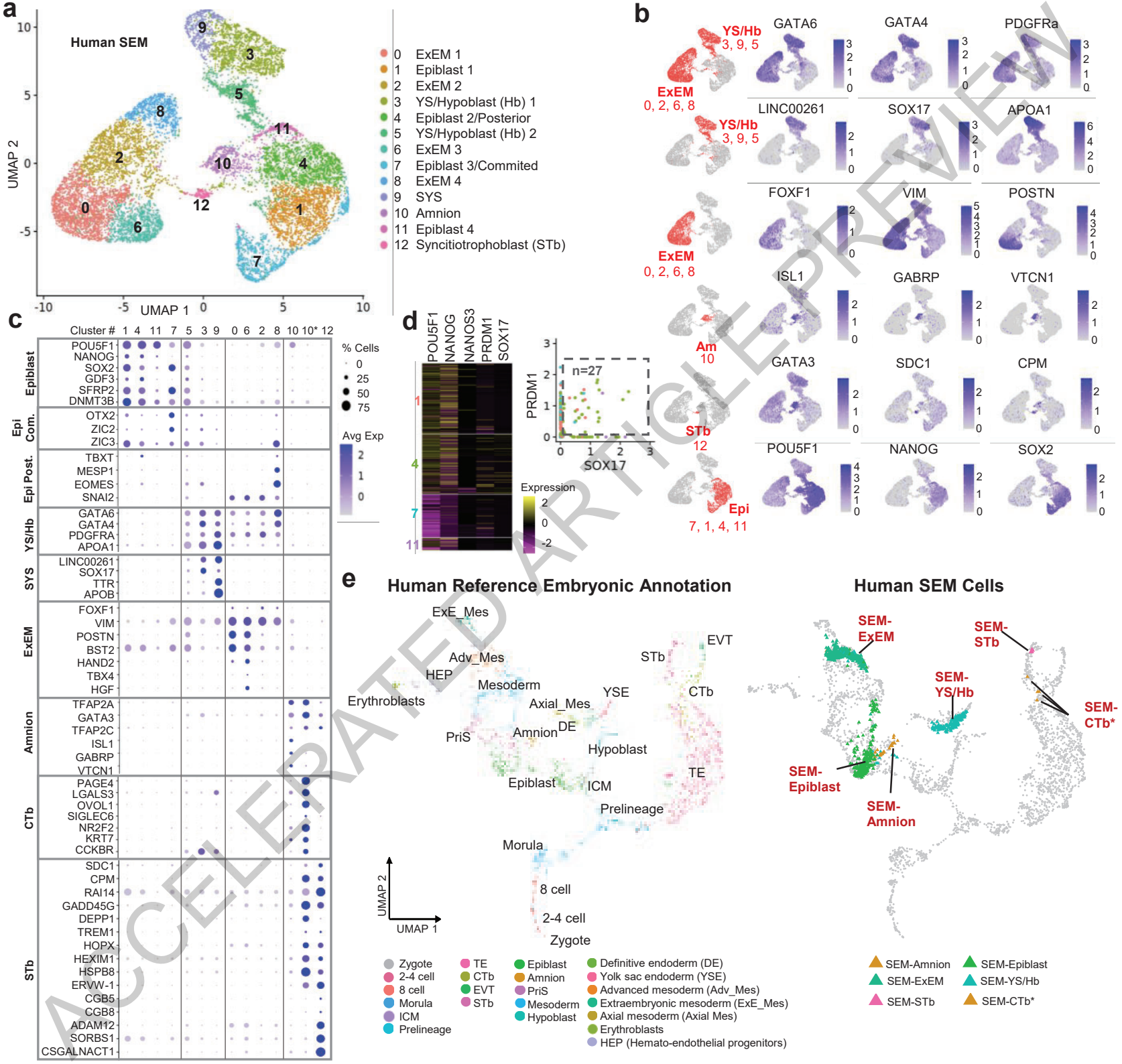


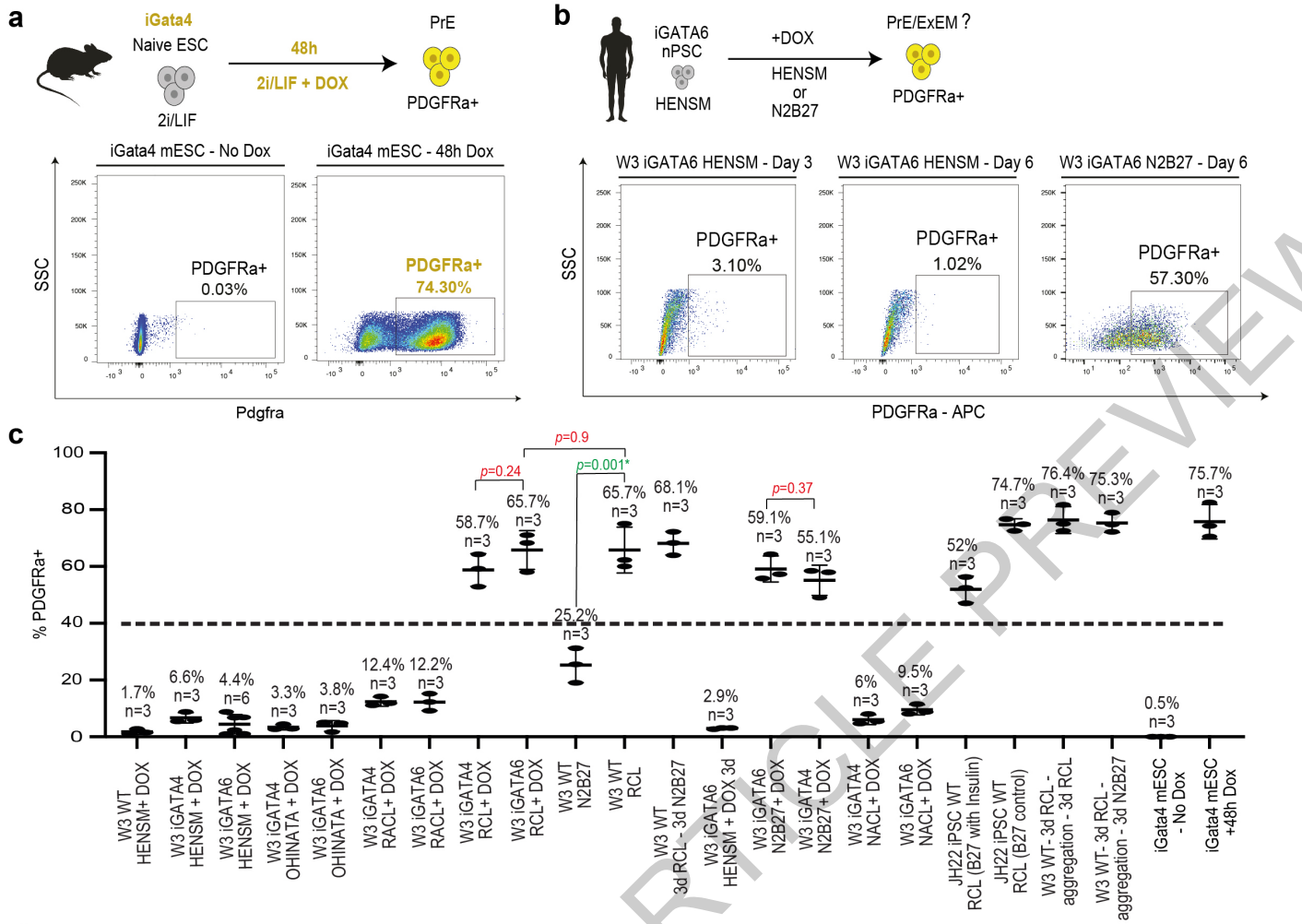


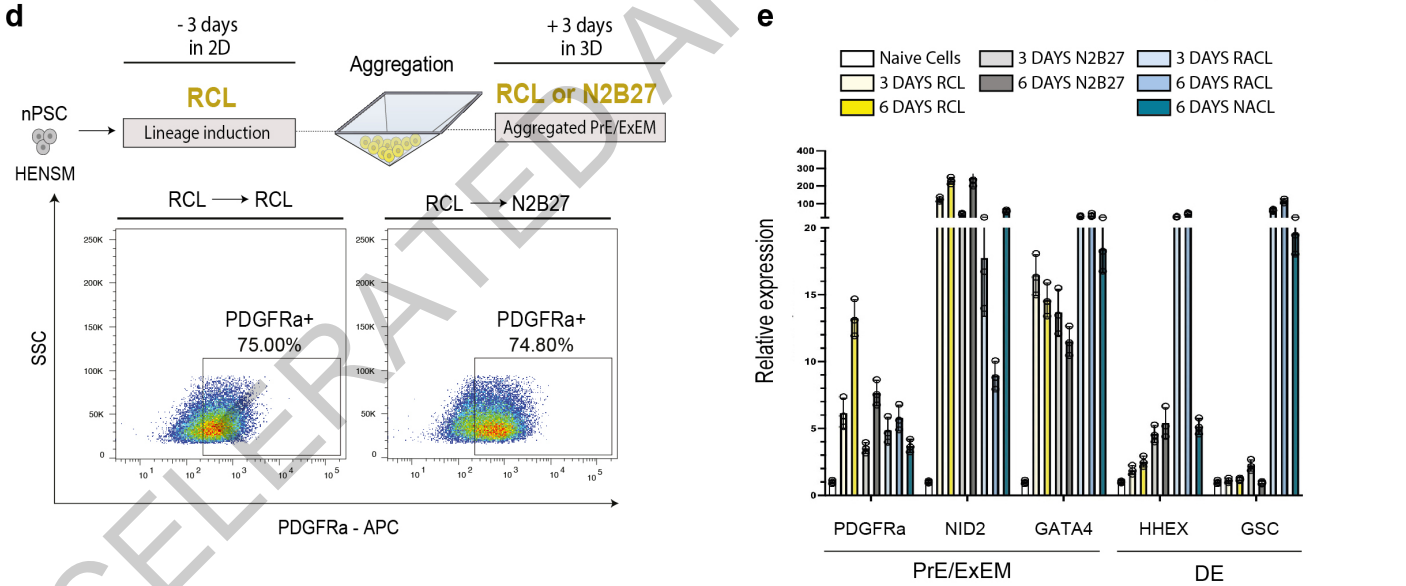


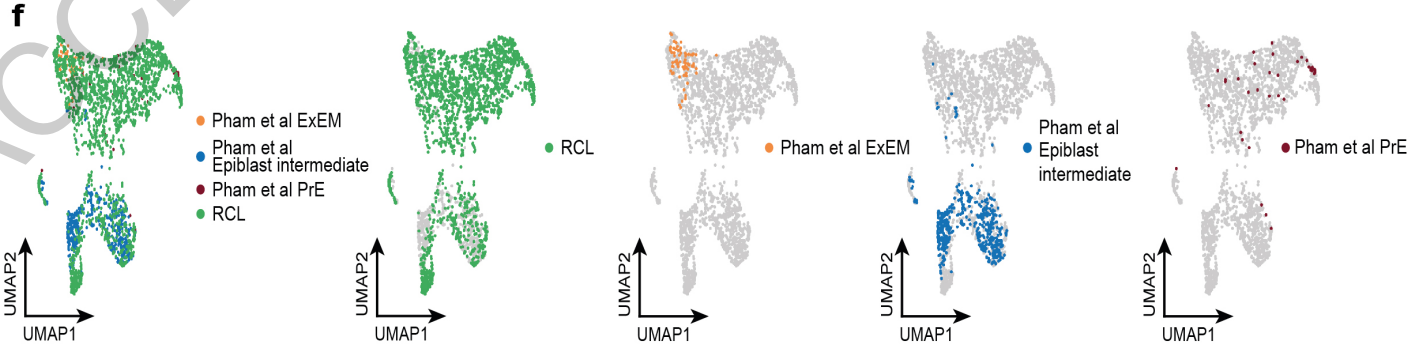




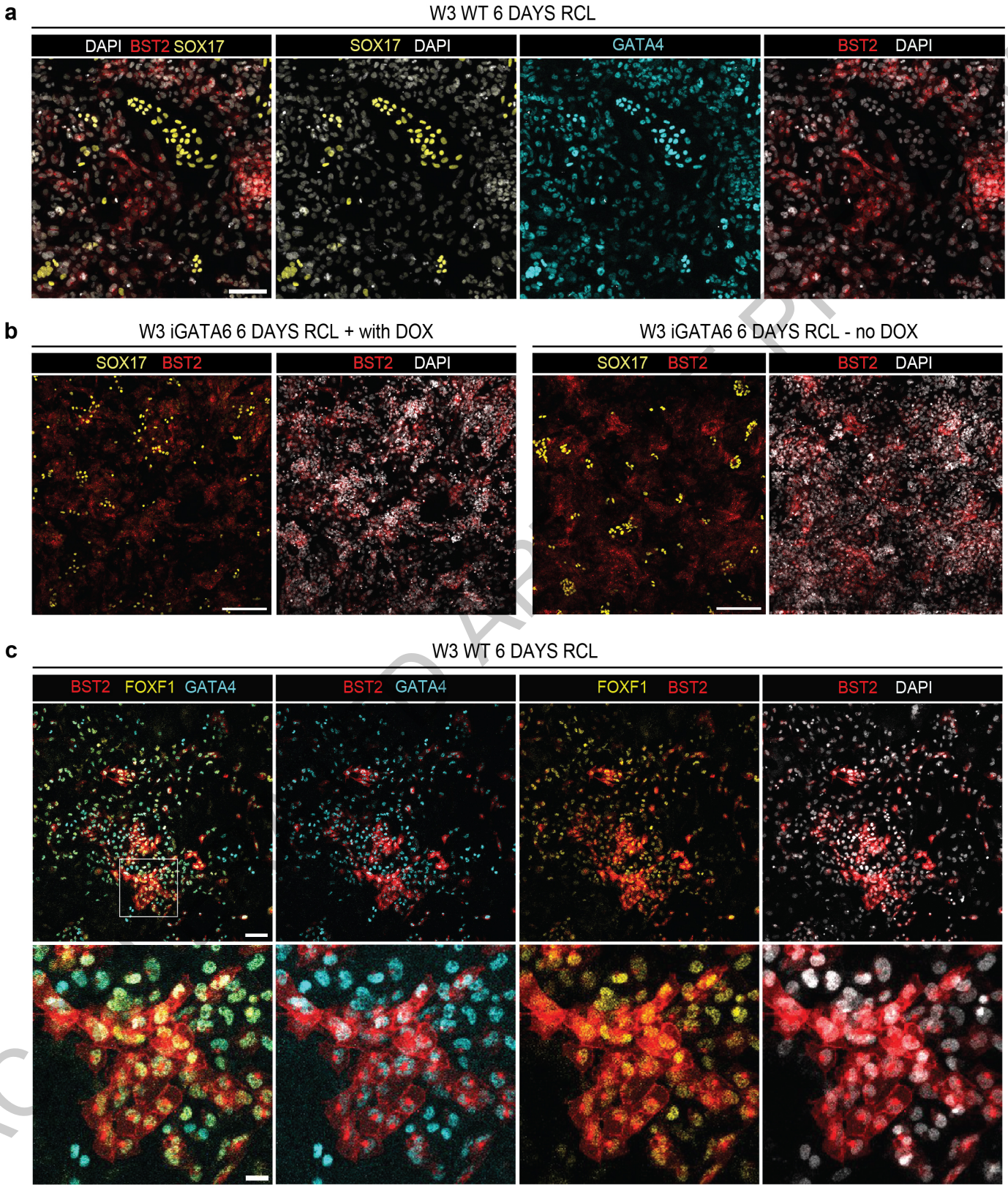




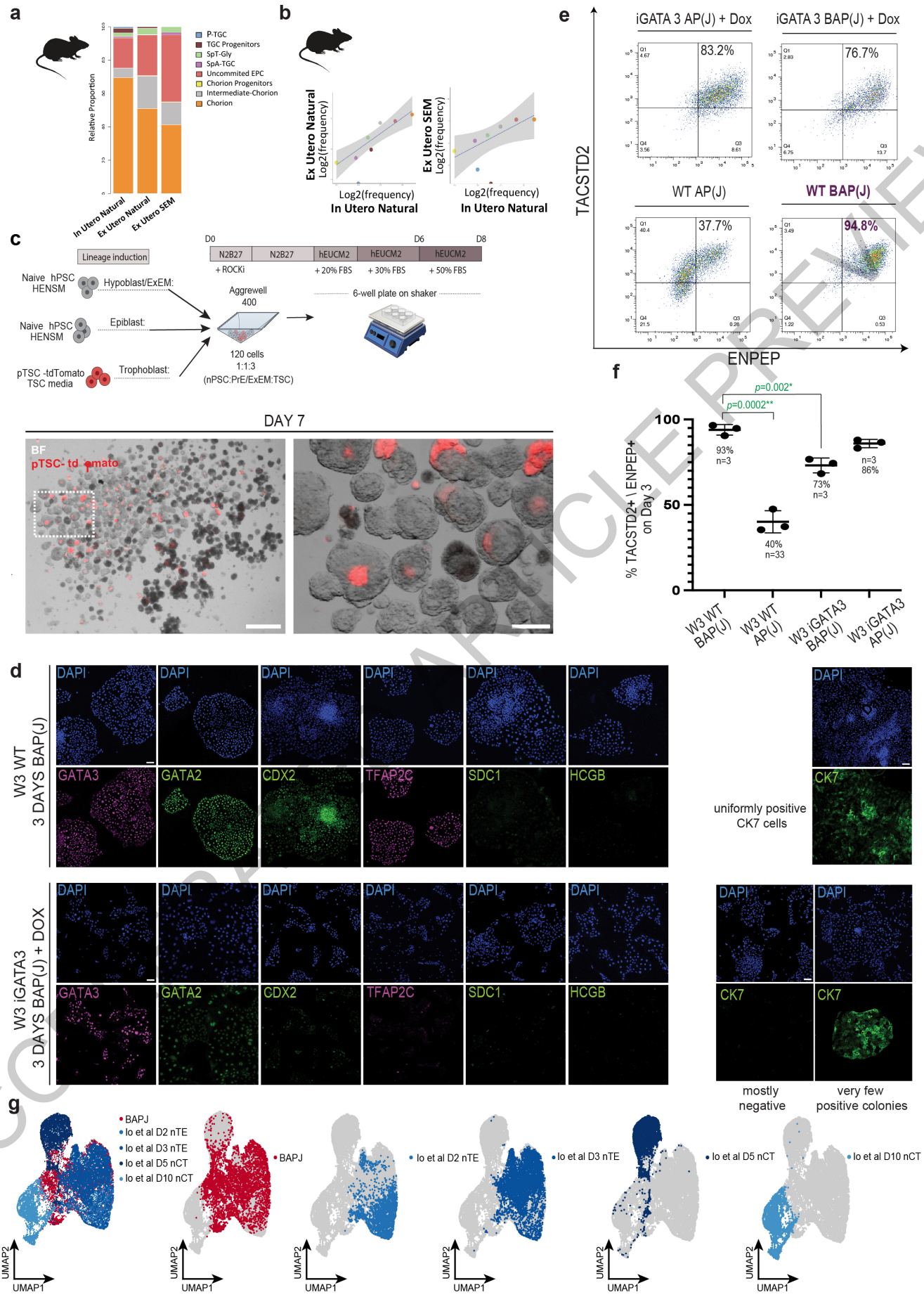




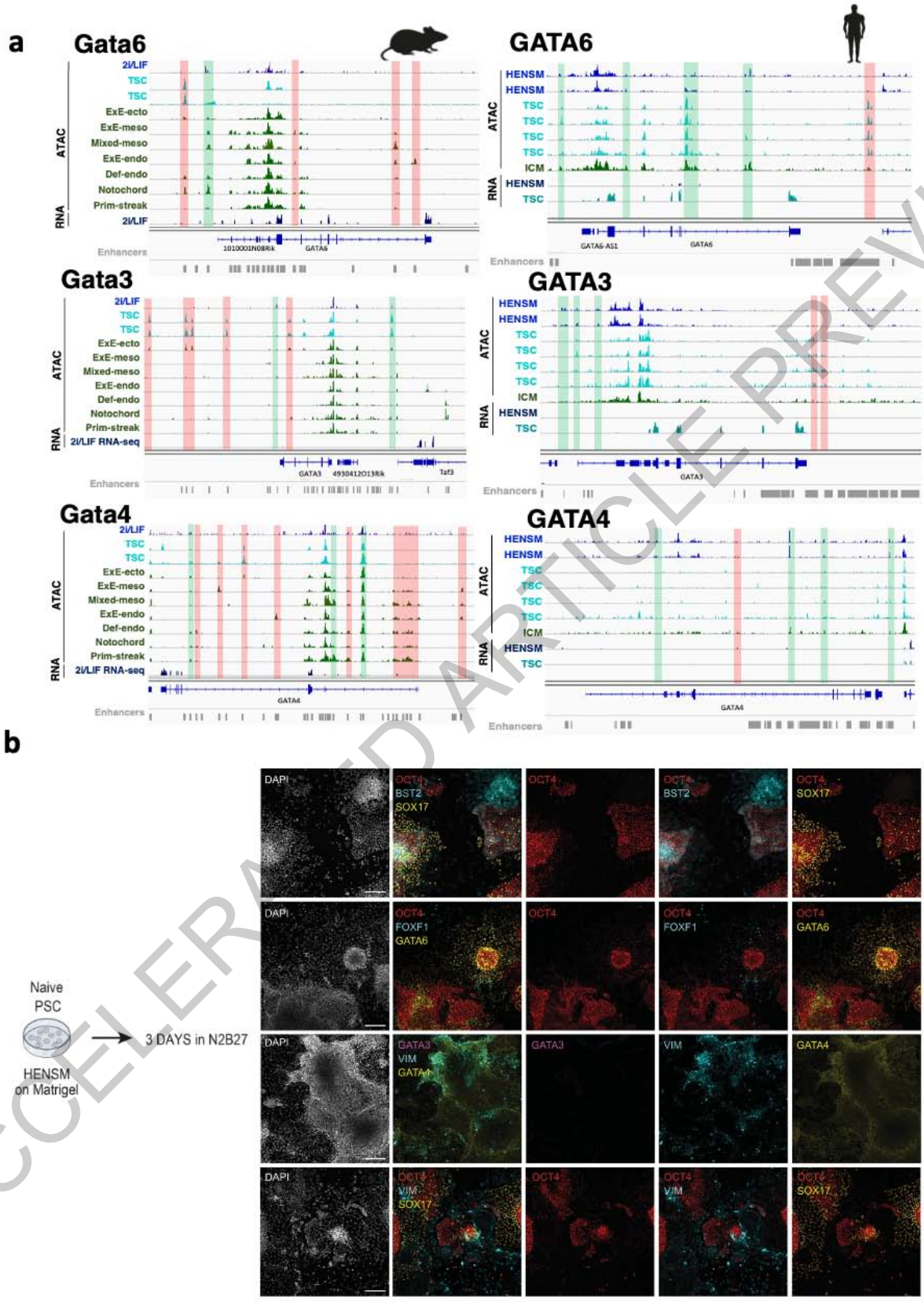
Extended Data Fig. 1



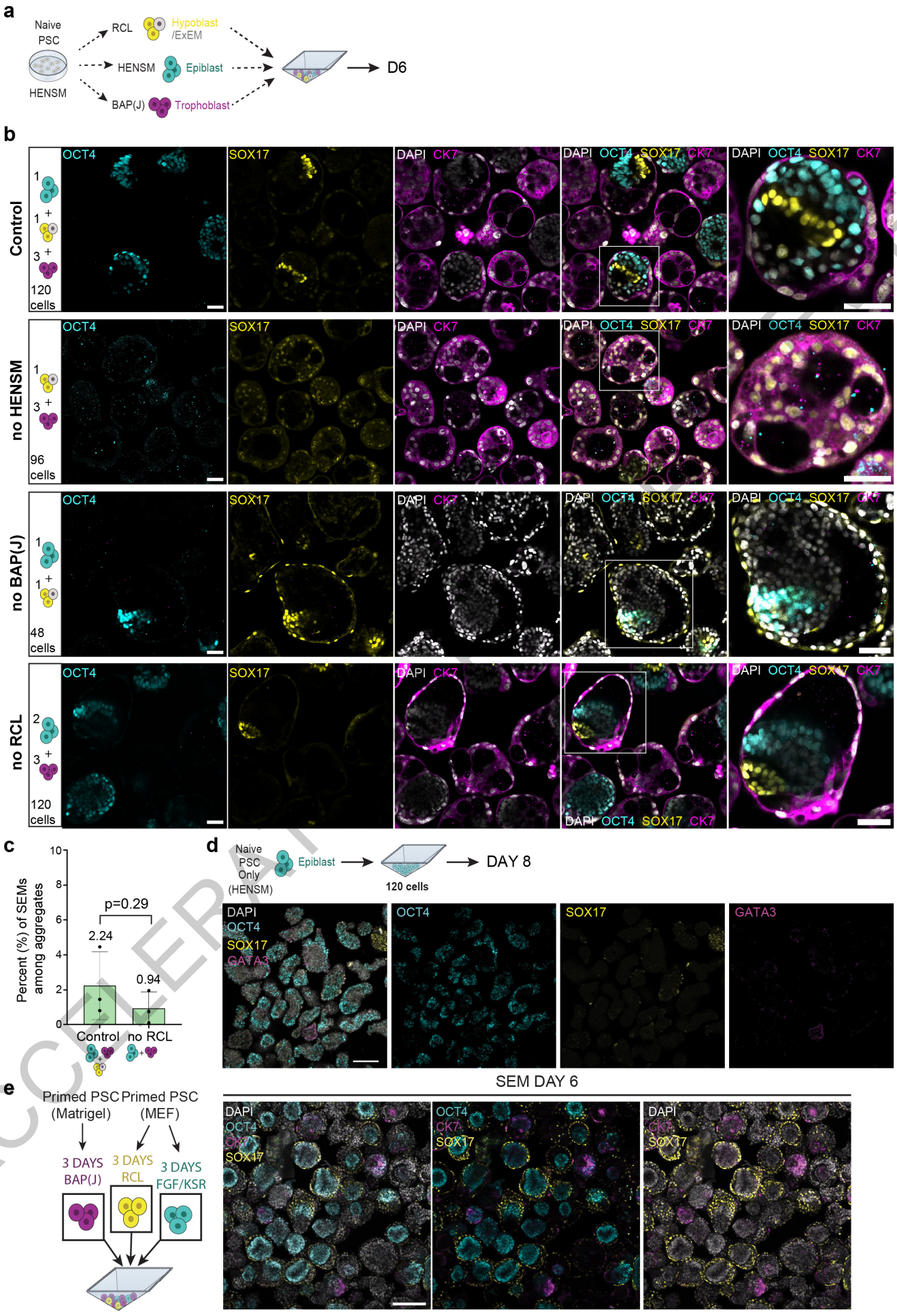
Extended Data Fig. 2



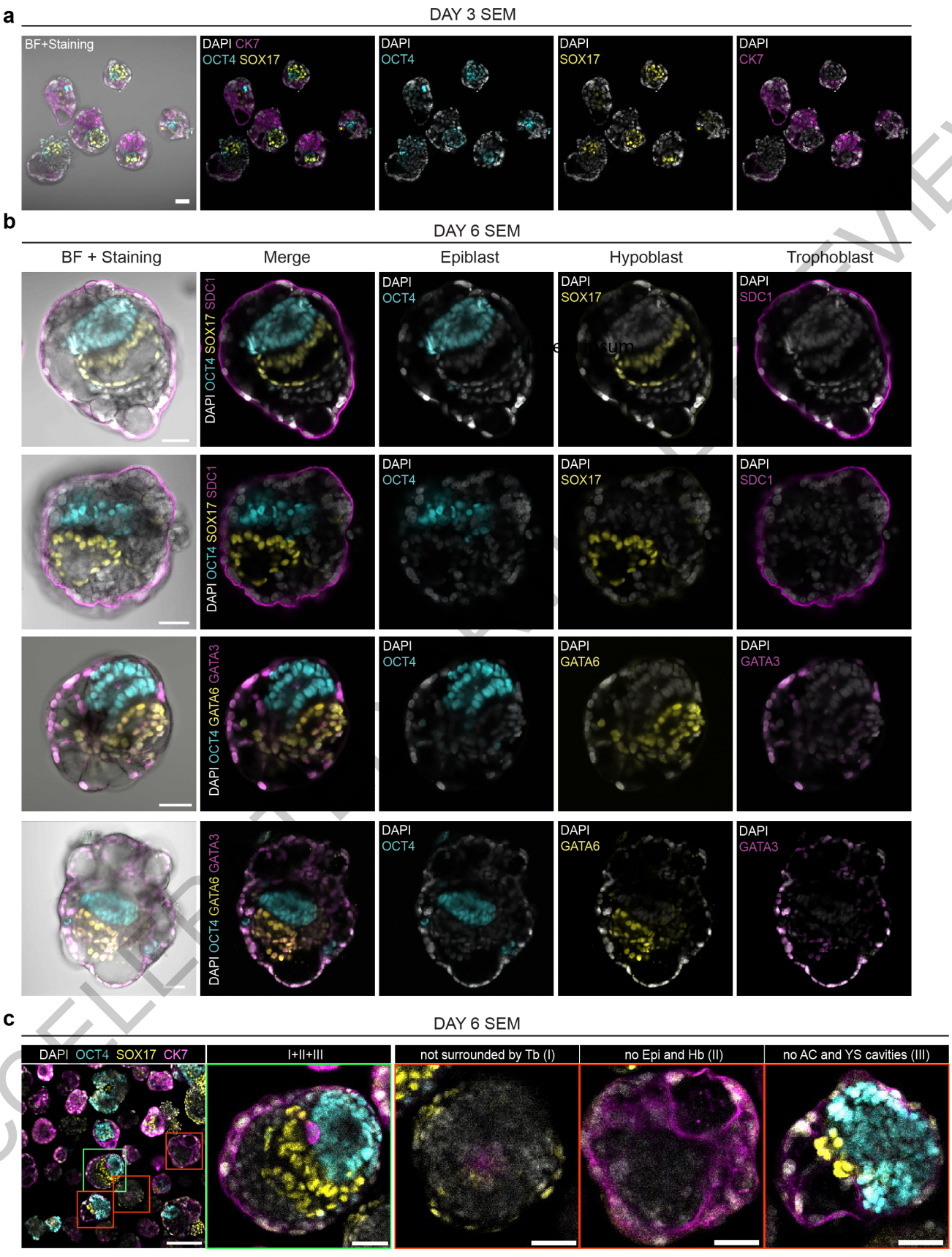
Extended Data Fig. 3



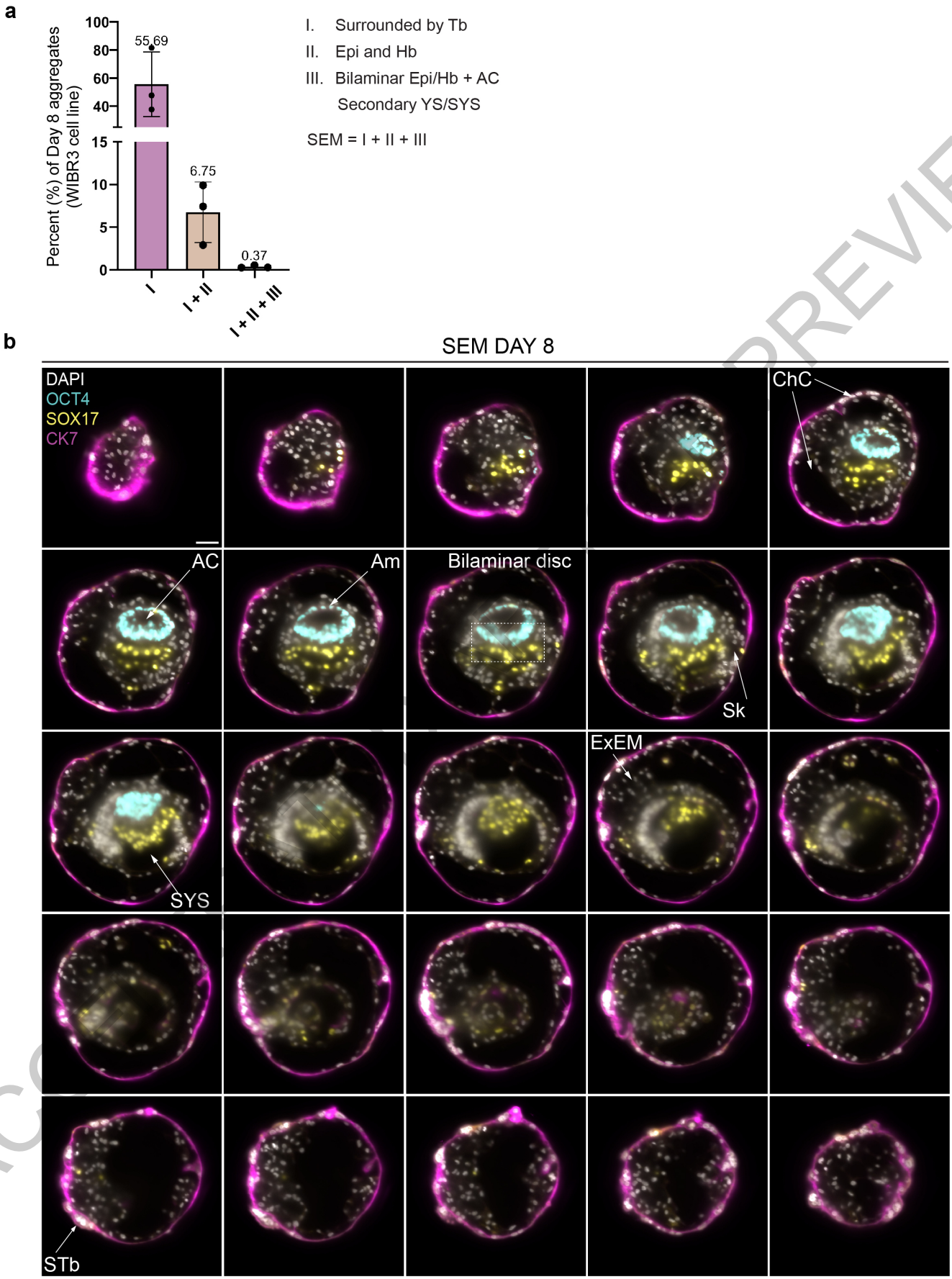
Extended Data Fig. 4



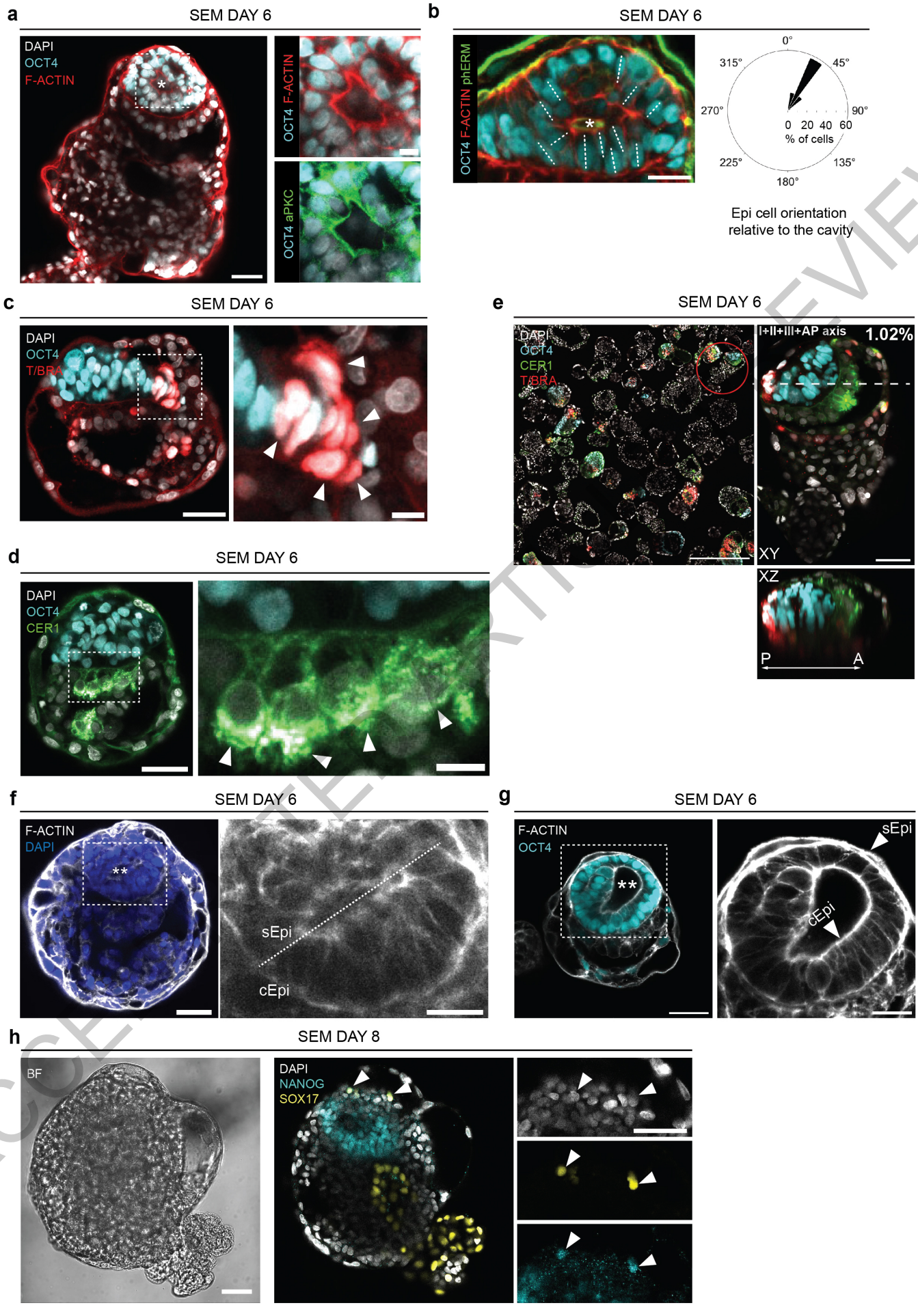
Extended Data Fig. 5



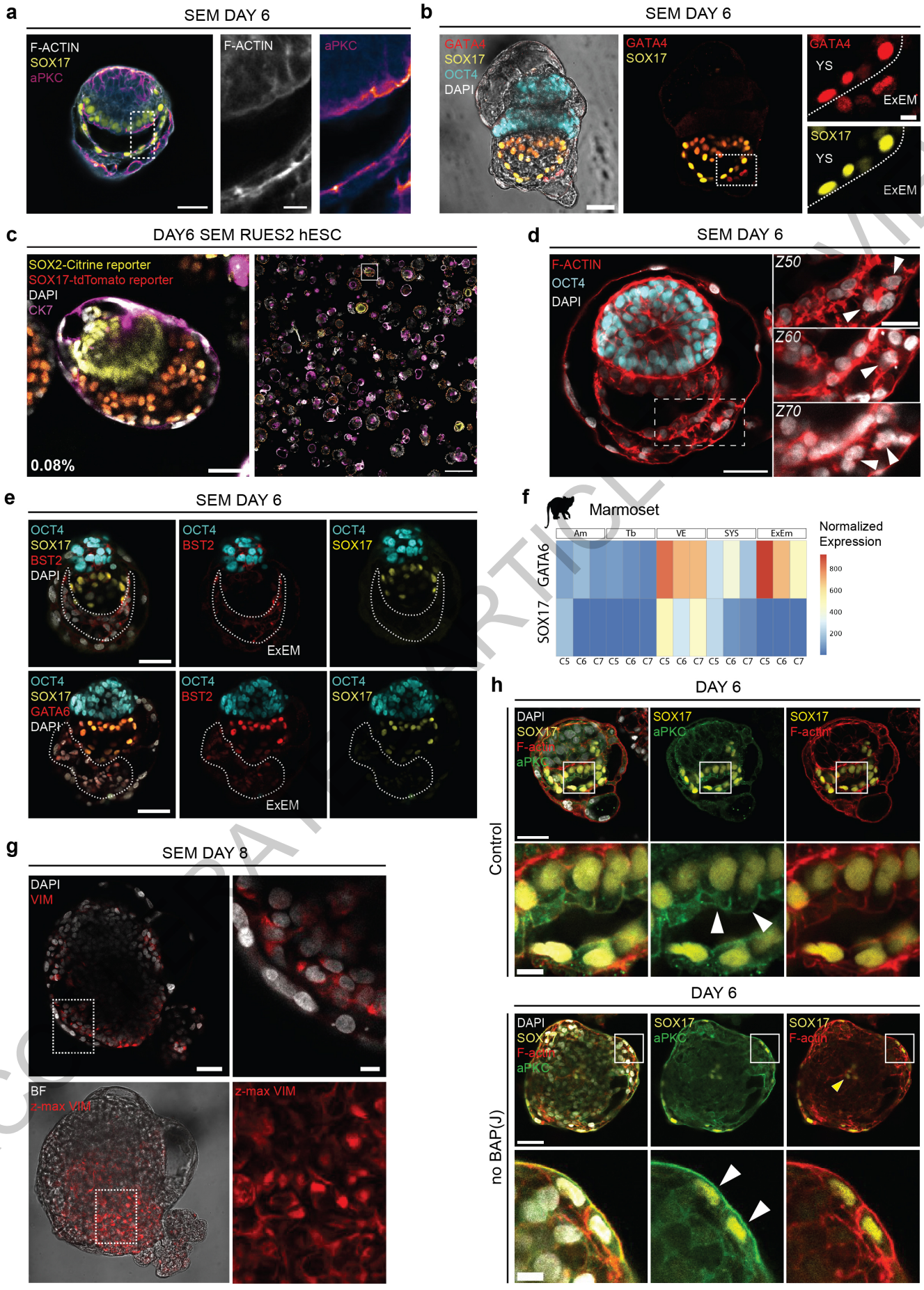
Extended Data Fig. 6



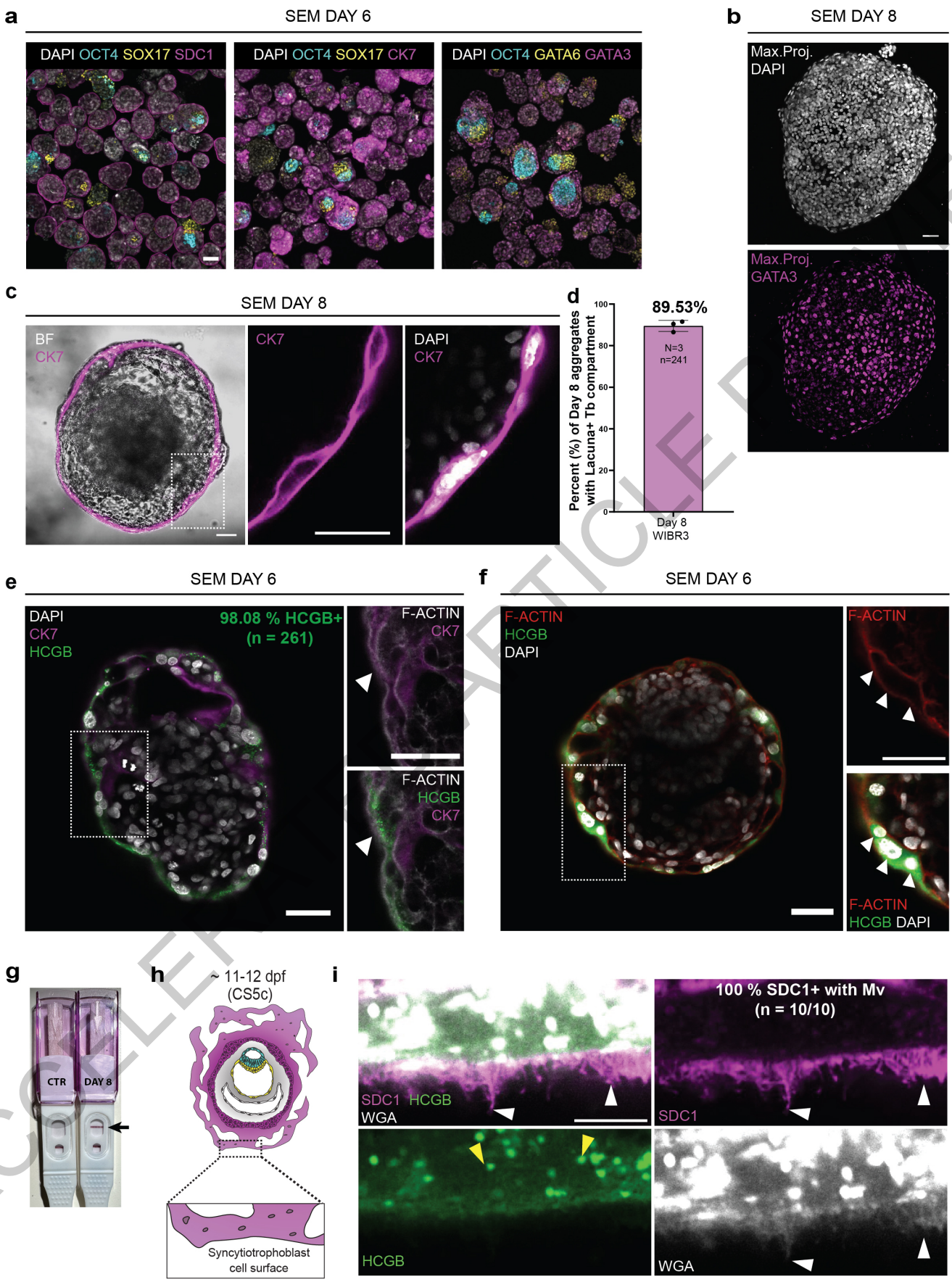
Extended Data Fig. 7



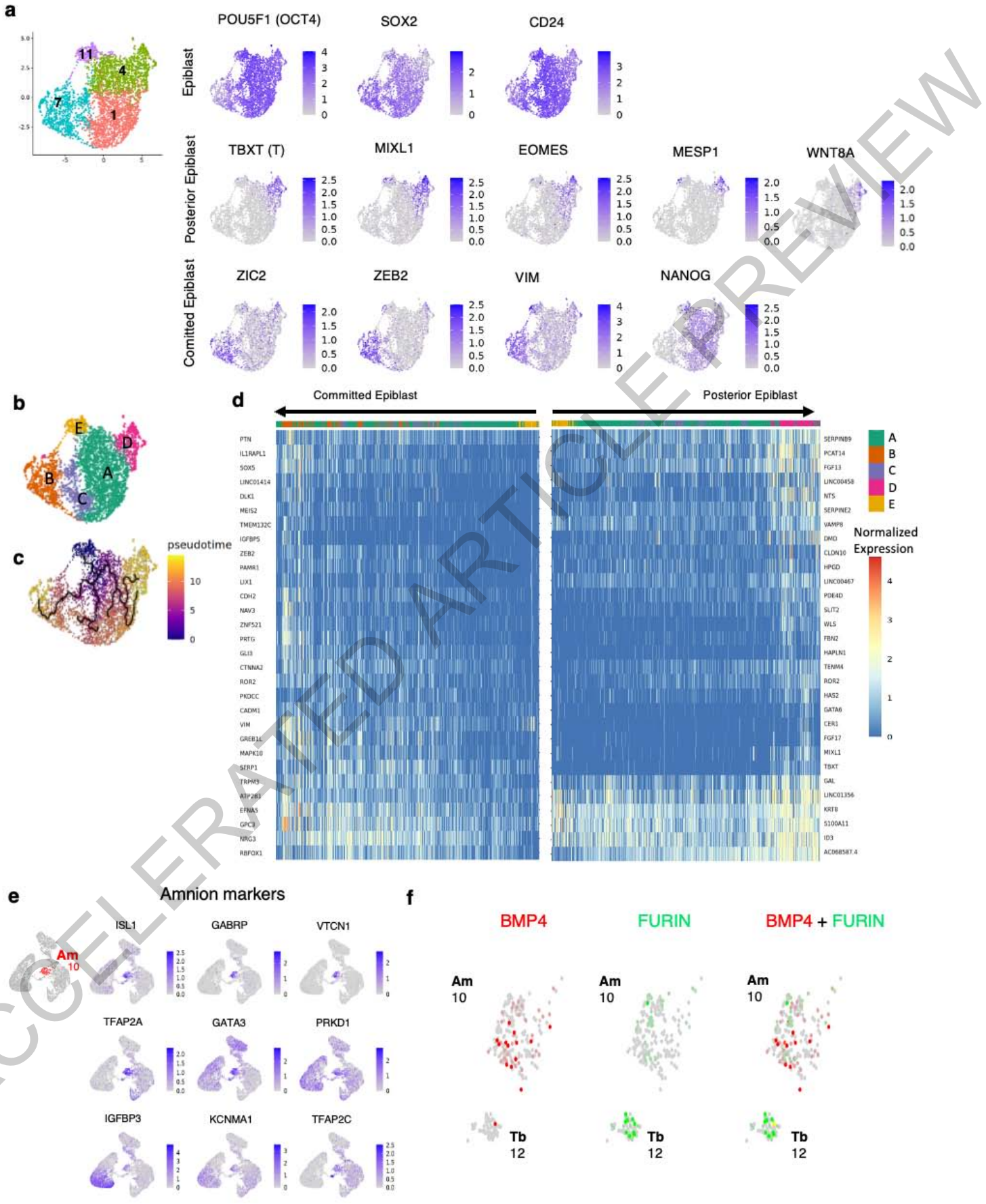
Extended Data Fig. 8



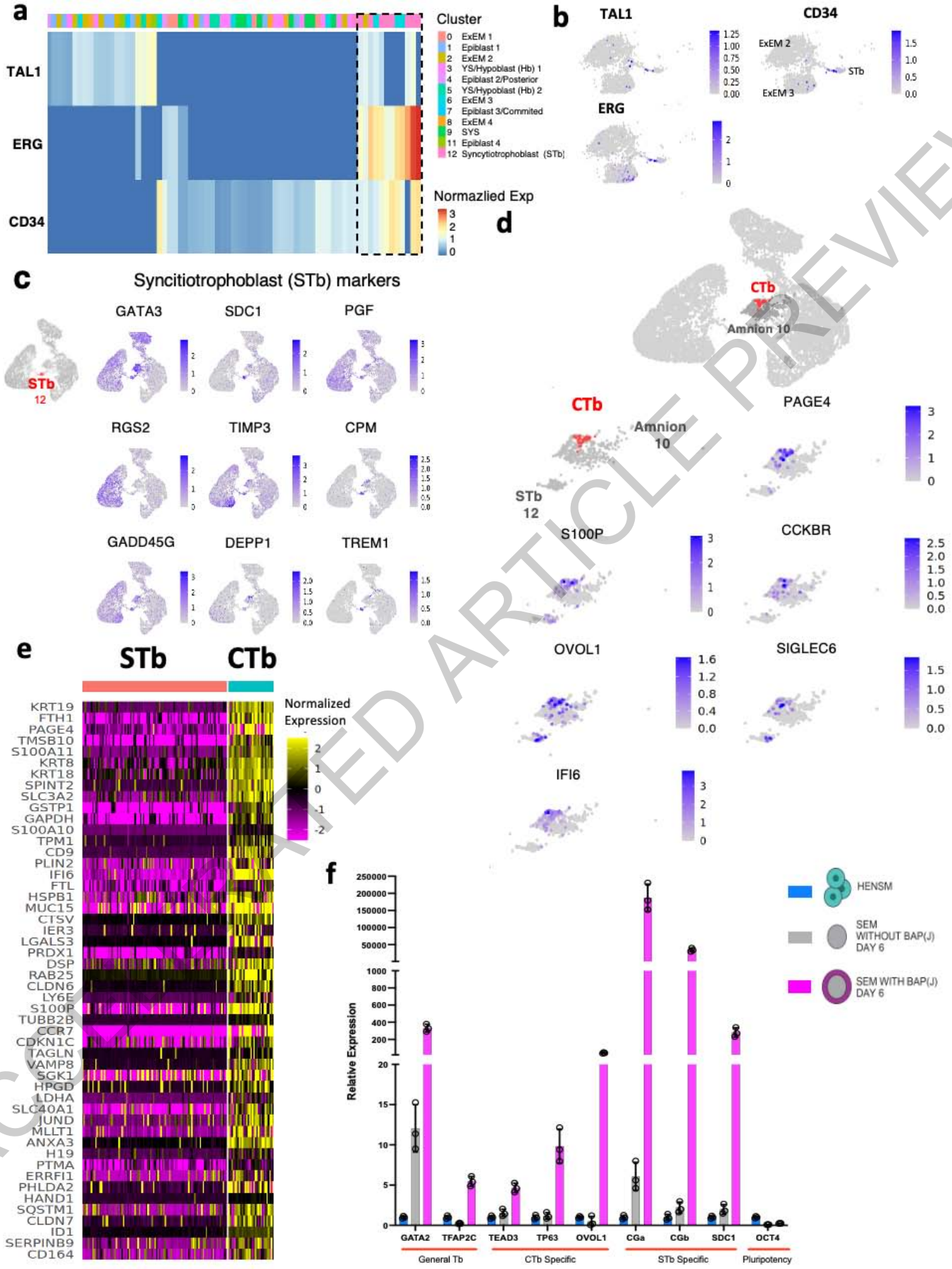
Extended Data Fig. 9



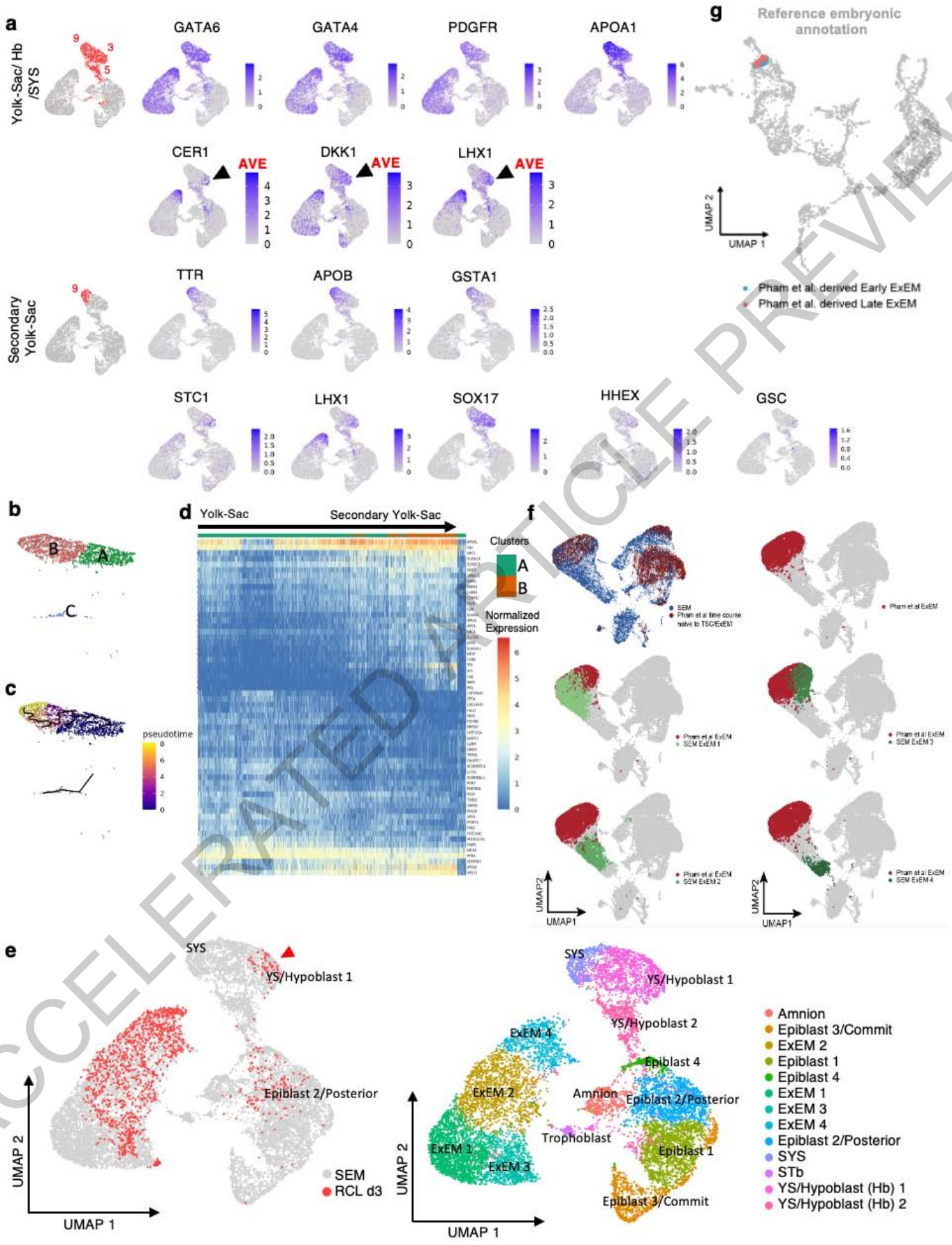
Extended Data Fig. 10



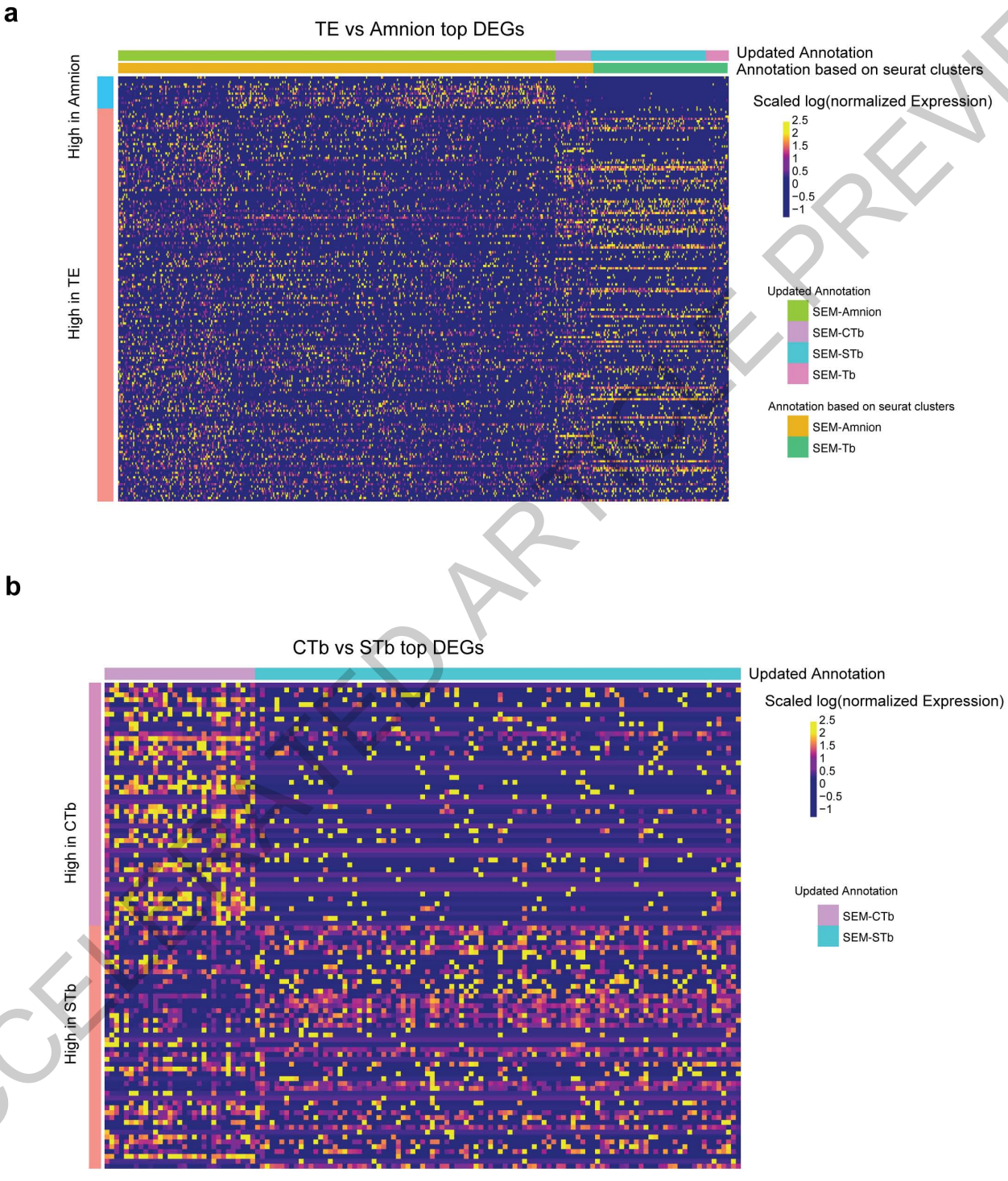
Extended Data Fig. 11



Extended Data Fig. 12



Extended Data Fig. 13



Extended Data Fig. 14

nature portfolio

Corresponding author(s): Jacob H. Hanna

Last updated by author(s): 8/14/2023

Reporting Summary

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

Statistics

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

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

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

Software and code

Policy information about availability of computer code

Data collection	Single cell sequencing data were collected using Novaseq platform, Illumina. qPCR data were obtained with the Viia7 platform (Applied Biosystems) via QuantStudio software (Version 7 Pro). Microscopy images were acquired with a Zeiss LSM 700, LSM800 inverted confocal microscopes (Carl Zeiss), and the light-sheet microscope (Z7, Carl Zeiss).
Data analysis	The statistical analysis besides single cell and bulk RNA-seq, was performed using the GraphPad Prism 8 software (La Joya, California) and Python v3.8.5, scipy v1.8.0 package; GraphPad Prism 8, Python's matplotlib v3.7.0 and seaborn v0.11.0 packages were used for plotting the data. Multiview fusion and deconvolution of the light-sheet microscopy data was performed in ZEN 3.5 software. Fiji/Image J (version 1.52p) was used for image analysis; manual cell counting and cell shape analysis were performed with Imaris v10.0.0 or v10.0.1 (Bitplane). 10X Genomics data analysis was performed with the Cell Ranger 7.1.0 software (10x Genomics) and Seurat 4.3.0, pheatmap 1.0.12, and Monocole3 R packages v1.3.1 Multiomics analysis was done also using Signac v1.6.0 and Harmony R V3 packages. Bulk ATAC-seq and RNA-seq were presented with Broad IGV software v2.16.2. Flow cytometry data was analyzed using FlowJo v10.7.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

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

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

Data availability
All newly generated scRNA-seq and 10x Chromium Single Cell Multiome ATAC + Gene Expression data are deposited under GEO: GSE239932. GSE number and reference are indicated for all other previously published and publicly available scRNA-seq and ATAC-seq data are indicated. Any other data is available upon request. All other information required to reanalyze the data reported in this work is available upon request from the corresponding author. Source data are provided with this paper.

Code availability
The custom code generated in this study is provided at GitHub: https://github.com/hannalab/Human_SEM_scAnalysis. The custom code was not essential to the main conclusions of this study.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. The number of SEMs used in each experiment was used with account of the data consistency/reproducibility and the available resources. Sample size for single cell RNA-Seq was determined when the main cell lineages at each developmental stages were captured.
Data exclusions	For scRNA-seq, to filter out low expressing single cells, possible doublets produced during the 10X sample processing or single cells with extensive mitochondrial expression, we filtered out cells with under 200 expressing genes, over 4000 expressing genes or over 10% mitochondrial gene expression.
Replication	The exact numbers of aggregates and biological replicates used for calculation of SEM protocol efficiency are indicated in the respective figure legends. All data refer to biological replicates and number of samples per biological replicate are indicated in figure legends and Methods section for all relevant panels and for all experiments.
Randomization	Human SEMs were chosen randomly when placed in different culture conditions. For efficiency calculations across conditions and developmental stages, multiple fields of view were imaged from randomly selected experiments and analyzed for an adequate contribution of each lineage with the relevant immunostaining. Number of biological samples/replicates and number of samples per biological replicate are indicated in figure legends and Methods section for all relevant panels. Other experiments were not randomized.
Blinding	The investigators were not blinded to allocation during experiments and outcome assessment. We had no relevant scientific reasons to conduct blinding.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Mouse monoclonal anti-Oct3/4 (clone C-10) (Santa Cruz Cat# SC-5279), 1:100;
Rabbit polyclonal anti-Oct3/4 (clone H-134) (Santa Cruz Cat# SC-9081), 1:100;
Goat polyclonal anti-Sox17 (R&D Cat# AF1924),1:100;
Rabbit monoclonal anti-Cytokeratin 7 (Abcam Cat# ab181598), 1:200;
Rabbit monoclonal anti-Cytokeratin 7 (Abcam Cat# ab68459), 1:200;
Goat polyclonal anti-Gata3 (R&D Cat# AF2605), 1:100;
Rabbit monoclonal anti-Syndecan1 (Abcam Cat# ab128936), 1:400;
Mouse monoclonal anti-Cdx2 (Biogenex Cat# MU392A-UC), 1:200;
Rabbit monoclonal anti-Phospho-Ezrin (Cell Signaling Cat# 3726), 1:400;
Rabbit monoclonal anti-Brachyury(D2Z3J) (Cell Signaling Cat# 81694), 1:100;
Goat polyclonal anti-Cer1 (R&D Cat# AF1075), 1:100;
Rabbit monoclonal Nanog (Abcam Cat# ab109250), 1:100;
Mouse monoclonal anti-PKC zeta Antibody (H-1) (Santa Cruz Cat# SC-17781), 1:200;
Mouse monoclonal anti-Podocalaxyn [clone 222328] (R&D Cat# MAB1658), 1:200;
Rabbit polyclonal anti-Gata4 (Abcam Cat# ab84593), 1:100;
Mouse monoclonal anti-Vimentin (Abcam Cat# ab8978), 1:100;
Rabbit monoclonal anti-BST2/Tetherin antibody [EPR20202-150] (Abcam Cat# ab243230), 1:100;
Rabbit monoclonal anti-hCG beta [5H4-E2] (Abcam Cat# ab9582), 1:200;
Rabbit monoclonal anti-Gata6 (clone D61E4) (Cell Signaling Cat# 5951), 1:100;
Rabbit monoclonal anti-Islet1 [EP4182] (Abcam Cat# ab109517), 1:100;
Mouse monoclonal anti- Anti-TFAP2a (AP-2α) (3B5) (Santa Cruz Cat# SC-12726), 1:100;
Goat polyclonal anti-Sox2 (R&D Cat# AF2018), 1:200;
Rabbit polyclonal anti-Dnmt3l (Imgenex/Novus Biologicals, Cat# IMG-6804A), 1:100;
Goat polyclonal anti-Otx2 (R&D Cat# AF1979), 1:200;
Mouse molonclonal anti-Stella (D-5 clone) (Santa Cruz Cat# SC-376862), 1:100;
Rabbit monoclonal anti- Blimp1/PDRI-BF1 [Clone C14A4] (Cell Signaling Cat# 9115), 1:100;
Goat polyclonal anti-FoxF1 (R&D Cat# AF4798), 1:100;
Goat polyclonal Nidogen2 (R&D Cat# AF3385), 1:100;
Rabbit monoclonal anti-Gata2 [EPR2822] (Abcam Cat# ab109241), 1:200.

FACS analysis:
Mouse monoclonal anti human TROP2-488 labeled (R&D Cat# FAB650G), 1:20;
Mouse monoclonal anti human CD249 (ENPEP)-BV421 labeled (BD Cat# 744872), 1:20;
Rat monoclonal anti mouse CD140a (PDFGR-a)-PE/Cy7 labeled (BioLegend Cat# 135912), 1:20;
Mouse monoclonal anti human CD140a (PDFGR-a)-PE/Cy7 labeled (BioLegend Cat# 323508), 1:20.
Mouse monoclonal anti human CD140a (PDFGR-a)-APC labeled (BioLegend Cat# 323512), 1:20.

Validation

All the antibodies have been validated by the companies from which they were obtained. Details of the validation statements, antibody profiles and relevant citations can be found on the manufacturer's website provided here.

Mouse monoclonal anti-Oct3/4 (clone C-10) (Santa Cruz Cat# SC-5279) has been referenced in 2450 publications: https://www.scbt.com/p/oct-3-4-antibody-c-10.

Rabbit polyclonal anti-Oct3/4 (clone H-134) (Santa Cruz Cat# SC-9081); has been referenced in 139 publications: https://www.scbt.com/p/oct-3-4-antibody-h-134.

Goat polyclonal anti-Sox17 (R&D Cat# AF1924); has been referenced in 288 publications: https://www.rndsystems.com/products/human-sox17-antibody_af1924?gclid=Cj0KCQjwoeemBhCfARIsADR2QCuGI49R7nTqXVxTxMmT2oKBdmlAHP7HGcpMELWWy1fve2cejy1VYMcaAk-YEALw_wcB&gclsrc=aw.ds

Rabbit monoclonal anti-Cytokeratin 7 (Abcam Cat# ab181598); has been referenced in 91 publications: https://www.abcam.com/products/primary-antibodies/cytokeratin-7-antibody-epr17078-cytoskeleton-marker-ab181598.html

Rabbit monoclonal anti-Cytokeratin 7 (Abcam Cat# ab68459); has been referenced in 31 publications: https://www.abcam.com/products/primary-antibodies/cytokeratin-7-antibody-epr1619y-cytoskeleton-marker-ab68459.html

Goat polyclonal anti-Gata3 (R&D Cat# AF2605); has been referenced in 19 publications: https://www.rndsystems.com/products/human-gata-3-antibody_af2605?gclid=Cj0KCQjwoeemBhCfARIsADR2QCuyed6_9X10TlOJlqu7BWtjQsRWKClUtoMxUbi_qSHy8gP4zR_PnBUaAhFaEALw_wcB&gclsrc=aw.ds

Rabbit monoclonal anti-Syndecan1 (Abcam Cat# ab128936); has been referenced in 39 publications: https://www.abcam.com/products/primary-antibodies/syndecan-1-antibody-epr6454-ab128936.html

Mouse monoclonal anti-Cdx2 (Biogenex Cat# MU392A-UC); has been referenced in 49 publications: https://www.labome.com/product/Biogenex/MU392A-UC.html

Rabbit monoclonal anti-Phospho-Ezrin (Cell Signaling Cat# 3726); has been referenced in 86 publications: https://www.cellsignal.com/products/primary-antibodies/phospho-ezrin-thr567-radixin-thr564-moesin-thr558-48g2-rabbit-mab/3726?_requestid=664905

Rabbit monoclonal anti-Brachyury(D2Z3J) (Cell Signaling Cat# 81694); has been referenced in 27 publications: https://www.cellsignal.com/products/primary-antibodies/brachyury-d2z3j-rabbit-mab/81694

Goat polyclonal anti-Cer1 (R&D Cat# AF1075); has been referenced in 3 publications: https://www.rndsystems.com/products/human-cerberus-1-antibody_af1075

Rabbit monoclonal Nanog (Abcam Cat# ab109250); has been referenced in 154 publications: https://www.abcam.com/products/primary-antibodies/nanog-antibody-epr20272-ab109250.html

Mouse monoclonal anti-PKC zeta Antibody (H-1) (Santa Cruz Cat# SC-17781); has been referenced in 134 publications: https://www.scbt.com/p/pkc-zeta-antibody-h-1

Mouse monoclonal anti-Podocalaxyn [clone 222328] (R&D Cat# MAB1658); has been referenced in 13 publications: https://www.rndsystems.com/products/human-podocalyxin-antibody-222328_mab1658

Rabbit polyclonal anti-Gata4 (Abcam Cat# ab84593); has been referenced in 63 publications: https://www.abcam.com/products/primary-antibodies/gata4-antibody-ab84593.html

Mouse monoclonal anti-Vimentin (Abcam Cat# ab8978); has been referenced in 507 publications: https://www.abcam.com/products/primary-antibodies/vimentin-antibody-rv202-cytoskeleton-marker-ab8978.html

Rabbit monoclonal anti-BST2/Tetherin antibody [EPR20202-150] (Abcam Cat# ab243230); has been referenced in 2 publications: https://www.abcam.com/products/primary-antibodies/bst2tetherin-antibody-epr20202-150-ab243230.html

Rabbit monoclonal anti-hCG beta [5H4-E2] (Abcam Cat# ab9582); has been referenced in 22 publications: https://www.abcam.com/products/primary-antibodies/hcg-beta-antibody-5h4-e2-ab9582.html

Rabbit monoclonal anti-Gata6 (clone D61E4) (Cell Signaling Cat# 5951); has been referenced in 91 publications: https://www.cellsignal.com/products/primary-antibodies/gata-6-d61e4-xp-rabbit-mab/5851

Rabbit monoclonal anti-Islet1 [EP4182] (Abcam Cat# ab109517); has been referenced in 52 publications: https://www.abcam.com/products/primary-antibodies/islet-1-antibody-ep4182-neural-stem-cell-marker-ab109517.html

Mouse monoclonal anti- Anti-TFAP2a (AP-2α) (3B5) (Santa Cruz Cat# SC-12726); has been referenced in 91 publications: https://www.scbt.com/p/ap-2alpha-antibody-3b5

Goat polyclonal anti-Sox2 (R&D Cat# AF2018); has been referenced in 195 publications: https://www.rndsystems.com/products/human-mouse-rat-sox2-antibody_af2018

Rabbit polyclonal anti-Dnmt3l (Imgenex Cat# IMG-6804A); has been referenced in 2 publications: https://www.novusbio.com/products/dnmt3l-antibody_nbp2-27098

Goat polyclonal anti-Otx2 (R&D Cat# AF1979); has been referenced in 81 publications: https://www.rndsystems.com/products/human-otx2-antibody_af1979

Mouse molonclonal anti-Stella (D-5 clone) (Santa Cruz Cat# SC-376862); has been referenced in 2 publications: https://www.scbt.com/p/stella-antibody-d-5

Rabbit monoclonal anti- Blimp1/PDRI-BF1 [Clone C14A4] (Cell Signaling Cat# 9115); has been referenced in 69 publications: https://www.cellsignal.com/products/primary-antibodies/blimp-1-prdi-bf1-c14a4-rabbit-mab/9115

Goat polyclonal anti-FoxF1 (R&D Cat# AF4798); has been referenced in 10 publications: https://www.rndsystems.com/products/human-mouse-foxf1-antibody_af4798?gclid=Cj0KCQjwoeemBhCfARIsADR2QCsjgD6WVshK8vZ2JC6h6tk6e4Wn2emSIPup0TER4GhFfDXQoWac2-AaAiJkEALw_wcB&gclsrc=aw.ds

Goat polyclonal Nidogen2 (R&D Cat# AF3385); has been referenced in 3 publications: https://www.rndsystems.com/products/human-nidogen-2-antibody_af3385

Rabbit monoclonal anti-Gata2 [EPR2822] (Abcam Cat# ab109241); has been referenced in 10 publications: https://www.abcam.com/products/primary-antibodies/gata2-antibody-epr28222-ab109241.html

Antibodies for flow cytometry: All the antibodies guarantee covers the use of the antibody for flow cytometry applications.

Mouse monoclonal anti human TROP2-488 labeled (R&D Cat# FAB650G). The antibody has been pre-titrated and tested by flow cytometry analysis of PC-3 human prostate cancer cell line. The antibody has been referenced in 2 publications: https://www.rndsystems.com/products/human-trop-2-alexa-fluor-488-conjugated-antibody-77220_fab650g

Mouse monoclonal anti human CD249 (ENPEP)-BV421 labeled (BD Cat# 744872). The production process of this antibody underwent stringent testing and validation to assure that it generates a high-quality conjugate with consistent performance and specific binding activity. This antibody has been referenced in 4 publications: https://www.bdbiosciences.com/en-eu/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/bv421-mouse-anti-human-cd249.744872

Rat monoclonal anti mouse CD140a (PDFGR-a)-PE/Cy7 labeled (BioLegend Cat# 135912). Each lot of this antibody is quality control tested by immunofluorescent staining with flow cytometric analysis. This antibody has been referenced in 5 publications: https://www.biolegend.com/en-us/products/pe-cyanine7-anti-mouse-cd140a-antibody-14822?GroupID=BLG8103

Mouse monoclonal anti human CD140a (PDFGR-a)-PE/Cy7 labeled (BioLegend Cat# 323508) and APC labeled (BioLegend Cat# 323512). Each lot of this antibody is quality control tested by immunofluorescent staining with flow cytometric analysis. This antibody has been referenced in 15 publications: https://www.biolegend.com/en-us/products/pe-anti-human-cd140a-pdgfralpha-antibody-3727?GroupID=BLG5119

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	WIBR1 human male, WIBR2, WIBR3 Human female embryonic stem cell lines were previously reported in Lenger et al. Cell 2010 and provided by the last author of that paper: Prof. Rudolf Jaenisch, Whitehead Institute of Science, Cambridge, MA, USA. RUES2 hESC line was previously described in Simunovich et al. Cell Stem Cell 2002 and provided by the last author, Prof. A Brivanlou, Rockefeller University, USA.
Authentication	Karyotype and sequencing data confirmed expected sex and karyotype, gene reporters and cell identity via SNPs.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination by using the MycoAlert plasma Detection Kit (Lonza, Cat# LT07-318) and were routinely screened every 1 month.
Commonly misidentified lines (See ICLAC register)	HEK293T cells were used for lentivirus generation only that was used to permanently labeled some cell lines as indicated in the paper. HEK293T cells were cultured in a dedicated tissue culture room that is separate from where ESC culture and SEM generation were performed.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Mus Musculus (mouse) ICR strain derived embryo samples were used as reference controls for mouse SEM related experiments. 4-10 week old male and female ICR mice were used for timed matings for natural embryo dissection.
Wild animals	The study did not involve wild animals
Reporting on sex	Sex of ES and iPS lines used (male and female) is indicated for all lines used and we do not report any sex bias of difference in result outcome. WIBR1 is a male hESC lines, WIBR2 and WIBR3 are female hESC line. JH22 and JH33 are male human iPSC lines. V6.5 and BVSC arei male mouse ESC line.
Field-collected samples	The study did not involve samples collected from the field
Ethics oversight	Mouse animal experiments pertained only to mouse SEM and comparing them to mouse embryos, and were performed according to the Animal Protection Guidelines of Weizmann Institute of Science and approved by the following Weizmann Institute IACUC (#01390120-1, 01330120-2, 33520117-2).

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

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were incubated for half an hour with fluorophore-conjugated antibodies (1:50) in PBS/0.5% BSA.
Instrument	BD FACS-Aria III
Software	FlowJo v10.7
Cell population abundance	Only one cell population was analyzed post-sorting, and the purity was verified by resampling.
Gating strategy	FSC and SSC singlets were gated to remove debris and aggregated cells, and only single cells were considering for all analyses. To determine the gating for positive or negative populations, an unstained control and naive PSCswere employed, making sure that approximately 100% of the unstained population was allocated on the negative area of the histogram/dot plot. Gating strategies are included in the last Supplementary Figure S17.

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