Essential amnion signals for primate primitive streak formation resolved by scRNA map

Ran Yang[1,#], Alexander Goedel[1,#], Yu Kang[2,#], Chenyang Si[2], Chu Chu[2], Yi Zheng[3], Zhenzhen Chen[2], Peter J. Gruber[1,4], Yao Xiao[1], Chikai Zhou[1], Chuen-Yan Leung, Yongchang Chen[2], Jianping Fu[3,5,6], Weizhi Ji[2,*], Fredrik Lanner[7,8,*], Yuyu Niu[2,9,*], Kenneth Chien[1,*]

1. Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm 171 77, Sweden.
2. Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, Yunnan, 650500, China.
3. Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.
4. Department of Surgery, Yale University, New Haven, CT 06520, USA.
5. Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA.
6. Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.
7. Department of Clinical Sciences, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden and Ming Wai Lau Center for Reparative Medicine, Stockholm node, Karolinska Institutet, Stockholm 171 77, Sweden.
8. Division of Obstetrics and Gynecology, Karolinska Universitetssjukhuset, Stockholm 141 86, Sweden.
9. Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, Yunnan 650500, China.

# Authors contributed equally.

* Correspondence: kenneth.chien@ki.se, fredrik.lanner@ki.se, niuyy@lpbr.cn, wji@lpbr.cn.

Abstract

The signaling network governing the formation of the primitive streak is well understood in mice, but largely unexplored in primates. Advances in single-cell technology and *in vitro* embryo culture have enabled to characterize the major cell populations involved. However, a detailed map of this process and insights into its regulatory networks are lacking. Herein, we generated a serial single cell atlas of over 30,000 cells spanning peri-implantation to early primitive streak stages in non-human primates (NHP) describing the emergence of the primitive streak, extraembryonic mesenchyme and amnion. We discovered that *ISL1*, a gene with a well-established role in cardiogenesis, controls a gene regulatory network in primate amniotic cells. Strikingly, CRISPR/Cas9-targeting of *ISL1* resulted in NHP embryos failing to form primitive streak. BMP4 was identified as a key signaling pathway in this process. This was confirmed in human embryonic stem cell lines, suggesting a conserved function in humans. Notably, no viable *ISL1* hypomorphic NHP mutant embryos could be recovered after embryo transfer confirming the essential requirement of *ISL1*. This highlights the importance of the amnion as a signaling center during primate embryogenesis and demonstrates the potential of *in vitro* primate model systems to investigate the genetics of early human development.

Introduction

One of the key steps during mammalian early development is the formation of the three primary germ layers. This occurs in a complex process termed gastrulation where cells from the columnar shaped epiblast undergo epithelial-to-mesenchymal transition and move anteriorly to form the primitive streak and the first mesodermal cells [1-3]. Since this step is crucial for further development of the embryo, disturbances in this process lead to embryonic lethality. It is believed that improper gastrulation occurs frequently in human embryos and accounts for a significant proportion of early miscarriages in the human population.

The tight regulatory network governing this process has been well studied during murine embryonic development [1, 4], but is largely elusive in humans. Despite the fact that mouse and primate embryos have a similar appearance before the implantation stage, the transcriptome already differs in key aspects [5, 6]. After implantation, the differences become more evident. Mouse embryos form a cup-like structure, while primate embryos acquire a disk-like shape and have a prominent amnion, a structure which is absent in mouse embryos before gastrulation [7, 8]. Notably, some evidence from *in vitro* models of early human development suggest that amniotic cells could be involved in the signaling network that governs primitive streak

formation [9]. Unravelling the pathways that form this regulatory network in primates is key for understanding primitive streak formation in humans and for identifying potential causes for congenital malformations and early pregnancy loss. Recently, two publications on cynomolgus embryogenesis [10, 11] and one publication on human embryogenesis [12] have created a framework of post-implantation development in primates and characterized the major cell populations involved in gastrulation. However, their interplay and the transcriptional networks guiding this essential step remain unknown.

Here, we created a high-resolution map of the peri-gastrulation development of early NHP embryogenesis with a focus on the signaling events leading to the formation of the primitive streak. We identified an *ISL1*-dependent gene regulatory network that is specifically active in amnion. Disturbance of this network in NHP embryos by CRIPSR/Cas9-mediated gene-editing of *ISL1* led to significant downregulation of BMP4 signaling from the amnion and subsequent failure to from the primitive streak. We confirmed these findings in a microfluidic-based model of amnion-epiblast interactions using *ISL1*-null human embryonic stem cells suggesting that these findings also apply to humans. Taken together, this study shows for the first time that signals from the amnion are indispensable for primitive streak formation in primate embryos.

Results

High-resolution transcriptomic map

We created a high-resolution transcriptomic map by single cell RNA (scRNA) sequencing of 11 *in vitro* cultured cynomolgus macaque embryos at three different time points (Day 10, Day 12 and Day 14) (Fig. 1A). 7194 cells passing quality control (Fig. S1A-B) were embedded for each day separately in low-dimensional space (Fig. 1b and Fig. S2A-B). In line with previous results [10, 11], the cells grouped into four main cell types, namely trophoblast, endoderm, epiblast with its derivatives and extraembryonic mesenchyme (Fig. 1B-C). Integration of our dataset with a published scRNA sequencing dataset of *in vivo* cynomolgus embryos [13] (Fig. 1D-E and Fig. S2C) revealed a striking difference in the transcriptomic profile between early (Day 10 + E08/E09) and late (Day 12/Day 14 + E13/E14) epiblast, reflecting the transition from a naïve to a primed state, which has been suggested before to happen during this time window [14]. Indeed, a published gene signature of naïve cells derived from scRNA sequencing analysis of *in vitro* cultured human embryonic stem cells (hESCs) [15], was highly enriched in the early epiblast (Fig. 1F), while genes belonging to the primed signature were enriched in the late epiblast (Fig. 1F). Aligning cells from the early and late epiblast in pseudotime disclosed a set

of differentially regulated genes that formed two distinct clusters based on their expression dynamics (Fig. 1G). Genes previously associated with a naïve state [14-16], including *DNMT3L*, *KHDC1L*, *NLRP7*, *OOEP* and *DPP4* were significantly downregulated over pseudotime (Fig. 1H). Notably, a number of those genes belong to the subcortical maternal complex, which is crucial for proper early development [17]. In contrast, genes associated with a primed state [16, 18, 19], including *CD24*, *CRABP2*, *SFRP1*, *USP44* and *VCAN* showed strong upregulation (Fig. 1H). This expression pattern was observed in cells from our dataset as well as in cells from the *in vivo* dataset, suggesting that the naïve to primed transition happening *in vivo* can faithfully be recapitulated in *in vitro* cultured embryos.

Gene regulatory network of early epiblast

Taking advantage of our high-resolution scRNA map, we next investigated the sequence of appearance of epiblast-derived cell populations. Strikingly, we identified cells expressing amnion marker genes *ISL1* and *TFAP2A* [10], although very low in numbers, as early as Day 10 (Fig. 2A and Fig. S2D). Embryos at this stage also consisted of a large cell population expressing genes typical of a mesodermal signature such as *MIXL1* and *MESP1* (Fig. S2E), which were previously annotated as early gastrulating cells [10]. However, in addition to their mesodermal signature, these cells show high expression of the transcription factor *ETS1* and the cell adhesion protein *PODXL* (Fig. S2F), which mark extraembryonic mesoderm in mice [20], while lacking expression of the receptor tyrosin-kinase *EPHA4* or the transcription factor *ZIC3* (Fig. S2G) expressed by murine embryonic mesoderm [20]. Thus, we provide strong evidence, that the epiblast-derived mesodermal-like cells present in Day 10 embryos are extraembryonic mesoderm, which appear to precede primitive streak formation.

Extraembryonic mesenchyme, a cell population unique to primates that contributes to a number of extraembryonic tissues [7, 13, 21, 22], was first present in Day 12 embryos (Fig. 2A). The close proximity of extraembryonic mesoderm and mesenchyme in the UMAP plot (Fig. 2A) as well as the expression pattern of *PODXL* and *ETS1* (Fig. S2H), advocate that extraembryonic mesoderm contributes to extraembryonic mesenchyme, as was suggested previously [23]. However, the large increase in cell number in the extraembryonic mesenchyme over a few days suggests, that this cell population might get additional contributions from other parts of the embryo, such as the trophoblast. The formation of the primitive streak, marked by expression of the transcription factors *MESP1* and *GSC* (anterior) [24, 25], as well as *HOXD9* and *CDX2* (posterior) [26] (Fig. 2B and Fig. S2I-J), was observed first at Day 14 (Fig. 2A). Late amniotic cells had a distinct expression profile with high expression of the transcription factor *ISL1*, the

Wnt-ligand *WNT6* and a subunit of the GABA-receptor *GABRP* (Fig 2B, Fig. S2K). Transcripts of these markers were specifically enriched in amniotic cells across the whole dataset including trophoblast, while other markers such as *HEY1* were also expressed in subpopulations of the trophoblast (Fig. S2L). Cells classified as early amniotic cells expressed amniotic markers in addition to marker genes of the posterior primitive streak, such as *CDX1* [26] (Fig. 2B, Fig. S2M).

Gene regulatory network (GRN) analysis at Day 14 using SCENIC (Single-cell regulatory network inference and clustering) [27] revealed sets of highly specific GRNs (Fig. 2C-D). The histone demethylase *KDM5B*, creating bivalent histone marks during development [28], controlled a GRN active in epiblast and all its derivatives, while the pluripotency factor *SOX2* [19] controlled a network specifically active in the epiblast (Fig. 2D). One of the most active GRNs in amniotic cells was an *ISL1*-dependent network (Fig. 2D), suggesting that *ISL1* is not only a specific marker of the amniotic cell population in primates, but also likely to play a functional role in amniotic cells. This finding is in sharp contrast to the mouse, where *Isl1* is first expressed in cardiac progenitor cells of the lateral plate mesoderm, but absent from the early embryo before E7.0 [29, 30].

ISL1 hypomorphs fail to form primitive streak

The functional role of *ISL1* in primate amniotic cells was investigated by generating *ISL1* hypomorphic mutant NHP embryos through one-cell stage CRISPR/Cas9 injections with gRNAs creating a long deletion in the *ISL1* locus (Fig. S3C). *In vitro* cultured hypomorphic mutant embryos grew normally up to Day 10 (Fig. 3A and Fig. S3A). From Day 12 onwards the mutant embryos progressively lost structure and at Day 14, the embryonic disk was no longer distinguishable (Fig. 3A and Fig. S3A).

Genotyping of the mutant embryos revealed a mosaic pattern with presence of ind/dels of different sizes in the targeted region of the *ISL1* locus and no alterations in selected off-targets from the in-silico prediction [31] (Fig. S3B-C). Interestingly, the frequency of mutations resulting in an internally shortened, but functional *ISL1* was significantly higher in the transcriptome (Fig. S3C). This could be due to nonsense-mediated decay of mutated *ISL1* message or suggest that cells with a complete loss of function of *ISL1* are not able to enter an amniotic fate. Overall, this resulted in a 50% reduction of functional *ISL1* message in hypomorphic mutant embryos.

To gain mechanistic insights, *in vitro* cultured *ISL1* hypomorphic mutant embryos were subjected to scRNA-sequencing at the same time points as for the wildtype (Day 10, Day 12,

and Day 14). Integrated analysis of 26136 cells passing quality control (Fig. S1A-B) with the wildtype dataset revealed the presence of similar populations at Day 10 and Day 12 (Fig. 3B-C, Fig. S4A-C). However, contribution to the cluster identified as extraembryonic mesoderm was reduced in the mutant embryos while amniotic cells were slightly more abundant (Fig. 3B-C). At Day 14, primitive streak cells were virtually absent from the mutant embryos accompanied by a drastic overrepresentation of amniotic cells (Fig. 4B-C). The absence of primitive streak in the mutant embryos was confirmed by a loss of expression of primitive streak markers *EOMES*, *MIXL1*, *MESP1*, *CDX1* and *CDX2* [24, 26, 32] across the embryos (Fig. 3D). Moreover, cells from the mutant embryos formed an additional cell cluster labeled "abnormal amnion" that was void in wildtype (Fig. 3B-C). In this cluster, expression of amniotic marker genes such as *HEY1* and *ISL1* [10] were present, but lower levels of *GABRP* or *TFAP2A* [10] (Fig. S4D). Instead, high levels of the mesenchymal marker *VIM* and extracellular matrix genes like *LUM*, *COL1A1* or *TGFBI* were found (Fig. S4E). Late amniotic cells of the mutant embryos showed a similar expression pattern indicating an abnormal function of amnion in mutant embryos. Differential gene expression analysis of amniotic cells in mutant versus wildtype embryos revealed that a high proportion of the significantly downregulated genes in the mutant amnion were members of the identified *ISL1* regulon (Fig. 3E-F). Among those was *BMP4,* a secreted member of the *TGFb*-signaling pathway and a known downstream target of *Isl1* in mice [33] that is essential for murine mesoderm formation [34] as well as for inducing primitive streak like cells from hESCs [32] (Fig. 3F). Taken together, this suggests that *BMP4* is secreted from the amnion in an *ISL1*-dependend manner to induce primitive streak formation in the early NHP embryo.

Loss of *ISL1* in amnion impairs BMP4 signaling

To validate this conclusion and to investigate whether these findings also apply to humans, *ISL1-null* hESCs (*ISL1*-null) harboring the most abundant long deletion in the *ISL1* locus found in the mutated NHP embryos (Fig. S3C and S5A) were generated and analyzed *in vitro* [35] (Fig. 4A). Amniotic ectoderm-like cells (AMLCs) derived from the *ISL1*-null showed a 50% reduction in *ISL1* mRNA (Fig. 4B) and absence of ISL1 protein (Fig. 4C) which was abundantly expressed in wildtype hESCs derived AMLCs (Fig. 4C). In concordance with the findings from the *in vitro* cultured NHP embryos, AMLCs derived from the *ISL1*-null showed higher mRNA and protein levels of the amnion markers *TFAP2A* and *GATA3* (Fig. 4D-E). In addition, we noticed a slight reduction in *WNT6* and a significant reduction in *BMP4* expression, suggesting a functional deficit in *ISL1*-null derived AMLCs (Fig. 4F). Indeed, AMLCs derived from the

ISL1-null failed to induce mesoderm-like cells (MeLCs) from hESCs shown by a lack of Brachyury (BRA) signal (Fig. 4G-H). Notably, *BRA* expression was similar between *ISL1*-null and wildtype using a directed differentiation protocol towards anterior and posterior primitive streak cells [32] (Fig. S5B). This highlights that the failure of mesoderm induction in the *ISL1*-null is a non-cell autonomous defect caused by altered signaling from AMLCs.

The capacity of *ISL1*-null and wildtype hESCs to self-organize into an embryonic-like sac was assessed in a microfluidic system that has been shown to faithfully recapitulate the peri-implantation development of the epiblast lineages [9]. With a reduced BMP4 dose as compared to the protocol used in the original publication, the wildtype cells still showed proper formation of embryonic-like sacs, adequate break of symmetry and formation of MeLCs in the epiblast-like region as shown by strong signals for BRA (Fig. 4I and Fig. S6). *ISL1*-null cells were capable of self-organizing into embryonic-like sacs similar to the wildtype but failed to develop further. The epiblast-like cells remained in a columnar shape with high levels of the pluripotency factor NANOG and absence of BRA [19] (Fig. 4I), indicating failure to form MeLCs similar to the findings observed in the mutant NHP embryos during *in vitro* culture.

ISL1 mutant NHP embryos are not viable

Finally, we investigated whether the observed alterations during embryonic development in the *ISL1* hypomorphic mutant NHP embryos lead to early embryonic lethality *in vivo*. The pregnancy rate per NHP surrogate mother after transfer of *ISL1* mutant embryos was 0% as compared to 58.3% with wildtype embryos (Table 1). Transfer of embryos that were targeted with injection of only a single gRNA leading to a slightly lower mutation rate resulted in a pregnancy rate of 28.6% (Table 1). Strikingly, genotyping of all 4 fetuses from this experiment showed an unmodified *ISL1* locus on both alleles, highlighting the importance of *ISL1* for proper early embryonic development.

Discussion

In this study, we generated a high-resolution developmental roadmap of post-implantation NHP embryos and identified the amnion as a key signaling structure essential for initiating gastrulation in primates. NHP embryos hypomorphic for the amnion marker *ISL1* failed to form primitive streak due to a lack of BMP4-signaling and did not give rise to viable offspring, demonstrating a novel, primate-specific role of *ISL1* in early embryogenesis (Fig. 5).

It is known that the initiation of primitive streak formation is largely dependent on BMP4-

signaling, which is provided by the extraembryonic ectoderm in mice [1, 36]. The findings from our study suggest that this role is taken over by the amnion during primate embryogenesis. Notably, the role of ISL1 acting upstream of BMP4-signaling seems to be a conserved pathway. The complete loss of Isl1 in mice leads to embryonic lethality at approximately E10.5 with severe cardiac defects accompanied by a strong reduction in Bmp4 [29], suggesting that Bmp4 is acting downstream of Isl1 in cardiac development. Similar observations were made in mice during genital development [37] and embryonic limb formation [38].

Isl1 has a well-established role in mammalian cardiac development and is expressed in multipotent cardiovascular progenitor cells in mice [39-41] and humans [42, 43]. Despite its established role in heart development in mice, loss of function variants in the *ISL1* locus are drastically underrepresented in large human cohorts of congenital heart malformations like the Pediatric Cardiac Genomics Consortium (PCGC) [28, 44]. In detail, among the 23,000 alleles reported in the PCGC cohort, 112 *ISL1* variants have been identified, none of which were damaging *de novo* mutations. We hypothesize that the exceptionally low frequency of damaging *ISL1* variants reported in this cohort is due to its essential role in initiating primitive streak formation. Importantly, we did not observe any pregnancy with an *ISL1* hypomorphic mutant primate embryo after transferring almost 50 blastocysts in over 15 surrogate NHP mothers. This highlights the importance of two intact *ISL1* alleles for proper embryonic development and suggests that mutations in *ISL1* could be a cause for early pregnancy loss in humans.

This study shows that the amnion is a crucial signaling center during early primate embryogenesis and demonstrates that *in vitro* cultured primate embryos are a powerful tool to model key steps of early human development. Further advances in the *in vitro* culture systems might enable us to support the embryo longer and to study early organogenesis, including the emergence of cardiac progenitor cells in lateral plate mesoderm. This would enrich our knowledge on human embryogenesis, help to identify causes for pregnancy loss and congenital malformations and, eventually, open the avenue for new therapies.

Methods

Cynomolgus macaque

Healthy cynomolgus monkeys (Macaca fascicularis), ranging from 5 to 12 years old, were used in this study. All animals were housed either at the facility of Yunnan Key Laboratory of Primate Biomedical Research in China, or at Astrid Fagræus laboratory in Karolinska Institutet in Sweden. Both facilities are accredited by AAALAC international. The ethics and all experimental protocols were approved in advance by the Institutional Animal Care and Use Committee of LPBR in China (KBI K001115033/01,01) and by the Jordbruksverket in Sweden (ethical permit number N277/14). Animals involved in this study were never used for other treatments.

Culture of NHP embryos

NHP embryos were collected as described in previously publication [45]. In brief, healthy female monkeys aged 5-8 years with regular menstrual cycles were selected as oocyte donors. Before superovulation, female animals were treated with rhFSH (recombinant human follitropin alfa, GONAL-F, Merck Serono) for 8 days, and administrated rhCG (recombinant human chorionic gonadotropin alfa, OVIDREL, Merck Serono) on day 9. Oocytes were collected by laparoscopic follicular aspiration 32-35 hours after rhCG administration. MII (first polar body present) oocytes were performed with intracytoplasmic sperm injection to generate zygotes and the fertilization was confirmed by the presence of two pronuclei.

Zygotes were cultured in embryo culture medium-9 (HECM-9) containing 10% fetal calf serum (Hyclone Laboratories) in 37°C incubator supplying 5% CO_2 until blastocyst stage. Blastocysts were then used for embryo transfer or post-implantation *in vitro* culture.

To culture blastocyst beyond implantation stage, we applied an optimized protocol based on the human embryo culture protocol from Zernicka-Goetz's group [46]. Frozen NHP blastocysts were thawed right before culturing by using the Thawing Media from Kizatato and cultured in blastocyst culture medium for at least 4 hours to recover. Blastocysts were then treated with Acidic Tyrode's solution (Sigma) to remove the Zona pellucida and transferred to an ibiTreat 8-well μ-plate (Ibidi) with 300 μL of pre-equilibrated in vitro culture medium 1 (IVC1). On the second day, 150 μL of IVC1 was carefully removed and 200 μL pre-equilibrated IVC2 was added. Blastocyst growth was monitored and medium was changed every two days until termination of experiments.

Culture of NHP embryonic stem cells

The in-house generation of NHP embryonic stem cells (ESC) was performed according to an established human embryonic stem cell protocol [47]. Briefly, MitC-treated mouse embryonic fibroblast (MEF) feeder cells were seeded at a density of 4 x 10^4 per cm^2. Media was changed to ESC medium before use. After removing the Zona pellucida the ICSI generated blastocysts were seeded on top of the prepared feeder cells and cultured without disturbance for three days in CO_2 incubator at 37°C. Fresh ESC medium was changed daily until primary ESC colonies formed. Colonies were cut into small pieces and passaged on fresh MEF feeder cells with ESC medium. Subsequently, ESC colonies were passaged with Collagenase type IV (STEMCELL Technologies) and seeded on fresh MEF feeder cells.

Culture of human embryonic stem cells

Human embryonic stem cells used in this study include HES-3 human ES cells (ES03, WiCell) and H9 human ES cells (WA09, WiCell). Genetic modification of the *ISL1* locus to generate *ISL1-null* hESCs was performed on HES-3 cells by applying CRISPR/Cas9 with the same guide RNAs used in NHP blastocysts. Two *ISL1* knockout cell lines were generated, named ISL1_ko_c15 and ISL1_ko_c51, and genotyped (Fig. S3c). All cell lines were authenticated as karyotypically normal by Cell Guidance Systems (United Kingdom) (Fig. S5a). Mycoplasma contamination test was performed regularly as negative (EZ-PCR Mycoplasma Detection Kit, Biological Industries). hESCs were maintained in a standard feeder-free culture system using mTeSR1 medium (STEMCELL Technologies) on 1% Matrigel (Corning) or Essential 8 medium (Thermo Fisher Scientific) on 1% Vitronectin (Thermo Fisher Scientific). Cells were passaged every 4-5 days and visually examined during each passage to ensure absence of spontaneously differentiation. Work with human embryonic stem cells was carried out according to Swedish legislation following the recommendations of the Swedish National Council on Medical Ethics.

NHP embryo transfer and pregnancy diagnosis

Embryos with high quality were transferred into the oviducts of the matched recipient monkey as described in previous study [48]. A total of 27 female monkey recipients with proper hormone level of β-estradiol and progesterone were used as surrogate recipients. Each recipient received 2-4 blastocysts. The pregnancy was primarily diagnosed by ultrasonography at 2-3 weeks after embryo transfer. Clinical pregnancy and the number of fetuses were confirmed by fetal cardiac activity and the presence of gestation sacs. When terminating pregnancy, caesarean section was performed. Tissue from umbilical cord, ear and tail was collected for genotyping.

Generation of *ISL1* hypomorphic mutant NHP embryos

NHP zygotes were injected with mix of Cas9 protein and guide RNAs. Intracytoplasmic injections were performed with a Nikon microinjection system under standard conditions. The embryos were cultured in HECM-9 supplemented with 10% fetal calf serum (Hyclone Laboratories) in 37°C incubator supplying 5% CO_2. Genetic modified embryos with high quality from morula to blastocyst stage were used for further studies.

Single cell dissociation and RNA sequencing

NHP embryos at Day 10, Day 12 and Day 14 were washed with PBS and treated with TrypLE Express Enzyme (Thermo Fisher Scientific) for 30 minutes at 37°C. After incubation, samples were gently dissociated into single cells by mouth pipetting. Single cells were transferred into a RNase-free, low-cell-adhesion 1.5 mL tube and centrifuged at 300 *g* for 5 minutes. Supernatant, containing some remaining cells was transferred into a new tube for genotyping. The cell pellet was resuspended with 40 µL of PBS containing 2% Bovine Serum Albumin (Sigma) in PBS. Cells were loaded into the 10x Genomics Chromium system within 30 minutes after dissociation. 10x Genomics v3 libraries were prepared according to the manufacturer's instructions. Libraries were sequenced with a minimum coverage of 30 000 raw reads per cell on an Illumina NovaSeq with paired-end sequencing.

Reads mapping, gene expression counting and correction

Sequencing data was aligned and quantified by using the Cell Ranger Pipeline v3.1.0 (10x Genomics) against the ensemble genome build Macaca_fascicularis_5.0 release 96 [49]. Ambient RNA contamination was estimated through the levels of choriogonadotropins expression in epiblast (POU5F1 positive) cells and removed from the count matrix using SoupX [50]. A gene was retained for analysis if it showed expression in at least 3 cells. Each sample was filtered based on expression level of mitochondrial genes (below 7.5%) and number of expressed genes. Details on the estimated contamination in each sample, the filtering criteria and number of cells retained for the analysis are provided in Fig. S1a.

Reads mapping and gene expression counting of in vivo dataset

The raw, archived single cell RNA sequencing data from in vivo cynomolgus embryos [13] was downloaded from the GEO database (GSE74767) and processed using TrimGalore v0.6.1. The reads passing quality control were aligned against the ensemble genome build macaca_fascicularis_5.0 release 96 using STAR v2.5.3 and counted using featureCounts v1.5.2. Cells expressing at least 1000 genes were kept for the integration with our dataset.

Data integration, dimensionality reduction and clustering

For analysis of the single cell RNA sequencing data from the wildtype in vitro cultured embryos, we integrated the filtered, corrected count matrices of the different batches for each day separately using the reciprocal principal component analysis (PCA) approach implemented in the Seurat package v3.1.3 [51, 52] based on 30 dimensions and 5000 anchor features. After integration, we performed PCA analysis on the integrated data followed by embedding into low dimensional space with Uniform Manifold Approximation and Projection (UMAP) as implemented in the R-package 'uwot'. For clustering, the shared Nearest Neighbor (SNN) Graph was constructed on the UMAP embedding by calling the FindNeighbors() function followed by the identification of clusters using the FindClusters() function, both part of the Seurat package. In some samples this clustering approach separated large, homogeneous cell groups into small sub-clusters with no distinct biological meaning. In these cases, the clusters were re-combined manually and both, the unsupervised and the manually adjusted clustering, was reported in the manuscript.

To integrated the single cell RNA sequencing data from in vivo cynomolgus embryos [13] with our dataset we combined the three time points (Day 10, Day 12 and Day 14) from our dataset for each batch separately and did the same for the in vivo data from embryonic day 8 (E08), E09, E13 and E14 resulting in three separate datasets. Normalization, Scaling and PCA was performed separately on each of these datasets after which they were combined using the reciprocal PCA approach described above based on 30 dimensions and 2000 anchor features. For the analysis of the wildtype and *ISL1* hypomorphic mutant embryos the different batches were integrated separately for each day by the same reciprocal PCA approach outlined above based on 30 dimensions and 5000 anchor features using the wildtype datasets as reference. Dimensionality reduction and clustering was performed as described above.

Differential gene expression analysis

Mainly due to the differences in cell numbers we observed a significant variation in sequencing depth between samples in our dataset (Fig. S1). It has recently been shown, that the effect of differences in read depth on differential gene expression analysis can be minimized by using regularized negative binomial regression as implemented in the R-package SCtransform [53]. Thus, all differential gene expression analysis was performed using a t-test on Pearson residuals after SCtransformation of the raw, filtered counts of the integrated Seurat object as implemented previously [53]. Gene expression data depicted throughout the manuscript in feature plots or violin plots are SCtransformed data. Expression data depicted in heatmaps are scaled, log-transformed expression values normalized to the total counts for each cell calculated

through running the NormalizeData() function followed by the ScaleData() function from the Seurat package.

Visualization of gene signatures

Scoring and visualization of gene signatures was performed using the Single Cell Signature Explorer v3.1 [54]. Gene signatures were created by identifying orthologues for the genes that have been previously described to mark the naïve and primed state of pluripotency in human embryonic stem cells [15] in the Macaca fascicularis genome using the according orthologue list provided from Ensemble through BioMart [55].

Pseudotime analysis

Pseudotime analysis was performed using Monocle3 v0.2 [56-58]. The principal graph was learned on the UMAP embedding extracted from the integrated Seurat object. Differentially expressed genes were calculated on the raw, filtered count matrix extracted from the integrated Seurat object using the Moran's I test implemented in the graph_test() function from the Monocle3 package. The genes were ranked according to their Moran's I and the top 100 genes were selected for display in the heatmap.

Gene regulatory network analysis

Gene regulatory network (GRN) analysis was performed using the R-package SCENIC (Single Cell rEgulatory Network Inference and Clustering) v1.1.2-2 [27] and the command line interface (CLI) of the python implementation pySCENIC. Macaca fascicularis gene IDs were converted to homo sapiens gene symbols using the according orthologue list from Ensemble. The raw, filtered count matrix extracted from the integrated Seurat object was pre-filtered and genes with at least 39 counts, equal to at least 3 UMI counts in 1% of the cells, present in at least 13 cells, equal to 1% of the cells, were used as input for the CLI. The human motif collection v9 and the cisTarget databases for hg38 were used in the pipeline and downloaded from https://resources.aertslab.org/cistarget/. Thresholds used for binarization were derived from the AUC values using Hartigan's Dip Test (HDT). After binarization, regulons showing activity in at least 1% of the cells were included in the downstream analysis.

Genotyping

Genomic DNA was extracted by Phenol-Chloroform method. DNA fragment covering both guide RNA target sites were PCR amplified and ligated to TOPO TA cloning vector (Thermo Fisher Scientific). At least 50 bacteria clones per sample were picked for Sanger sequencing and used to estimate the genomic mutation rate. The transcriptomic mutation rate of *ISL1* hypomorphic mutants was also calculated. cDNA libraries of each scRNA-sequencing sample

were used to amplify the *ISL1* mRNA fragment covering both guide RNA target sites. PCR products were ligated into TOPO TA cloning vector (Thermo Fisher Scientific). At least 50 clones per cDNA library sample were picked and performed Sanger sequencing.

Off-target assay

Cas-OFFinder was applied to search for potential off-target sites with maximal two mismatches and two bulges [31]. Among all off-target candidates of both gRNAs, targets located on gene exons were selected for test. The DNA fragments of target sites were PCR amplified and the sequences were confirmed by Sanger sequencing.

RNA extraction and quantitative real-time PCR

Total mRNA was extracted by Direct-zol RNA miniprep kits (Zymo) and reverse transcription to cDNA library was prepared by GoScript Reverse Transcriptase (Promega). Quantitative real-time PCR was performed by PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) on ABI 7500Fast machine.

Transwell assay

The transwell assay was performed based on previous work by Zheng and colleagues [9]. In brief, it was performed on Transwell 12-well plates with permeable polyester membrane inserts (0.4 μm, Corning). The membrane inserts were coated with 1% Geltrex diluted in DMEM/F12 for 1 hour before use. hESCs were collected and re-suspended in culture medium containing Y-27632 (10 μM, Tocris), and seeded onto the membrane insert at a density of 3 x 10^4 cells per cm^2. Eighteen hours after seeding, culture medium was changed to E6 medium supplemented with bFGF (20 ng/mL, R&D System) and BMP4 (50 ng/mL, PeproTech) and cultured for 48 hours. Undifferentiated hESCs were collected, re-suspended in E6 supplemented with bFGF (20 ng/mL) and seeded at a density of 9 x 10^4 per well on freshly coated 12-well plates. The membrane inserts were washed with E6 + bFGF and transferred on top of the re-seeded hESCs. Cells were collected after 48 hours for analysis. Two wildtype hESC-lines (HES-3 and H9) and two *ISL1*-null lines were used in this assay. Both of the wildtype cell lines showed comparable results, as did the two *ISL1*-null lines.

Primitive streak induction from hESCs

Differentiation of hESCs to primitive streak-like cells was done in chemically defined media (CDM) as previously described [32]. In brief, posterior primitive streak was induced by the addition of bFGF (20 ng/ml, R&D System), the phosphoinositide 3-kinases (PI3K)-inhibitor LY294002 (10 μM, Tocris) and BMP4 (10 ng/ml, PeproTech). Anterior primitive streak was

induced with the same factors and, additionally, Activin A (50 ng/ml, PeproTech). After 40h cells were harvested. RNA extraction, reverse transcription and quantitative real-time PCR were performed as detailed below with 200 ng RNA as input for RT-reaction. All experiments were performed in at least biological triplicates.

Microfluidic assay

This assay was performed as previously described [35]. Briefly, the microfluidic device is fabricated by bonding a PDMS structure layer to a coverslip. Geltrex is diluted to 70% using E6 medium and loaded into the central gel channel separated from the side channels by trapezoid-shaped supporting posts. Upon gelation, Geltrex matrix would generate concave Geltrex pockets between supporting posts for cell seeding. hESCs suspended in mTeSR1 medium was introduced into the cell loading channel and allowed to settle and cluster in the gel pockets. After hESCs cluster formation, mTeSR1 medium was replaced by a basal medium (E6 and 20 ng/mL bFGF), and 20 ng/mL BMP4 was supplemented only into the cell seeding channel. After 18 hours of BMP4 stimulation, the BMP4 medium was replaced by the basal medium. The microfluidic devices were fixed at 48 hours since the hESCs clusters were exposed to BMP4.

Immunohistochemistry

Immunohistochemistry of cells from the transwell assay was performed following standard procedures. Briefly, cells were fixed in 2% paraformaldehyde for 30 minutes at room temperature and washed with PBS. Cells were blocked in blocking buffer (serum diluted in PBS with 0.1% Triton X-100) for one hour and then incubated with primary antibodies diluted in blocking buffer overnight at 4°C. Cells were washed with PBS supplemented with 0.1% Tween-20 (PBS-T) and incubated with secondary antibodies diluted in blocking buffer for 2 hours at room temperature. After incubation, secondary antibodies were washed off by PBS-T, and the samples were mounted for imaging. Staining of embryonic-like sac structure was performed as previously described [35]. Confocal micrographs were acquired by Zeiss 700 LSM confocal microscope or Olympus spinning-disc confocal microscope (DSUIX18) equipped with an EMCCD camera (iXon X3, Andor). The bright-field morphologic images of embryonic-like sacs were acquired by Zeiss Observer.Z1 microscope equipped with a monochrome CCD camera (AxioCam, Carl Zeiss MicroImaging). Images were analyzed by iMaris (Bitplane).

Quantification and Statistical Analysis

Values are shown as the mean value plus SEM. Continuous data was analyzed using student's t-test. P-values or adjusted p-values (where appropriate) below 0.05 were considered statistically significant. Details on the samples (e.g. number of biological replicates) are indicated in figure legends. Graphs were generated using Prism or R.

Data and code availability

The raw data, unfiltered count matrix and processed count matrix are deposited in the Gene Expression Omnibus (GEO) database with the accession number GSE148683 and will be publicly released upon publication. All code is available from the authors upon request.

Acknowledgments

We thank the Eukaryotic Single Cell Genomics facility, hosted at SciLifeLab, for single cell RNA sequencing service. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2019/8-234. Martin Dahlö at UPPMAX is acknowledged for assistance concerning technical and implementational aspects in making the code run on the UPPMAX resources. We would like to thank Xiaobing He and Jesper Sohmer for technical support and want to acknowledge Dr. Xidan Li for advice on the bioinformatic analysis. This study was supported by grants from the National Key Research and Development Program of China (2016YFA0101401) and a Grant for Swedish-Chinese collaboration from the Swedish Research Council (Dnr 539-2013-7002). K.C. was supported by a Grant for distinguished professors of the Swedish Research Council (Dnr 541-2013-8351). F.L. was supported by Ming Wai Lau Centre for Reparative Medicine, Knut and Alice Wallenberg Foundation, Centre for Innovative Medicine, Swedish Research Council, Ragnar Söderberg Foundation and Wallenberg Academy. A.G. was supported by a grant from the German research foundation (grant number: GO3220/1-1). Y.Z. and J.F. were supported by the National Institutes of Health (R21 NS113518) and the National Science Foundation (CMMI 1917304 and CBET 1901718).

Author Contributions

K.C. and W.J established the initial joint project and the specific research plan was conceived by K.C. and R.Y.. R.Y., A.G., F.L. and K.C. planned experiments, analyzed data interpreted the results and wrote the manuscript. C.L. and R.Y. designed the guide RNAs. C.C. tested the guide RNA efficiency on cynomolgus cells and embryos. Y.K. and C.S. generated the wildtype and mutant NHP blastocysts. Z.C. and Y.C. performed the NHP embryo collection and transfer.

C.Z thawed the NHP blastocysts. R.Y. performed the *in vitro* culture of embryos, collected samples for scRNA sequencing, did genotyping and off-target assay. A.G. analyzed the scRNA sequencing data, performed the primitive streak induction and related quantitative real-time PCR. Y.X. maintained human cell lines, prepared samples for karyotyping and assisted in many experiments. R.Y. did the transwell assay and related immunofluorescent staining and PCR amplification. Y.Z. and J.F. performed the microfluidic assay and related immunofluorescent staining. P.G. analyzed the genetic *ISL1* variants in the human population. Y.N., W.J., F.L. and K.C. supervised the study. R.Y., A.G. and Y.K. contributed equally to the study. Correspondence can be addressed to K.C., F.L., Y.N., or W.J..

Competing interests

The authors declare no competing interest.

References

1. Tam, P.P. and D.A. Loebel, *Gene function in mouse embryogenesis: get set for gastrulation.* Nat Rev Genet, 2007. **8**(5): p. 368-81.

2. Wamaitha, S.E. and K.K. Niakan, *Human Pre-gastrulation Development.* Curr Top Dev Biol, 2018. **128**: p. 295-338.

3. Kessler, D.S. and D.A. Melton, *Vertebrate embryonic induction: mesodermal and neural patterning.* Science, 1994. **266**(5185): p. 596-604.

4. Rossant, J., *Genetic Control of Early Cell Lineages in the Mammalian Embryo.* Annu Rev Genet, 2018. **52**: p. 185-201.

5. Boroviak, T., et al., *Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development.* Development, 2018. **145**(21): p. dev167833.

6. Petropoulos, S., et al., *Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos.* Cell, 2016. **165**(4): p. 1012-26.

7. Carlson, B.M., *Formation of Germ Layers and Early Derivatives*, in *Human Embryology and Developmental Biology*. 2014. p. 75-91.

8. Pereira, P.N., et al., *Amnion formation in the mouse embryo: the single amniochorionic fold model.* BMC Dev Biol, 2011. **11**: p. 48.

9. Zheng, Y., et al., *Controlled modelling of human epiblast and amnion development using stem cells.* Nature, 2019. **573**(7774): p. 421-425.

10. Ma, H., et al., *In vitro culture of cynomolgus monkey embryos beyond early gastrulation.* Science, 2019. **366**(6467): p. eaax7890.

11. Niu, Y.Y., et al., *Dissecting primate early post-implantation development using long-term in vitro embryo culture.* Science, 2019. **366**(6467): p. 837-+.

12. Xiang, L., et al., *A developmental landscape of 3D-cultured human pre-gastrulation embryos.* Nature, 2020. **577**(7791): p. 537-542.

13. Nakamura, T., et al., *A developmental coordinate of pluripotency among mice, monkeys and humans.* Nature, 2016. **537**(7618): p. 57-62.

14. Liu, D., et al., *Single-cell RNA-sequencing reveals the existence of naive and primed pluripotency in pre-implantation rhesus monkey embryos.* Genome Res, 2018. **28**(10): p. 1481-1493.

15. Messmer, T., et al., *Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution.* Cell Rep, 2019. **26**(4): p. 815-824 e4.

16. Shahbazi, M.N., et al., *Pluripotent state transitions coordinate morphogenesis in mouse and human embryos.* Nature, 2017. **552**(7684): p. 239-243.

17. Amoushahi, M., L. Sunde, and K. Lykke-Hartmann, *The pivotal roles of the NOD-like receptors with a PYD domain, NLRPs, in oocytes and early embryo developmentdagger.* Biol Reprod, 2019. **101**(2): p. 284-296.

18. Shakiba, N., et al., *CD24 tracks divergent pluripotent states in mouse and human cells.* Nat Commun, 2015. **6**: p. 7329.

19. Collier, A.J., et al., *Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States.* Cell Stem Cell, 2017. **20**(6): p. 874-890 e7.

20. Saykali, B., et al., *Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo.* Elife, 2019. **8**: p. e42434.

21. Knofler, M., et al., *Human placenta and trophoblast development: key molecular mechanisms and model systems.* Cell Mol Life Sci, 2019. **76**(18): p. 3479-3496.

22. Rossant, J. and P.P.L. Tam, *New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation.* Cell Stem Cell, 2017. **20**(1): p. 18-28.

23. Boroviak, T. and J. Nichols, *Primate embryogenesis predicts the hallmarks of human naive pluripotency.* Development, 2017. **144**(2): p. 175-186.

24. Kitajima, S., et al., *MesP1 and MesP2 are essential for the development of cardiac mesoderm.* Development, 2000. **127**(15): p. 3215-26.

25. Blum, M., et al., *Gastrulation in the mouse: the role of the homeobox gene goosecoid.* Cell, 1992. **69**(7): p. 1097-106.

26. Chawengsaksophak, K., et al., *Cdx2 is essential for axial elongation in mouse development.* Proc Natl Acad Sci U S A, 2004. **101**(20): p. 7641-5.

27. Aibar, S., et al., *SCENIC: single-cell regulatory network inference and clustering.* Nat Methods, 2017. **14**(11): p. 1083-1086.

28. Jin, S.C., et al., *Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands.* Nat Genet, 2017. **49**(11): p. 1593-1601.

29. Cai, C.L., et al., *Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart.* Dev Cell, 2003. **5**(6): p. 877-89.

30. Pijuan-Sala, B., et al., *A single-cell molecular map of mouse gastrulation and early organogenesis.* Nature, 2019. **566**(7745): p. 490-495.

31. Bae, S., J. Park, and J.S. Kim, *Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.* Bioinformatics, 2014. **30**(10): p. 1473-1475.

32. Mendjan, S., et al., *NANOG and CDX2 pattern distinct subtypes of human mesoderm during exit from pluripotency.* Cell Stem Cell, 2014. **15**(3): p. 310-325.

33. Gao, R., et al., *Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate.* Cell Res, 2019. **29**(6): p. 486-501.

34. Winnier, G., et al., *Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse.* Genes Dev, 1995. **9**(17): p. 2105-16.

35. Zheng, Y. and J. Fu, *Protocol for controlled modeling of human epiblast and amnion development using stem cells.* PROTOCOL (Version 1) available at Protocol Exchange, 2019.

36. Arnold, S.J. and E.J. Robertson, *Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo.* Nat Rev Mol Cell Biol, 2009. **10**(2): p. 91-103.

37. Ching, S.T., et al., *Isl1 mediates mesenchymal expansion in the developing external genitalia via regulation of Bmp4, Fgf10 and Wnt5a.* Hum Mol Genet, 2018. **27**(1): p. 107-119.

38. Yang, L., et al., *Isl1Cre reveals a common Bmp pathway in heart and limb development.* Development, 2006. **133**(8): p. 1575-85.

39. Domian, I.J., et al., *Generation of functional ventricular heart muscle from mouse ventricular progenitor cells.* Science, 2009. **326**(5951): p. 426-9.

40. Laugwitz, K.L., et al., *Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages.* Nature, 2005. **433**(7026): p. 647-53.

41. Moretti, A., et al., *Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification.* Cell, 2006. **127**(6): p. 1151-65.

42. Bu, L., et al., *Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages.* Nature, 2009. **460**(7251): p. 113-7.

43. Sahara, M., et al., *Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract.* Dev Cell, 2019. **48**(4): p. 475-490 e7.

44. Zaidi, S., et al., *De novo mutations in histone-modifying genes in congenital heart disease.* Nature, 2013. **498**(7453): p. 220-3.

45. Niu, Y., et al., *Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos.* Cell, 2014. **156**(4): p. 836-43.

46. Shahbazi, M., et al., *Culture of human embryos through implantation stages in vitro.* PROTOCOL (Version 1) available at Protocol Exchange, 2016.

47. Liu, X., et al., *Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming.* Nat Methods, 2017. **14**(11): p. 1055-1062.

48. Liu, H., et al., *TALEN-mediated gene mutagenesis in rhesus and cynomolgus monkeys.* Cell Stem Cell, 2014. **14**(3): p. 323-328.

49. Cunningham, F., et al., *Ensembl 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D745-D751.

50. Young, M.D. and S. Behjati, *SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data.* bioRxiv, 2020: p. 303727.

51. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nat Biotechnol, 2018. **36**(5): p. 411-420.

52. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.

53. Hafemeister, C. and R. Satija, *Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.* Genome Biol, 2019. **20**(1): p. 296.

54. Pont, F., M. Tosolini, and J.J. Fournié, *Single-Cell Signature Explorer for comprehensive visualization of single cell signatures across scRNA-seq datasets.* Nucleic Acids Research, 2019. **47**(21): p. e133-e133.

55. Smedley, D., et al., *The BioMart community portal: an innovative alternative to large, centralized data repositories.* Nucleic Acids Res, 2015. **43**(W1): p. W589-98.

56. Cao, J., et al., *The single-cell transcriptional landscape of mammalian organogenesis.* Nature, 2019. **566**(7745): p. 496-502.

57. Qiu, X., et al., *Reversed graph embedding resolves complex single-cell trajectories.* Nat Methods, 2017. **14**(10): p. 979-982.

58. Trapnell, C., et al., *The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.* Nat Biotechnol, 2014. **32**(4): p. 381-386.

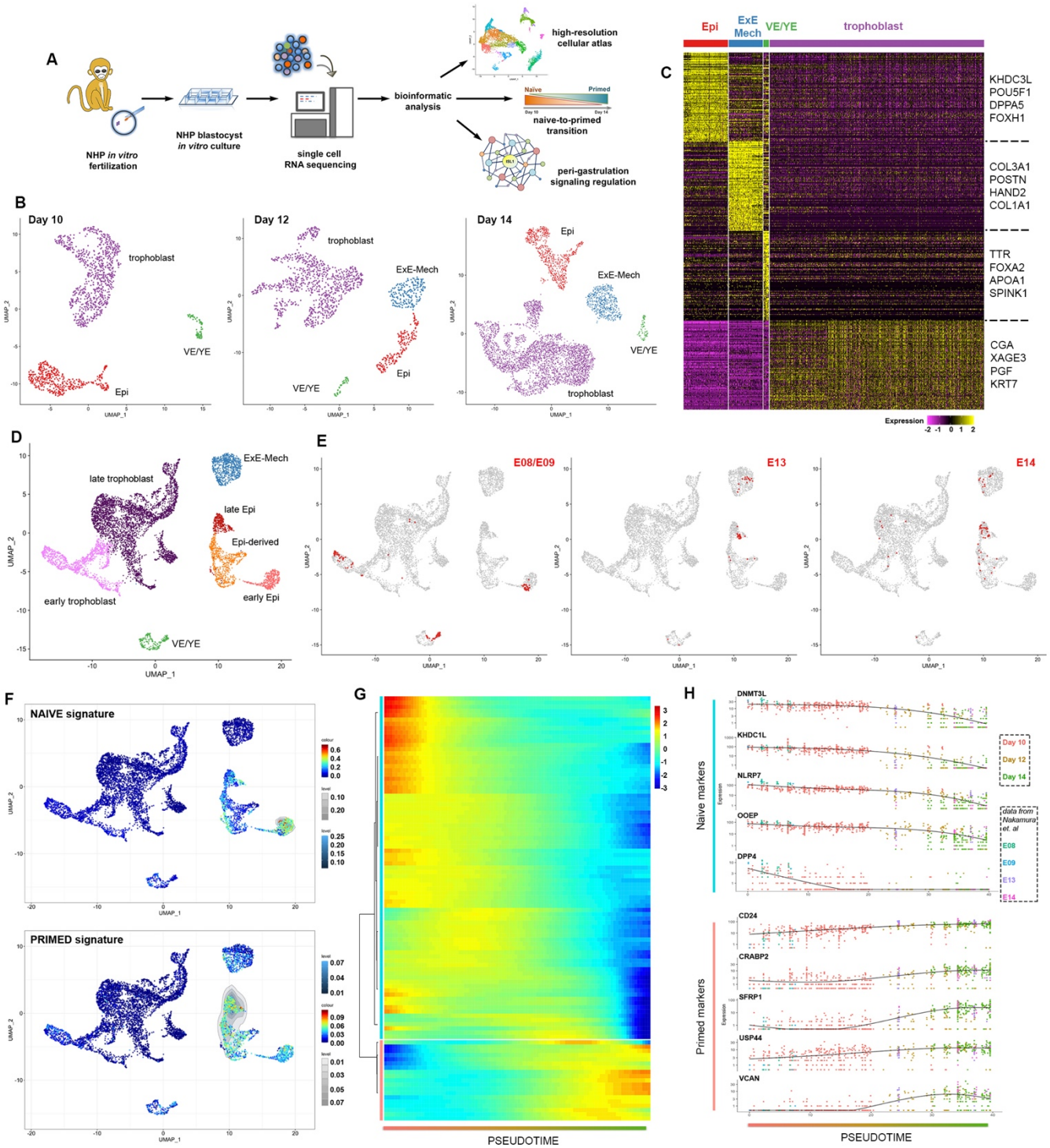

Fig. 1 High resolution transcriptomic map of peri-gastrulation events in wildtype *in vitro* cultured NHP embryos. A, scheme of the workflow. **B**, UMAP plot of all cells from the *in vitro* cultured embryos at the different time points (Day 10, Day 12 and Day 14) coloured by cell type. **C**, Heatmap showing the scaled expression at Day 14 of 100 differentially expressed genes (DEGs) for each cell type identified in (b) selected by adjusted p-value. **D**, UMAP plot of all cells of the integrated dataset from the *in vitro* cultured embryos (Day 10, Day 12 and Day 14) with cells from the *in vivo* NHP embryos (E08, E09, E13 and E14) colored by cell type. **E**, position of the cells from the *in vivo* cultured embryos (E08/E09, E13 and E14) in the UMAP plot of the integrated dataset highlighted in red. **F**, visualization of the single cell signature score reflecting the naïve and the primed state of pluripotency on the UMAP plot of the integrated dataset. **G**, heatmap showing the scaled expression of 100 DEGs across pseudotime in epiblast development selected by the Moran's I statistics. Unsupervised clustering of genes into a blue cluster containing genes downregulated over pseudotime and a red cluster containing genes upregulated over pseudotime. **H**, gene expression of 5 exemplary genes downregulated over pseudotime (blue cluster) and 5 exemplary genes upregulated over pseudotime (red cluster). Black line showing the fitted expression trend for each gene over pseudotime. Epi, epiblast; ExE-Mech, extraembryonic mesenchyme, VE/YE, visceral endoderm/yolk sac endoderm.

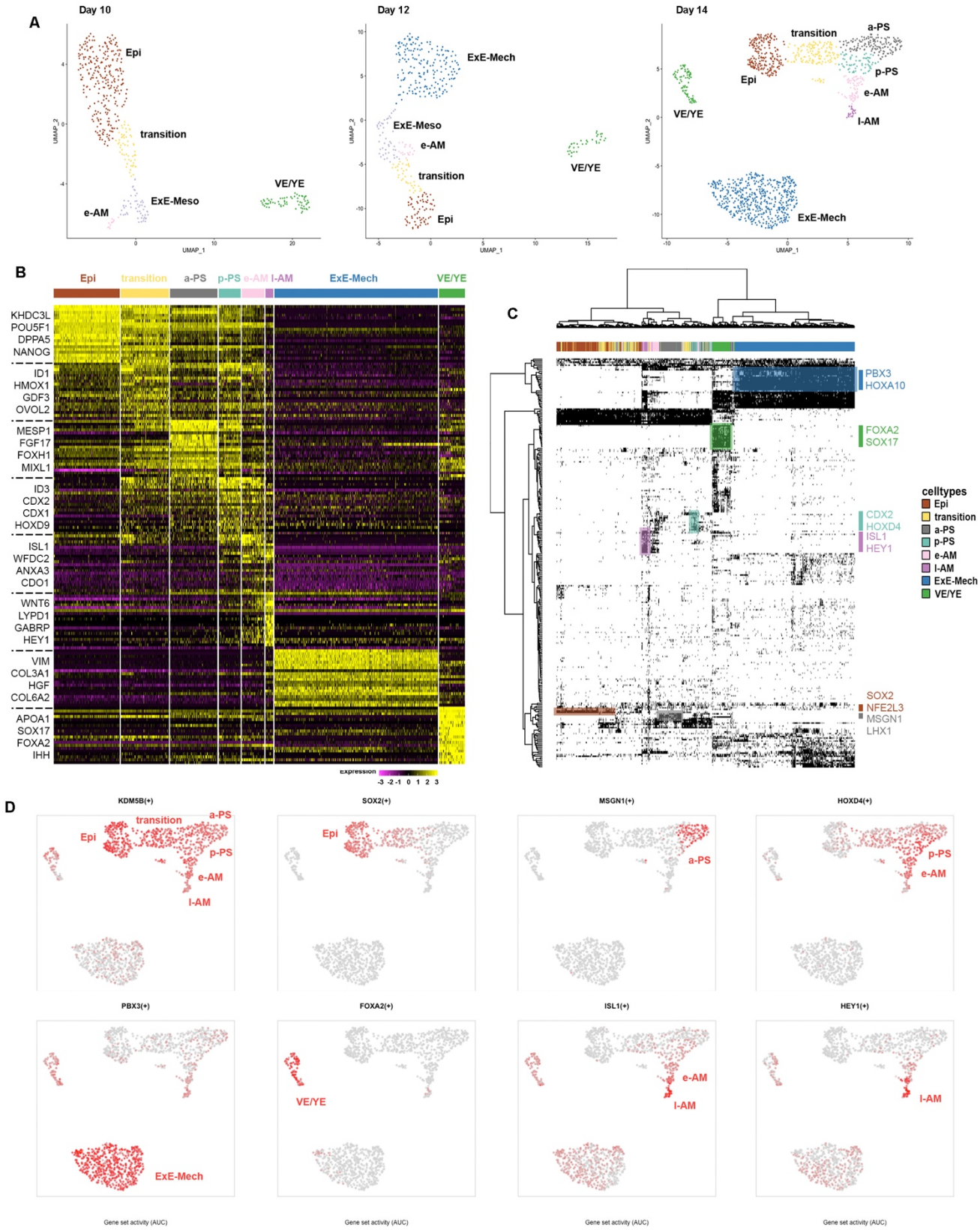

Fig. 2 Gene regulatory networks in epiblast and its derivatives, endoderm and extraembryonic mesenchyme. A, UMAP plot of cells from the *in vitro* cultured embryos at the different time points (Day 10, Day 12 and Day 14) mapping to the Epi and its derivatives, VE/YE and ExE-Mech coloured by cell type. **B**, Heatmap showing the scaled expression at Day 14 of 20 DEGs for each cell type identified in (A) selected by adjusted p-value. **C**, binary activity matrix of regulons identified at Day 14 by gene regulatory network interference active in at least 1% of the cells clustered unsupervised. Selected master regulators are depicted in the colour corresponding to the cell type they show activity in. **D**, gene set activity of the selected regulons at Day 14 in the different cell types shown as "Area Under the Curve" (AUC) depicted on the UMAP plot from (A). Epi, epiblast; VE/YE, visceral endoderm/yolk sac endoderm; ExE-Meso, extraembryonic mesoderm; ExE-Mech, extraembryonic mesenchyme; e-AM, early amnion; l-AM, late amnion; ab-AM, abnormal amnion; a-PS, anterior primitive streak; p-PS, posterior primitive streak.

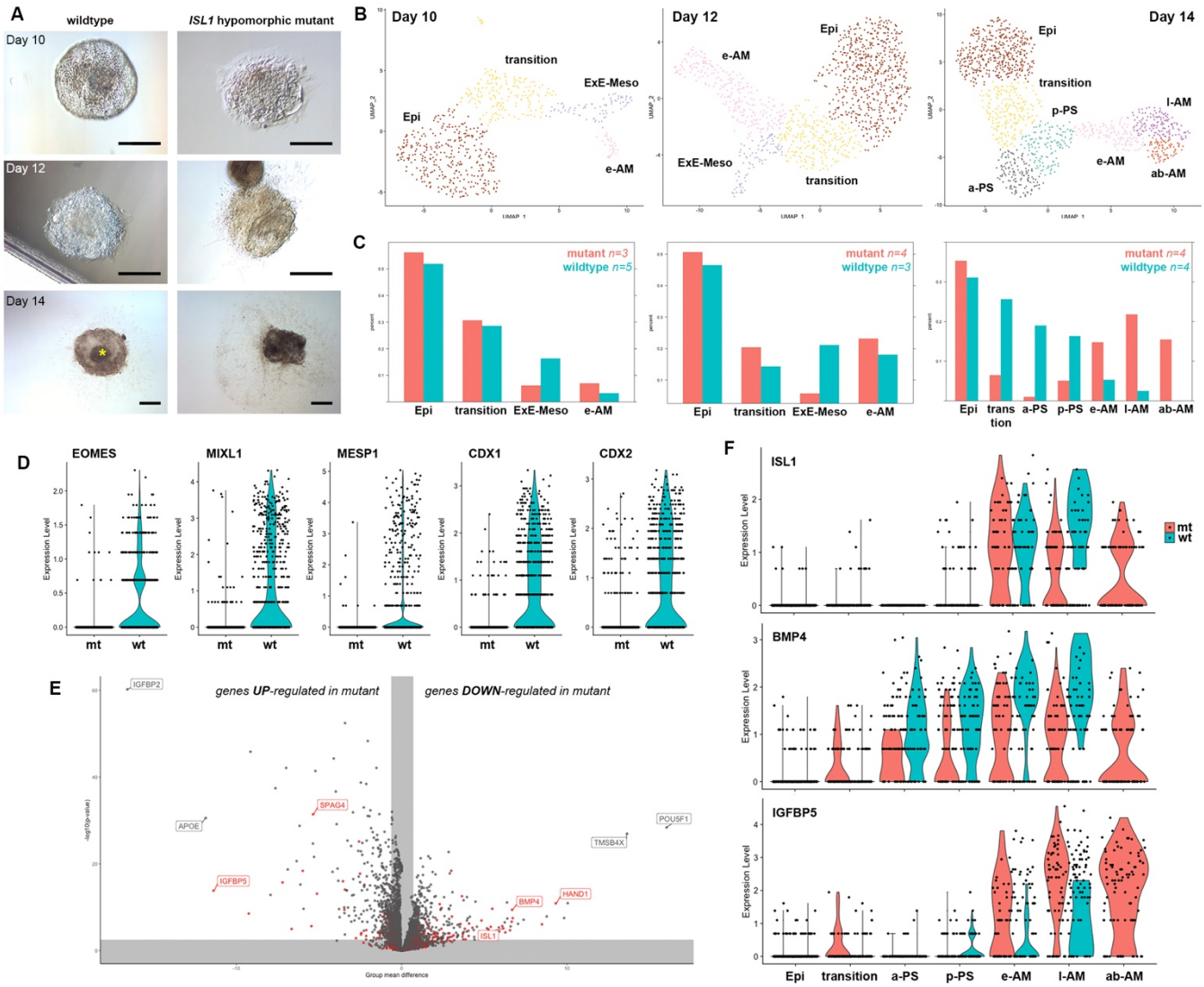

Fig. 3 *ISL1* hypomorphic mutants fail to form primitive streak. A, the morphology of wildtype (left) and *ISL1* hypomorphic mutants (right) at Day 10, Day 12 and Day 14. Scale bar, 200 µm. Asterisk indicates the embryonic disk. **B**, UMAP plot of cells mapping to the epiblast and its derivatives from the integrated dataset of wildtype and mutant embryos at the different time points (Day 10, Day 12 and Day 14) coloured by cell type. **C**, relative contribution of cells to the various cell types in wildtype (blue) and mutant (red) embryos. **D**, violin plot of the expression of genes involved in primitive streak formation in cells from the epiblast and its derivatives of Day 14 wildtype (blue) and mutant (red) embryos. **E**, volcano plot showing the DEGs between amnion (e-AM, l-AM and ab-AM) of the wildtype and mutant embryos. Grey areas show expected group mean difference from a random cell subset and a false discovery rate (FDR) of 1%. Red labelling indicates that the gene is part of the ISL1 regulon identified by the gene regulatory network interference. **F**, violin plot of the expression of three selected DEGs from (E) in the different cell types identified in (B) separated between wildtype (blue) and mutant (red) embryos. Epi, epiblast; VE/YE, visceral endoderm/yolk sac endoderm; ExE-Meso, extraembryonic mesoderm; ExE-Mech, extraembryonic mesenchyme; e-AM, early amnion; l-AM, late amnion; ab-AM, abnormal amnion; a-PS, anterior primitive streak; p-PS, posterior primitive streak; wt, wildtype; mt, mutant.

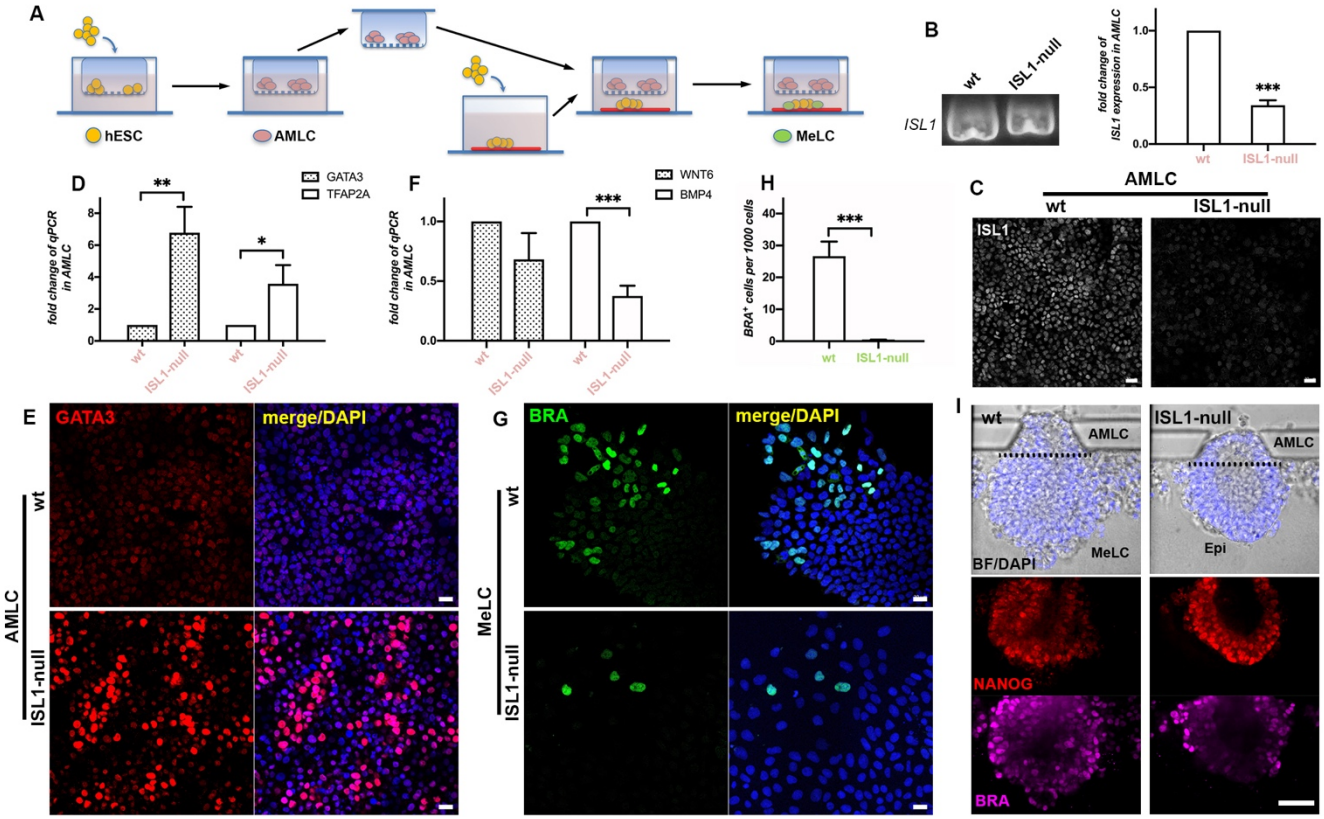

Fig. 4 *ISL1* regulates human mesodermal cell formation through BMP4 pathway. A, a diagram of the transwell assay. **B**, *ISL1* mRNA expression in AMLCs. Analysed by student's t-test, p-value <0.0001. Wildtype n=4; *ISL1*-null are aggregated values from two *ISL1* knockout cell lines n=4 for each. **C**, ISL1 protein in AMLCs. Scale bar, 50 µm. **D**, the expression of the amnion markers *GATA3* (p-value 0.0038) and *TFAP2A* (p-value 0.0403) in AMLCs. Analysed by student's t-test. Wildtype n=4; *ISL1*-null are aggregated values from two knockout lines n=4 for each. **E**, GATA3 (red) protein in AMLCs. Nucleus shown by DAPI (blue). Scale bar, 50 µm, n=3. **F**, the expression of *WNT6* (p-value 0.0752) and *BMP4* (p-value <0.0001) in AMLCs. Analysed by student's t-test. Wildtype n=4; *ISL1*-null are aggregated values from two knockout lines n=4 for each. **G**, Mesoderm marker Brachyury (BRA, green) protein in MeLCs. Nucleus shown by DAPI (blue). Scale bar, 50 µm. **H**, Quantification of BRA+ cells from (G). Analysed by student's t-test, p-value <0.0001. Wildtype n=3; *ISL1*-null are aggregated values from two knockout cell lines n=3 for each. **I**, the morphology (top panel) of embryonic-like sacs overlayed with nucleus shown by DAPI (blue). NANOG (red) protein and BRA (magenta) protein in lower panels. Scale bar 50 µm, n = 30 for each. Significance levels are depicted by asterisks: * p-value <0.05, ** p-value <0.01, and *** p-value <0.001.

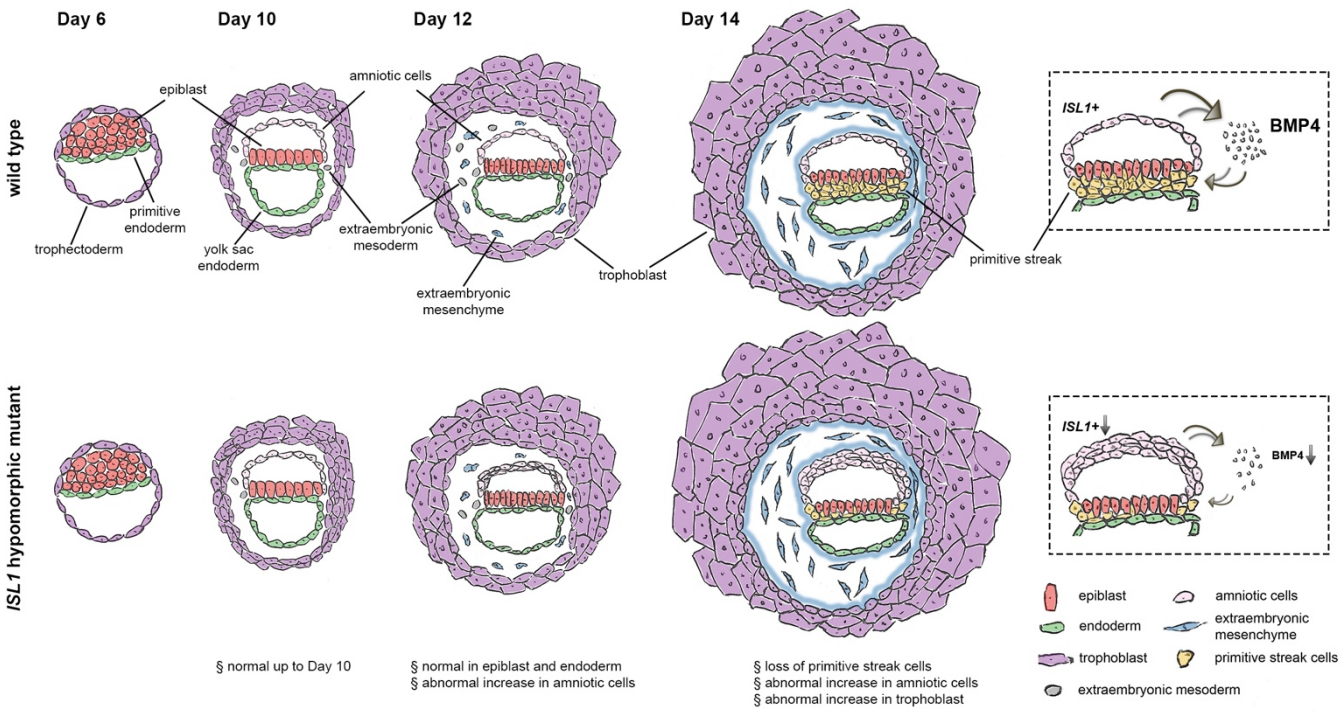

Fig. 5 A summary scheme depicting the embryogenesis of wildtype and *ISL1* hypomorphic mutant embryos.

Table 1. Summary of embryo transfer

	Oocytes	Zygotes	Blastocysts for ET	Surrogate mothers	Pregnancies	Fetuses
wildtype	48	45	35	12	7	10
single gRNA injection	48	45	24	7	2	4*
double gRNAs injection	53	50	25	8	0	0

* genotyping revealed two intact *ISL1* alleles in all fetuses