Reconstituting epiblast–extraembryonic endoderm interactions restores anterior–ventral patterning in stem cell–based embryo models

Natalia P. Smirnova[1,2], Sergey V. Ponomartsev[1,2*], Tharvesh M. Liyakat Ali[4], Max Lycke[2], Brian K. Chung[5], Jonas Øgaard[5], Espen Melum[5, 2], Jesse Veenvliet[3], Stefan Krauss[1,2]*

[1]Department of Immunology and Transfusion Medicine, Oslo University Hospital, P.O. Box 4950, 0424 Oslo, Norway.

[2]Hybrid Technology Hub Centre of Excellence, Institute of Basic Medical Science, University of Oslo, P.O. Box 1110, 0317 Oslo, Norway.

[3]Stembryogenesis Lab, Max Planck Institute of Molecular Cell Biology and Genetics, 01307, Dresden, Germany.

[4]Liyakat Ali Bioinformatics, Løytnant Møllers veg 21A, 2080 Eidsvoll, Norway

[5]The Experimental Liver Research Group, Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway.

* Authors contributed equally to this work.

Corresponding author: Stefan Krauss, s.j.k.krauss@medisin.uio.no

Key words: extraembryonic endoderm, mouse stem cell-based embryo models, mSCBEM, gastruloids, aggregoids, primitive streak, notochord, cardiopharyngeal mesoderm

Abstract

During mouse embryogenesis, the interactions between the epiblast and extraembryonic endoderm are critical for germ layer specification and body plan development. Gastruloids recapitulate aspects of gastrulation in the absence of morphogenic signals from the extraembryonic environment. This favors a predominantly posteriorized and dorsalized phenotype with a limited representation of embryonic lineages. Here, we develop and employ a co-aggregation (aggregoid) approach combining embryonic and extraembryonic endoderm-like cells to mimic spatial interactions in developing mouse embryos. The obtained embryo models show the appearance of a regionalized primitive streak-like structure that results in the formation of node and notochord, enriched endoderm populations, increased mesoderm diversity including cardiopharyngeal lineages and vascular endothelium in an overall anteriorized and ventralized phenotype. The work delineates a versatile strategy for refining stem cell-based embryo models to achieve specific morphotypes.

Introduction

Mouse 3D gastruloids represent a sub-group of stem cell-based embryo models (mSCBEM) able to recapitulate key features of peri-gastrulation embryonic development: the establishment of three body axes (anterior-posterior, dorsal-ventral and left-right), the emergence of three germ layers (ectoderm, mesoderm and endoderm) and the initiation of early organogenesis[1-3]. In gastruloids, the complex cascade of symmetry breaking and morphogenic events is induced by WNT up-regulation due to exposure of mouse embryonic stem cells (mESC) aggregate to CHIR99021 (CHIR). However, *in vivo*, mammalian gastrulation is orchestrated by precise spatial and temporal coordination of several major signaling pathways (WNT, BMP, NODAL, FGF, HH), tightly regulated by the extraembryonic environment[4-6] .Hence, despite the spectacular ability for reproducible self-organization in absence of pre-patterned signaling gradients and instructive extraembryonic cues, gastruloids

exhibit significant developmental constraints, including a post-occipital and overall posteriorized identity, a lack of anterior neural structures, and a limited lineage potential towards endodermal and anterior mesodermal fates[3,7].These effects can be explained by the total exposure of the 3D cell aggregate to CHIR that results in the prevalence of a WNT3a/BRA late primitive streak program driving formation of posterior structures and suppression of the early primitive streak NODAL/EOMES program giving raise to mesendoderm and more anterior structures[8,9].

The cell types corresponding to extraembryonic endoderm (ExEnd) such as Primitive Endoderm (PrE), Visceral Endoderm (VE), and Parietal Endoderm (PE) play a multifaceted and dynamic role in early mouse embryogenesis supporting epiblast maturation, axis formation while providing structural support, transport of nutrients[10]. A subpopulation of the VE, termed Anterior Visceral Endoderm (AVE) is a critical signaling center that spatially restricts primitive streak formation by expressing Wnt antagonists like Dkk1, and TGFβ/Nodal antagonists such as Lefty1 and Cerberus 1(Cer1) and thus defines the anterior region of the embryo[11]. The posterior visceral endoderm (PVE) collaborates with BMP4-expressing extraembryonic ectoderm and the adjacent epiblast to induce WNT3, which initiates primitive streak formation and gastrulation[12,13].

To explore how extraembryonic endoderm (ExEnd) could improve development of mSCBEMs, while providing a platform to study the regulation of structure, patterning and lineage formation by ExEnd in the gastrulating embryo, we developed a co-aggregation platform (aggregoids) whereby we combine pre-differentiated ExEnd-like cell types (PrE-like, VE-like and AVE-like) with naïve mESCs. We found that in addition to providing cellular diversity and structure, the presence of ExEnd cells favorably modulates the balance of NODAL and WNT signaling with significant implications on the gastrulating embryo model. Taken together, the here developed aggregoids show hallmarks of a primitive streak, a notochord-like structure and an overall enriched lineage specification, including endodermal lineages, axial mesoderm and cardiopharyngeal mesoderm in an overall ventralized and anteriorized morphotype when compared to “classical” gastruloids. Furthermore, the differential choice of including PrE-like, VE-like or AVE-like cells in the aggregates allows fine tuning of lineage deployment. The robustness and versatility of the aggregoid strategy, in particular the ability to finetune the state, timing, amount and surrounding of the aggregated cells makes it a platform for engineering the composition and potential of mSCBEMs.

Results

Derivation of extraembryonic endoderm-like cells.

To generate mouse stem cell-derived embryo models containing extraembryonic endoderm (ExEnd) compartment, we employed and adapted protocols for derivation cell types resembling pre- and post-implantation ExEnd cell types from mouse embryonic stem cells (mESC) without induced transgene overexpression. Previous works[14,15] has shown that culturing naïve mESC in RACL medium (RPMI supplemented with B27 without insulin, 100 ng/ml Activin A, 3 μM CHIR99021, and Lif) induces their differentiation into a cell type resembling primitive endoderm (PrE). We first maintained mESCs for 5-7 days in 2i/N2B27 (2i) and then converted them for 6-8 days in RACL medium. Additionally, we tested two subsequent modifications of the RACL protocol to derive later ExEnd-like types (Fig. 1a, Supplementary 3c). To obtain VE-like cells, NACLB medium (N2B27, 100 ng/ml Activin A, 3 μM CHIR99021, Lif, and 50ng/ml BMP4) was used for 3 days following RACL conversion[14]. To obtain AVE-like cells, based on findings that WNT inhibition facilitates AVE induction[16], we withdrew CHIR99021 from the media and replaced it with the WNT pathway inhibitor XAV939 creating XAL medium (N2B27, 10uM XAV939, 100 ng/ml Activin A, Lif).

According to bulk RNA sequencing, 2i cells demonstrated the signature of naïve mESC including dome-shaped morphology of colonies and expression of *Zfp42, Klf2, Klf5, Dppa3,*

Nanog, Alpl, Tcl1, Nkx1-2, Esrrb, Fzd10, Prdm14 and low expression level of primed markers such as *Fgf5, Pou3f* and *Zic2* (Fig.1b, c, Supplementary 3d). Upon RACL conversion, cells adopted an epithelial phenotype and an expression profile typical for extraembryonic endoderm appeared including an up-regulation of *Sox17, Gata6, Gata4, Hnf4a, Pdgfra, Igf1, Dab2* (Fig, 1b Supplementary 1a, Supplementary 3 a, Supplementary, 4 a-g) and the expression of the PrE markers *Igf1, Itgb2, Sox7* and *Dusp4* was registered (Fig. 1d; Supplementary 3a)The expression profile of RACL cells was enriched with transcripts of *Zscan4, Pramel* and *Gm* families (Fig. 1 c) associated with pluripotency, the germline and early embryogenesis[17].

Differentiation towards VE-like and AVE-like cells was accompanied with morphological changes: the appearance of lipid droplets in the cytoplasm soon after start of NACLB treatment and characteristic shape changes in XAL cells including columnar shaped cells and cells with a migratory phenotype (Fig. 1a, Supplementary 3e, 4g). All three ExEnd-like cell types (RACL, NACLB and XAL) shared a plethora of pan-ExEnd markers such as *Hnf4a, Foxa2, Hnf1b, Ttr, Dab2, Cubn, Lama1 Ctsl* and, *Srgn* (Fig. 1c, Supplementary 1d, 2a, b, 3a), confirming their general resemblance to extraembryonic endoderm. All ExEnd-like cell types further showed increased expression of *Wnt11*, *Bmp4, Ihh, Vegfa* (Supplementary 1d, 3a) - as well as laminins (*Lama1, Lamb1*). However, we did not observe any sufficient down-regulation of markers such as *Sox17* and *Sox7* that occurs upon VE derivation from PrE *in vivo*[18,19] (Supplementary 3a), indicating limitations of the protocol. Despite a high similarity in gene expression between NACLB and XAL (Fig.2e, f), immunohistochemistry revealed that XAL cells expressed markers associated with the anterior visceral endoderm (AVE), including OTX2, LEFTY1, and EOMES (Supplementary 1b, e; 4d, f). Bulk RNA sequencing and qPCR showed the upregulation of specific AVE markers such as *Hhex, Sfrp1, Lhx1* and inhibitory molecules such as *Lefty1, Dkk1 Cer1* (Fig 1b, c, e, Supplementary 1b, d, f). Among XAL-specific transcripts were also genes associated with cell migration: *Sema6a, Sema6d, Plat, Enpp2, Nid1, Sox11* (Fig. 1c), which are also specific for AVE[20].

The expression of *Dkk1* was detected in all three cell types, but in RACL its level was significantly higher than in XAL. *Cer1* and *Lefty1* were only expressed at relevant levels in XAL and RACL cells (Supplementary 1b). Overall, the ExEnd-like cell types demonstrated a diverse repertoire of signaling molecules including inhibitors of TGF-b and WNT signaling, known for their role in axis establishment and lineage regulation[21,22].

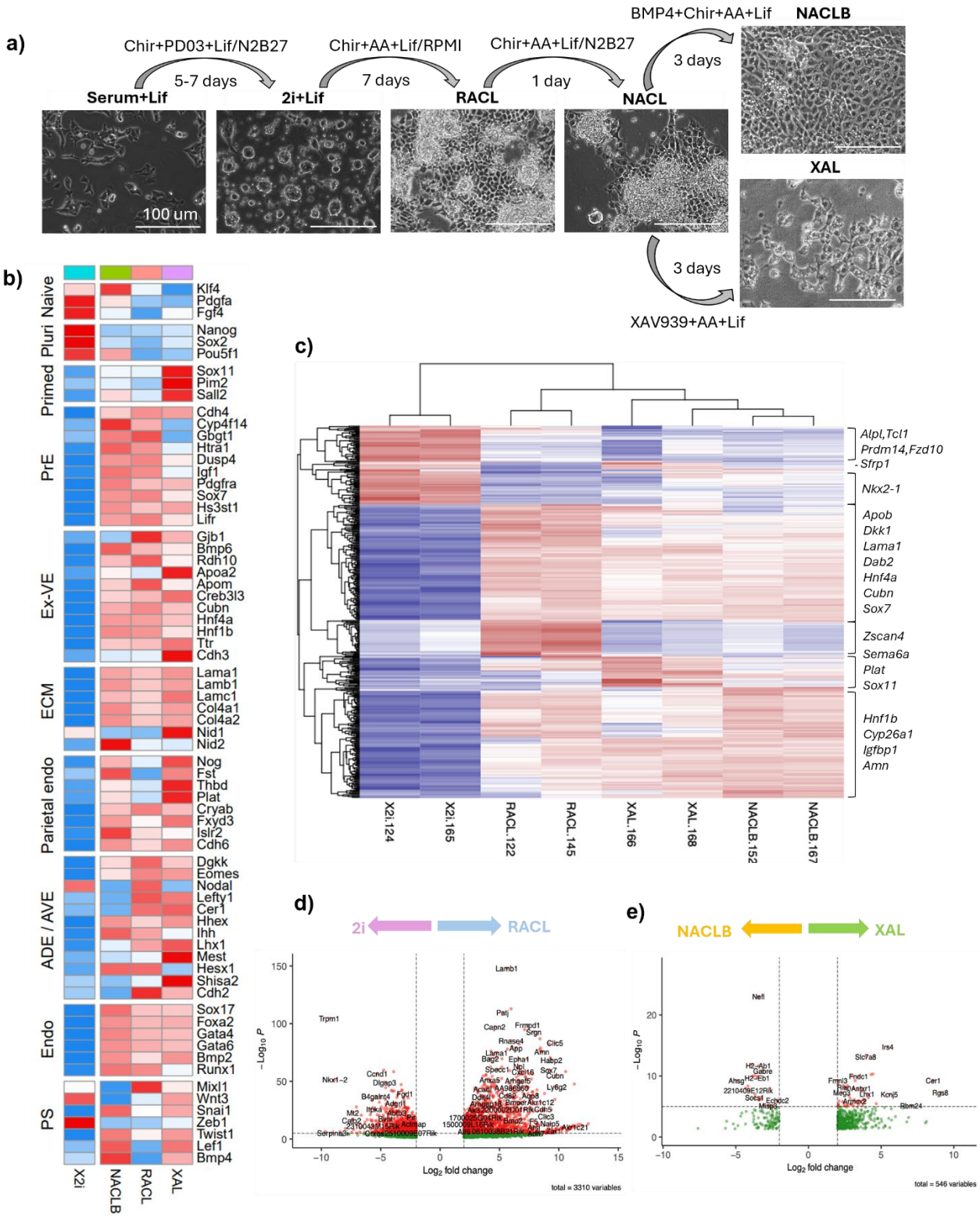


Figure 1. Derivation of extraembryonic endoderm-like cells. a) Figure 1. Derivation of extraembryonic endoderm-like cells. a) *Schematic representation of steps and timing of RACL, NACLB and XAL conversion protocols with representative phase contrast images of mESC, 2i and ExEnd-like cell types;* ***b)*** *Heatmap with averaged expression of selected markers based on bulk RNA-seq (n=3);* ***c)*** *Correlation based hierarchical clustering of top 500 genes expressed in 2i, RACL, NACLB and XAL cells based on bulk RNA sequencing;* ***d, e)*** *Volcano plots showing the differences between pairwise 2i and RACL or NACL and XAL cells based on bulk RNA sequencing.*

Co-aggregation of ExEnd-like cells with 2i cells drives the early aggregoid regionalization.

To investigate the influence of ExEnd-like cells on early patterning of mouse gastruloids, we co-aggregated 2i cells with either of the three pre-differentiated ExEnd-like cell pools in 96-well U-bottom plates (Figure 2a). A pulse of CHIR (3 µM) was given between 48 and 72h, as in the canonical gastruloid protocol[1,3]. The resulting structures are termed RACL-, NACLB-, and XAL-aggregoid, depending on the specific cell type that was used for the co-aggregation, and collectively referred to as ExEnd-aggregoids. Structures formed from 2i cells alone are referred to as control gastruloids. Co-aggregation of 300 2i cells with 100 RACL cells resulted in 87% of "successful" structures (72% and 92% for NACLB and XAL, respectively), characterized by partial coverage with an ExEnd-like cell cap (Figure 2d, e, Supplementary 5e) and further ability to undergo elongation and axis establishment. (Supplementary 5a-c). The "unsuccessful" structures, fully incapsulated into extraembryonic cells failed to break symmetry and develop (Supplementary 5e).

We employed the BRA-GFP;SOX17-RFP[23] double reporter line for the 2i-derived part to track the dynamics of symmetry breaking in control and ExEnd-gastruloids using time-lapse microscopy. We also performed single-cell RNA-sequencing (scRNA-seq) of control, RACL, NACLB and XAL gastruloids at time-points 48h, 72h, 96h and 120h. We used automated annotation of cell types by transferring labels from the mouse gastruloid reference atlas[24] (Methods; Supplementary 8 a-d).

Before the CHIR pulse, at 48 hours post-aggregation, BRA-GFP expression was undetectable in both ExEnd aggregoids and control gastruloids (Figure 2e, Supplementary 6a). At 48h, scRNA-seq similarly showed minimal or no expression of *Brachyury* (*Bra*) at this stage (Supplementary 6 b, c). Furthermore, no significant differences between the conditions in the proportion of cells expressing *Nodal*, *Eomes* of *Fgf5* was revealed (Supplementary 6b, c). The scRNA-seq showed that at 48h control gastruloids and ExEnd-aggregoids were predominantly composed of *Otx2, Fgf5, Utf1, Rasgrp2,* and *Dppa5a*-positive cells, corresponding to a pre-gastrulation epiblast (Figure 2b, c; Supplementary 6 a-c), and a cluster automatically annotated as "PGC" containing a mix of cells spanning various pluripotency states. As expected, RACL-, NACLB- and XAL-aggregoids also contained clusters corresponding to ExEnd cell types which were absent in control gastruloids (Figure 2b, c). Taken together, these observations suggest that detectable variations in 2i-derived cell compositions between control gastruloids and ExEnd-aggregoids emerge after 48 h, with no initial condition-specific heterogeneity detectable.

By 72h, the BRA-GFP expression in control gastruloids appeared in a radially symmetrical pattern, emerging from the periphery and spreading toward the center (Figure 2d, 3a). In contrast, ExEnd-aggregoids showed early polarized BRA expression, emerging on average from 67 hours at a single pole (Figure 2d, e, 3a). Time-lapse imaging with the use of lines containing ROSA26-BFP/Flk1-GFP[25] or SOX17-RFP reporters for ExEnd revealed that: i) BRA expression consistently emerged at the pole opposite to the "cap" of ExEnd-like cells (Figure 2d); ii) The region in proximity of the ExEnd-like cap remained BRA-negative; iii) The initial BRA-positive pole corresponded to the future posterior end of the structure, while the ExEnd-covered side marked the anterior (Movie 1). Immunocytochemistry and scRNA-seq demonstrated that at 72h ExEnd-like cell types still expressed WNT inhibitor DKK1 (Figure 2d, supplementary 6d). Altogether, these data show that complementation of canonical gastruloids with ExEnd-like cells results in early symmetry breaking and regionalization, likely resulting from differential responses to WNT signaling, modulated by the ExEnd-like cells.

At 72h timepoint, scRNA-seq revealed that in control gastruloids, only 54% of cells expressed *Bra*, compared to 89% in RACL-aggregoids, 91% in NACLB-aggregoids and 84% in XAL-aggregoids (Figure 2f). Considering the transcription-to-translation delay, we propose that it might reflect *de novo* polarization of the initially uniform BRA expression in control gastruloids and appearance of an additional expression domain at the anterior pole for ExEnd-aggregoids that can be registered between 76–84h on the level of BRA-GFP reporter (Figure 3a). Also, at 72 h, 17% of cells in RACL and NACLB-aggregoids and 5% in XAL co-expressed *Bra* and *Eomes* (Figure 2f), a transcriptional signature of early primitive streak[26]. In contrast, a comparable population comprised only 0.5% in control gastruloids (Figure 2f). These findings indicate that in ExEnd-aggregoids, RACL, NACLB, and XAL cells do not completely inhibit the gastrulation-like process but rather delay its progression, allowing both early and late primitive streak (PS) programs to operate within the same structure.

The scRNA-seq data further revealed caudal epiblast (CE) as the most prevalent cell type at 72h in both control and RACL-, NACLB, and XAL-aggregoids (Figure 2g, j). However, in contrast to control gastruloids, ExEnd-aggregoids contained cells annotated as primitive streak (PS), anterior primitive streak (APS) and nascent mesoderm (NM). Moreover, both control gastruloids and XAL-aggregoids featured populations annotated as ectoderm and neural tube (NT), suggesting an early neural bias under these conditions. Importantly, these identities were less represented in RACL- and NACLB aggregoids. Taken together, our data demonstrate that co-aggregation with DKK1-expressing ExEnd-like cells induces early regionalization of the aggregoid into domains that differently respond to the CHIR pulse and transcriptionally correspond to early and late PS.

The activation of a primitive streak (PS) and anterior primitive streak (APS) like transcriptional program in RACL-, NACLB- and XAL-aggregoids, culminating in the co-emergence of posterior and anterior cell types, is further supported by the up-regulation of TGF-β–associated genes such as *Nodal, Eomes, Cer1, Lefty1, Lefty2, Nkx2-1, Tdgf1,* and *Prrx2* at 72h (Supplementary 7b, c).

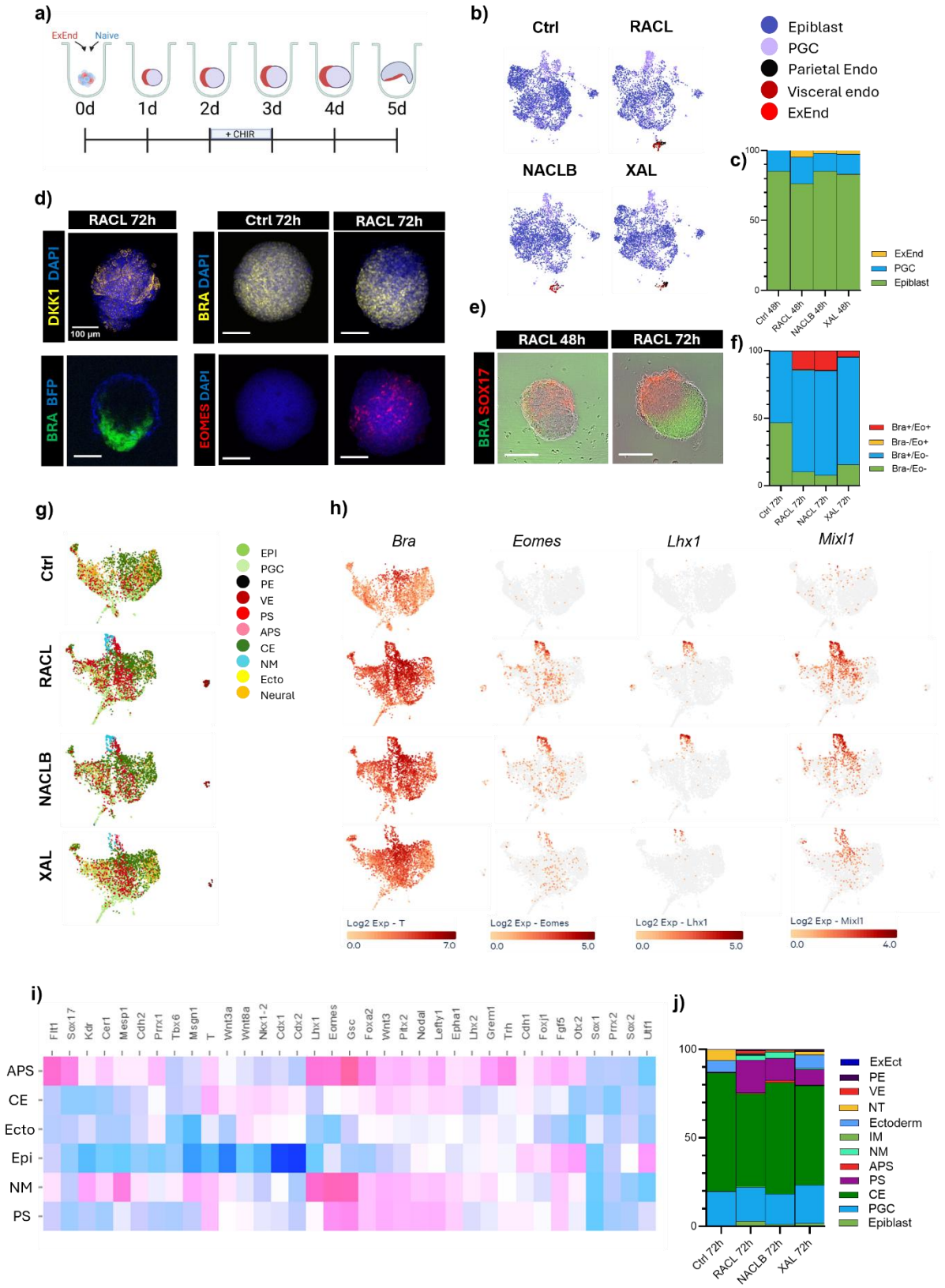


Figure 2. Co-aggregation of ExEnd-like cells with 2i cells drives the early aggregoid regionalization a) *Experimental outline of ExEnd co-aggregation with 2i/N2B27 mESC in gastruloid protocol.* ***b)****UMAP representing predicted cell clusters in 48h control gastruloid and RACL, NACLB and XAL aggregoids, scRNA-seq.* ***c)*** *Bar plot showing the proportions of predicted cell types in control, RACL, NACLB, and XAL-aggregoids at 48h post-aggregation.*

d) *Immunostaining of BRA, DKK1, and EOMES expression in 72 h control gastruloids and RACL-aggregoids 72h control gastruloid, immunostaining; scale bar = 100 µm.* ***e)*** *Phase contrast images with BRA-GFP and SOX17-RFP reporters demonstrating typical BRA expression in RACL-aggregoids before and after CHIR pulse; at this stage, SOX17 marks added RACL cells. Scale bar = 100 µm* ***f)*** *Bar plots showing the ratio of Bra-positive, Bra-negative and Bra/Eomes double-positive cells in 72h control gastruloid and RACL-, NACLB, XAL-aggregoids;* ***g)*** *scRNA-seq based UMAP comparing cell populations in 72h control gastruloids and RACL, NACLB and XAL aggregoids;* ***h****) scRNA-seq based UMAP of T/Bra, Eomes, Lhx1 and Mixl1 expression in control gastruloids and RACL-, NACLB- and XAL-aggregoids at 72 h; e* ***i)*** *Heatmap showing differential expression of selected marker genes in the predicted clusters at 72h.* ***j)*** *Bar plot showing the proportions of predicted cell types in control, RACL, NACLB, and XAL-aggregoids at 72h post-aggregation based on scRNA-seq.*

Co-aggregation with ExEnd-like cells results in axial mesoderm formation.

From 72 to 96 hours, control gastruloids displayed progressive polarization of the BRA-GFP domain, with expression becoming restricted to the posterior half and subsequently localized to the posterior tip as elongation proceeded (Figure 3a). In contrast, in ExEnd-aggregoids, the BRA-GFP domain began to narrow after 84h, and by 96h, got a shape of longitudinal streak in the midline of the aggregoid (Figure 3a). Concomitantly, 77% of RACL and 89% of NACLB aggregoids had anteriorly positioned BRA-GFP+ domains not typically observed in canonical gastruloids (Supplementary 10 a,b).

At 72h, high expression of BRA was registered in the CE cluster, characterized by expression of *Bra* alongside *Wnt3a*, *Cdx1*, *Cdx2*, and *Nkx1-2* (Figure 2i, Supplementary 9a). This cluster were represented in both control and ExEnd-gastruloids. Additionally, in RACL-, NACLB-, and XAL-gastruloids, *Bra* was expressed within same clusters as *Eomes:* in a cluster of cells with a signature of the early primitive streak (PS) and its derivatives: anterior primitive streak (APS) (*Otx2*-, *Gsc*-, *Flt1*-, *Sox17*-, *Foxa2*-, *Tdgf1*-positive) (Supplementary 7a, 9a) and nascent mesoderm (NM (*Mesp1* and *Msgn1*-positive)) (Figure 2h, i, Supplementary 7a, 9a).

Next, we performed live imaging and immunostaining of fixed samples to spatially locate the identified cell populations. At 96h BRA expression was condensed along the midline and absent from the lateral regions (Figure 3 a-d). The EOMES-positive region consistently corresponded the anterior area (likely corresponding NM and APS) covered by DAB2-marked extraembryonic endoderm cells (Supplementary 10c) and opposed posterior CDX2-positive region probably corresponding CE or neuromesodermal progenitors (NMPs) that appeared at this time-point (Figure 3e, Supplementary 9b).

At 96 hours, some RACL and NACLB aggregoids displayed a distinct domain of BRA-, CDH1-, and FOXJ1 (marker of node), surrounded anteriorly by a broad SOX17-positive region (resembling endoderm) and posteriorly by mesodermal FOXF1 and BRA-expressing cells (Figure 3f-h). At 96 hours, a *Bra*, *Foxj1*, *Sox9*, *Prrx2*, and *Lefty1*-positive cluster transcriptionally resembling the node (Figure 3i) was represented in RACL- and NACLB-aggregoids. Based on the morphology, developmental timing, and concordance between scRNA-seq and immunostaining data, we propose the resemblance of this domain to a node-like structure. Notably, BRA- and FOXJ1-positive node-like structures were observed in the anterior or medial regions at 96h and closer to the posterior tip at 120 h, possibly recapitulating the cranial-to-caudal sequence of node regression[27] (Figure 3f-h).

At 96 hours, a *Bra*, *Foxj1*, *Sox9*, *Prrx2*, and *Lefty1*-positive cluster transcriptionally resembling the node was represented in RACL- and NACLB-aggregoids. Based on the morphology, developmental timing, and concordance between scRNA-seq and immunostaining data, we propose the resemblance of this domain to a node-like structure. Notably, BRA- and FOXJ1-positive node-like structures were observed in the anterior or medial regions at 96h and closer to the posterior tip at 120 h, likely recapitulating the cranial-to-caudal sequence of node regression[27] (Figure 3f).

The node is a source and instructive niche for axial mesoderm (notochord and prechordal plate).[28]. Between 96 and 120 hours, a midline domain co-expressing BRA and FOXA2 emerged in RACL-, NACLB-, and XAL- aggregoids (Figure 3k). The scRNA-seq data for the corresponding time-points revealed a population expressing both *Bra* and *Foxa2* alongside *Shh, Noto, Chrd* that was annotated as notochord (Figure 3i) and was observed in all three experimental conditions, but not in control gastruloids (Supplementary 8a-d). Confocal microscopy and time-lapse imaging revealed that the axial mesodermal domain began to emerge between 96 and 120 hours, aligning along the midline at the surface and flanked laterally by definitive endoderm (Figure 3a). This pattern resembled the formation of notochordal plate, - a transient embryonic structure where axial mesoderm and endoderm co-develop within one continuous layer[29] (Figure 3j). At 120–144 hours, the notochord-like domain appeared as a cord of BRA- and FOXA2-positive cells (Figure 3k-l).

By expressing morphogens such as Sonic hedgehog *(Shh*), the notochord serves as a key signaling center for neural tube, gut and somite patterning[30,31]. Although dissection of these interactions requires additional experiments, there is indirect evidence of signaling activity originating from the notochord-like cells. RACL-aggregoids contained cell populations resembling the hindbrain floor plate and sclerotome (Supplementary 8b, 10h-i), both of which depend on inductive ventralizing Shh signals from the notochord[32,33] which were absent in control gastruloids.

Altogether, our data suggest that by inducing the emergence of APS-like fates in the aggregoid, ExEnd-like cells initiate a cascade of developmental processes leading to the formation of node- and notochord-like cells.

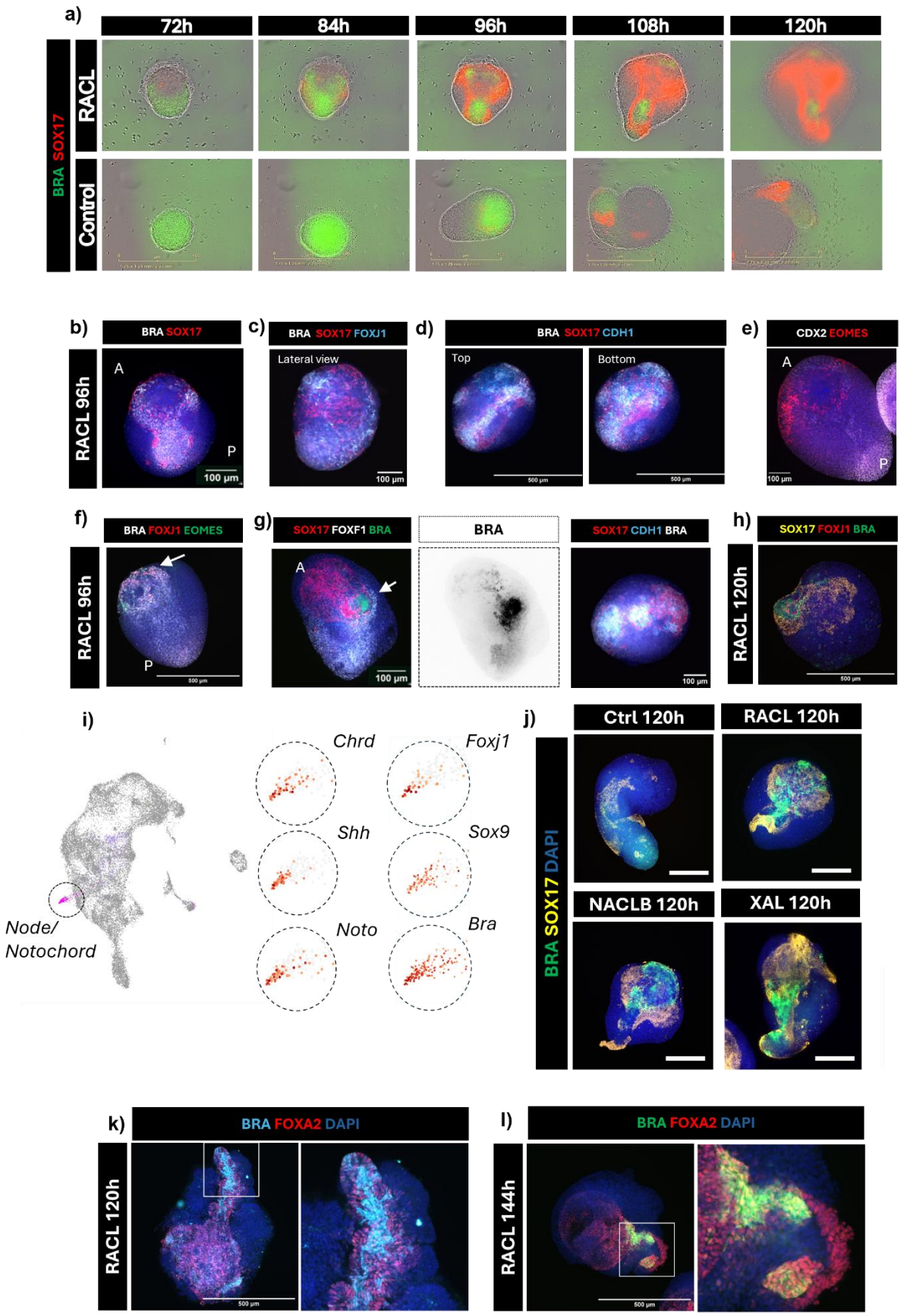

a)
72h
84h
96h
108h
120h
BRA SOX17
RACL
Control
b)
BRA SOX17
RACL 96h
A
P
100 µm
c)
BRA SOX17 FOXJ1
Lateral view
100 µm
d)
BRA SOX17 CDH1
Top
Bottom
500 µm
e)
CDX2 EOMES
A
P
100 µm
f)
BRA FOXJ1 EOMES
RACL 96h
P
500 µm
g)
SOX17 FOXF1 BRA
A
100 µm
BRA
SOX17 CDH1 BRA
100 µm
h)
SOX17 FOXJ1 BRA
RACL 120h
500 µm
i)
Node/
Notochord
Chrd
Foxj1
Shh
Sox9
Noto
Bra
j)
Ctrl 120h
RACL 120h
NACLB 120h
XAL 120h
BRA SOX17 DAPI
k)
BRA FOXA2 DAPI
RACL 120h
500 µm
l)
BRA FOXA2 DAPI
RACL 144h
500 µm

Figure3. The co-aggregation with ExEnd-like cells drive axial mesoderm formation. a) *Time-lapse imaging of control gastruloid and RACL-aggregoid development from 72h and 120h post-aggregation. Phase contrast, fluorescent double reporter mESC line BRA-GFP:SOX17-RFP. Scale bar = 400 µm. b) Confocal image of RACL-aggregoid at 96h, BRA and SOX17 signals from reporter, DAPI.* ***c)*** *Confocal image of RACL-aggregoid, lateral view. SOX17-RFP and BRA-GFP signals from fluorescent reporters; FOXJ1 – immunocytochemistry****. d)*** *Confocal images of RACL-aggregoids at 96h, immunofluorescence; BRA-GFP and SOX17 signals from reporter, CDH1- immunofluorescence;* ***e)*** *Confocal image of RACL-aggregoid at 96h, immunofluorescence****. f-h)*** *Confocal images of RACL-aggregoids at 96h and 120h; SOX17-RFP and BRA-GFP reporters, CDH1, FOXJ1 and FOXF1 – immunofluorescence. Arrowhead points a possible node-like structure.* ***i)*** *UMAP projection of RACL-aggregoid at timeline 96-120h, the Nodo/Noto cluster is highlighted with red and expanded. UMAP projection of Chrd, Noto, Shh, Sox9, Foxj1 and Bra expression within Node/Notochord cluster.* ***j)*** *Confocal images of control gastuloids and RACL, NACB, and XAL-aggregoids at 120h. SOX17-RFP and BRA-GFP signals from reporters.* ***k-l)*** *Confocal images of RACL-aggregoids at 120h and 144h respectively. FOXA2 – immunostaining, BRA-GFP reporter. Co-expression of BRA and FOXA2 marks Notochord.*

Co-aggregation of 2i cells with ExEnd-like cells promotes mesoderm bias and increases mesoderm diversity.

Control gastruloids derived from the FLK-GFP cells grown in 2i/N2B27 conditions displayed a pronounced bias toward neuroectodermal differentiation, exhibiting limited mesodermal differentiation capacity (Figure 4a, b). At 120 hours of development, scRNA-seq analysis revealed that control gastruloids consisted of approximately 5% mesodermal and 72% neural cell types (Figure 4a). Co-aggregation with RACL, NACLB, or XAL cells significantly altered the mesoderm/neuroectoderm ratio, yielding distributions of 49%/32%, 45%/37%, and 28%/61%, respectively (Figure 4a). Immunostaining demonstrated the expansion of FOXC1-positive domain in RACL-aggregoids compared to control gastruloids (Figure 4c).

At 96 hours, control gastruloids typically contained neuromesodermal progenitors (NMPs), a bipotent population co-expressing *Sox2* and *Bra*, which can differentiate into either *Tbx6*+, *Msgn1*+ mesodermal or *Sox1*+, *Sox2*+ neuroectodermal derivatives[34,35] (Supplementary 10e). In RACL- and NACLB-aggregoids, a substantial proportion of NMPs - 37% and 23%, respectively - exhibited mesodermal bias (Supplementary 10b), as indicated by the up-regulation of *Bra*, *Tpx2*, *Lef1*, *Tbx6*, *Prr11*, and *Rspo3* (Supplementary 10f). In contrast, only 5% and 8% of NMPs in control and XAL-gastruloids, respectively, were mesodermally biased at 96 hours, corresponding with reduced mesodermal fates seen at 120 hours (Supplementary 10b).

The diversity of mesodermal lineages differed notably across conditions. In control gastruloids, mesodermal populations were predominantly composed of paraxial mesoderm derivatives, with only a minor contribution from intermediate mesoderm (IM) cells (Supplementary 10g, h). In contrast, ExEnd-aggregoids displayed a broader spectrum of mesodermal lineages, including an increased representation of IM, as well as the emergence of lateral plate mesoderm (LPM) and axial mesoderm (notochord), indicating an increase of mesodermal diversity (Supplementary 8 b-d, 10g-h).

In vivo, the lateral plate mesoderm (LPM) gives rise to a wide array of tissues, particularly cardiac mesoderm and vasculature - lineages that are typically underrepresented in standard gastruloids in the absence of added cardiogenic factors[36]. We used the FLK1(Kdr)-GFP

reporter line[37] to trace cardiovascular development and registered the low frequency (4%) of spontaneous FLK1-GFP signal appearance (Figure 4h). In contrast, ExEnd-aggregoids showed a substantial increase in FLK1-GFP expression detected in 67% of RACL-aggregoids, 77% of NACLB-aggregoids, and 30% of XAL-aggregoids at 120 hours (Figure 4f). Spontaneous contractile activity (Figure 4d) and appearance of CTNT-expressing cardiac cells (Figure 4g) was observed in 70% of RACL aggregoids between 144 and 168 hours. qPCR analysis of pooled RACL-aggregoids revealed an upregulation of cardiogenic (*Gata4*, *Tbx1*, *Ryr2*, *Hcn*) and vasculogenic markers (*Pecam*, *Kdr*/*Flk1*) (Figure 4e).

scRNA-seq further identified a distinct cluster corresponding to cardiopharyngeal mesoderm (Figure 8b-d). This cluster expressed key cardiac lineage genes such as *Prrx2*, *Hand1*, *Hand2*, *Foxf1*, *Gata6*, *Gata4*, as well as first heart field (FHF) markers (*Mesp1*, *Tbx5*) and second heart field (SHF) markers (*Isl1*, *Tbx1*, *Tbx18*). Additionally, two endothelial clusters corresponding to Venous Endothelium and Embryo Proper Endothelium were identified, characterized by expression of *Kdr (Flk1), Flt4*, *Etv2*, *Tal1*, and *Fli1*. (Figure 4g, l, j).

Those results indicate that the addition of ExEnd-like cells enhances LPM development and significantly promotes the emergence of a cardiovascular transcriptional program in ExEnd-aggregoids, even in the absence of exogenous cardiogenic stimuli.

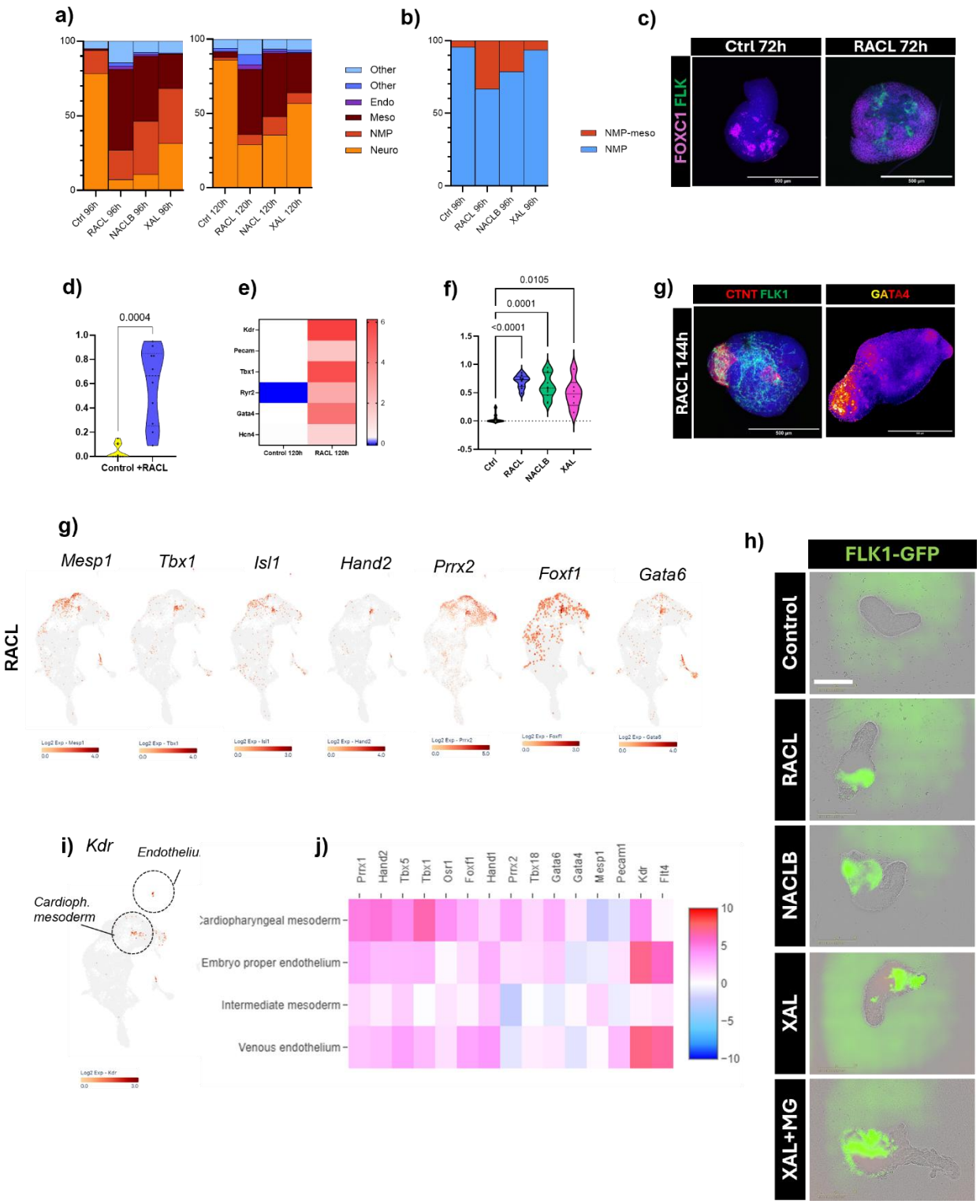


Figure 4. Co-aggregation of 2i cells with ExEnd-like cells promotes mesoderm bias and increases mesoderm diversity. a) *Bar plots demonstrating the proportions of cell populations in control gastuloids, RACL-, NACLB, and XAL- aggregoids at 96h and 120h based on scRNA-seq data.* ***b)*** *Repesentative confocal images demonstrating mesodermal populations marked by FOXC1 in control gastruloid and RACL aggregoid. Immunostaining (FOXC1), repoter (FLK1-GFP), DAPI.* ***c)*** *Violin plot demonstrating the percentage of gastuloids or aggregoids with beaing activity at 144h. (n=10). Y axis is proportion of structures with the registerd beating activity in independent experiments. For statistical comparisons non-*

parametric t-test used. ***d)*** *Heatmap of cardiovascular markers expression, normalized by control gastuloids, qPCR (n=3).* ***e)*** *Violin plots showing the percentage of embryo-like structures having FLK1-GFP reporter signal at 120h. Y axis means the proportion of aggregoids or gastruloids with FLK1-GFP signal. Kruskal-Wallis test.* ***f)*** *Confocal images of 144h RACL-aggregoid; CTNT – immunostaining, FLK1 – fluorescence from reporter, DAPI.* ***g)*** *UMAP projection showing cardiovascular markers expression in RACL-aggregoids (integrated dataset 48-120h).* ***h)*** *Phase contrast and FLK1-GFP reporter expression in control gastruloids and aggregoids at 120h.* ***i)*** *UMAP projection of Flk1/Kdr expression in cardioharingeal and endothelium clusters.* ***j)*** *Heatmap of cardiovascular markers expression in Cardiopharingeal mesoderm, Venous Endotheliun, Embryo proper endothelium and intermediate mesoderm clusters*

ExEnd-like cells promote DE induction and proper organization.

In the mouse embryo, endoderm arises from two distinct origins: the early extra-embryonic endoderm, known as primitive endoderm (PrE), that is specified during the pre-implantation stage, whereas the definitive endoderm (DE) emerges during gastrulation from the APS10. Cells from this region migrate through the streak and displace the visceral endoderm (VE) originating from the PrE ultimately giving rise to the epithelial lining of the gut tube and associated organs such as the liver, pancreas, and lungs[38].

In scRNA-seq dataset, we detected extraembryonic endoderm clusters in RACL-, NACLB- and XAL-aggregoids, at all time points from 48 to 120h, but not in control gastruloids (Figure 5b). The extra-embryonic and definitive endoderm clusters shared pan-endodermal markers such as *Sox17, Foxa2, Cdh1,* and *Gata6* (Figure 5d). However, those populations could be segregated by the expression of *Ttr*, *Apoe*, *Gata4* and *Dab2* for extra-embryonic endoderm and *Foxa1*, *Trh, Shh, Ceacam1, Tgfb2* and *Kitl* for definitive endoderm. The ExEnd-like population was further sub-divided into two clusters corresponding to visceral endoderm (VE) and parietal endoderm (PE). The VE cluster contained cells expressing *Amn, Podxl, Dab2, Foxa2, Igf2, Krt8, Krt18, Krt19, Car4*, and *Cdh1*. However, it lacked mature VE markers such as *Apoa*, *Sfrp1,* and *Fgf8.*while retaining expression of primitive endoderm (PrE)-associated genes including *Dkk1, Col4a1, Cubn, Dppa5a, Lamb1, and Lama1. Notably, Hnf1b,* a key regulator of VE maturation and epiblast survival, was upregulated in this cluster. The PE cluster was characterized by expression of *Mycn, Cubn, Vim, Nid1,* and *Snai1*, although it lacked some canonical PE markers such as *Thbd* and *Grem2*. At 48–72 hours, the VE cluster still contained *Dkk1*-positive cells, suggesting a potential role in WNT inhibition (Supplementary 6d).

The capacity of gastruloids to form DE is generally considered to be limited and variable, showing substantial variability depending on the cell line and conditions[39]. In control gastruloids generated from the Flk1-GFP cell line, the deficiency of endoderm was especially pronounced: SOX17+ and CDH1+ cells were either absent or underrepresented, often failing to form tubes (Figure 5a). In contrast, co-aggregation with ExEnd-like cell types led to formation of continuous structure composed of $SOX17^{+}$, $CDH1^{+}$ cells and, in the scRNA-seq dataset, the expression of *Sox17, Kitl, Foxa2, Foxa1* and *Trh*-positive cells indicative for DE populations, which were absent in the control gastruloids (Figure 5d). In RACL-aggregoids, the DE clusters were predicted as foregut (including pharyngeal endoderm), midgut, gut tube, and hindgut, while NACLB-aggregoids contained foregut, gut tube and midgut. XAL aggregoids exhibited DE restricted to the Foregut and Gut tube (Figure 5e).

To investigate additional effects of ExEnd-like cells on DE formation, the mESC lines BRA-GFP: SOX17-RFP was selected for their robust capacity to generate endoderm in canonical gastruloids. Despite the presence of a substantial number of SOX17-positive cells in control

gastruloids at 120 hours, definitive endoderm (DE) mostly failed to organize into a continuous epithelial tube. Instead, the endodermal compartment consisted of a large, posteriorly localized domain adjacent to the BRA$^+$ region at the tip, along with small, scattered islets exhibiting weak CDH1 staining (Figure 5c). Upon co-aggregation with RACL-, NACLB-, or XAL-derived cells, however, the endoderm at 120 hours formed a continuous tubular structure, with a profound anterior domain. (Figure 5c).

Comparative live imaging from 72 to 96 and 120h using BFP-ROSA26 cell line for RACL cells and BRA-GFP SOX17-RFP double reporter line as 2i cells (for the gastruloid part) revealed the differences in DE dynamics. In control gastruloids, SOX17-RFP expression typically initiated between 88–96 hours, emerging deep within the aggregate adjacent to the posterior BRA-GFP domain, and gradually extending anteriorly (Figure 3a). In contrast, in RACL-aggregoids, SOX17-RFP–positive cells were observed as early as 76–80 hours, shortly after the onset of BRA-GFP expression (Figure 3a). In RACL-aggregoids, these cells initially appeared at the surface in regions adjacent to the ExEnd-like cap, specifically at the interface between the cap and the emerging BRA-GFP–positive domain. Over time, they moved anteriorly and began to integrate into the ExEnd-like layer. (Movie 1).

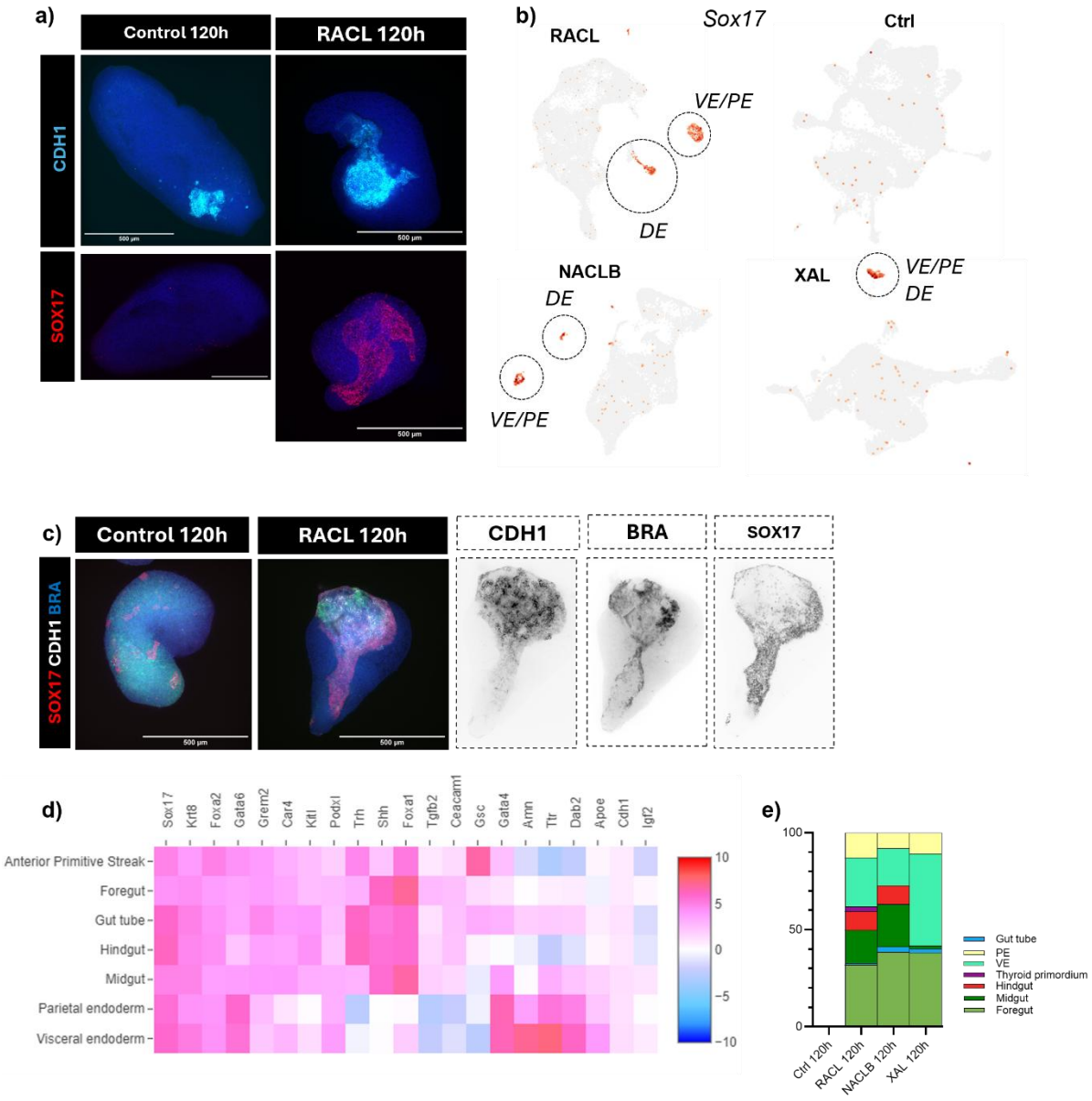


Figure 6 ExEnd-like cells promote DE induction and proper organization. a) *Representative immunofluoresent images showing CDH1 and SOX17 expresion in control gastruloids and RACL-aggregoids made of FLK1-GFP reporter line, which shows poor abilities for endoderm formation in standard conditions.* ***b)*** *UMAP projection of endodermal clusters in RACL-aggregoids (Definive endoderm, Visceral endoderm, Parietal endoderm) marked by Sox17 expression in control, RACL, NACLB and XAL conditions.* ***c)*** *Immunostaining for SOX17 and CDH1 showing the differences in endoderm organization between control gastruloids and RACL-aggregoids made of BRA-GFP SOX17-RFP cell line.* ***d)*** *Heatmap showing expression of DE and ExEnd markers in all endodermal populations present at 120h (APS, foregut, gut tube, hindgut, midgut, PE and VE.* ***e)****Bar plot showing proportions of endodermal cell types in aggregoids. Data for ctrl are absent because there was no endoderm in the control dataset.*

Discussion

Gastruloids represent a 3D model of embryonic development, generated by aggregating pluripotent stem cells and applying a CHIR pulse to mimic key features of the complex process of gastrulation such as symmetry breaking, pattern formation, and lineage segregation, in the absence of extra-embryonic structures traditionally considered essential for these developmental processes in the embryo. The reproducible yet modular nature of gastruloids make them well-suited for targeted testing of the effect of specific features of the natural embryo on embryonic development.

Berenger-Currias et al.[40], combined the mESC and mouse embryo-derived XEN cells in structures termed XEGs (XEN-gastruloids), proposing a model for studying crosstalk between extraembryonic endoderm and epiblast and demonstrating the appearance of anterior neuroepithelial structures, probably, due to local WNT inhibition and laminin-mediated signaling. However, prior to co-aggregation, XEN cells transcriptionally resembled parietal endoderm (PE) and did not express AVE markers, features that might cause some limitations of their model. Indeed, studies using mouse SCBEMs, such as ETX[41-44], have demonstrated that the origin and developmental state of the extra-embryonic endoderm cells are critical determinants of model outcome.

Here, we tested three different extra-embryonic endoderm-like cell populations derived from 2i mESCs as an alternative to the above used XEN cells. For this, we employed published differentiation protocols based on a modulation of the WNT, Nodal, JAK/STAT3 signaling pathways[14] and its modifications. Bulk RNA sequencing analysis of those populations revealed the general resemblance of those cell types to mouse extraembryonic endoderm with signatures of PrE, VE and AVE including a distinct repertoire of signaling molecules; however, co-aggregation, incubation in N2B27 and CHIR pulse caused the changes in expression profile of RACL, NACLB and XAL cells and full or partial loss of AVE markers. This might be consistent with the transient expression of *Cer1* and *Lefty in vivo* and in embryo models[16,45]. Probably due to the described co-aggregation conditions, the introduced extraembryonic endoderm-like cells underwent changes that limited their functional and molecular distinctions as AVE-, PrE-, and VE-like cell types, resulting in highly similar phenotypes and restricting the potential of this model to recapitulate the full spectrum of signaling interactions. For example, we did not observe the reproducible formation of anterior forebrain structures upon co-aggregation with XAL cells, - an outcome that would be expected from co-aggregation with cells possessing AVE-like identity[46,47]. The induction of anterior neural tissues depends not only on inhibition of WNT signaling but also on suppression of TGFβ and BMP pathways[48,49]. At the time of CHIR exposure, however, XAL cells within the aggregoids had lost expression of *Lefty1* and *Cer1* and displayed upregulation of *Bmp* genes, that might have compromised the anteriorizing potential of this cell type. This limitation could potentially be mitigated through an alternative co-aggregation strategy, such as adding ExEnd-like cells shortly before the CHIR pulse to maintain their signaling capacity. The ExEnd derived populations within the aggregoids expressed the WNT signaling inhibitor *Dkk1*, along with ligands of several key developmental signaling pathways, including Indian hedgehog (*Ihh*), non-canonical WNTs, and BMPs. This expression profile suggests that extraembryonic endoderm–derived cells in ExEnd-aggregoids not only modulate local WNT activity but may also provide paracrine cues that influence epiblast-derived tissues, potentially affecting patterning events and mesoderm diversity upon co-aggregation. Thus, *Ihh* signaling is critical for LPM specification[50], and non-canonical *Wnt11* and *Wnt5a* participate in heart induction[51].

In contrast to XEGs, ExEnd-aggregoids did not exhibit a pronounced anterior neuroepithelial phenotype. Although RACL-, NACLB-, and XAL cells expressed laminins, their endogenous

levels appear to be insufficient to reproduce the effects of the extracellular matrix (ECM) that have been observed in XEG models. Consequently, while somitic mesoderm and neural tube–like cell populations emerged in ExEnd-aggregoids, *a priori* those structures remained largely unorganized and failed to establish epithelialized structures.

The incorporation of ExEnd-like cells in the aggregoids efficiently promoted early anteroposterior (AP) patterning, resulting in asymmetric BRA expression as early as 72 hours. From the onset, BRA expression was confined to the posterior pole, establishing a clearly defined BRA-negative anterior domain. In contrast to XEGs, this BRA-negative region was not limited to a narrow stripe adjacent to the extraembryonic compartment but instead extended over approximately one-half to one-third of the ExEnd-aggregoid, indicating an expanded anterior territory. Notably, BRA restriction did not lead to the emergence of SOX2-positive anterior neural structures; instead, this area displayed EOMES-positive anterior primitive streak–like features. By 84–96 hours, an additional anterior BRA domain appeared within this EOMES-expressing region, suggesting dynamic reorganization and possible anterior streak activity reminiscent of early mesendoderm formation. This might show the role of ExEnd-like cells not only in anterior neuroectoderm formation but also in structuring the primitive streak and promoting diversity of cell lineages within ExEnd-aggregoids as underscored by the scRNA-seq analysis.

In this work, we show that co-aggregation with ExEnd-like cells causes the shift towards pro-mesodermal bias of NMPs fate and general prevalence of mesodermal cell types. This can be explained either by pro-mesodermal effect of BMPs (2/5/6/7) expressed in ExEnd-like cells[52]

The ExEnd-aggregoids also displayed an up-regulation of Nodal and associated genes, the appearance of an anterior primitive streak-like structure, a node-like structure and derivates, the specification of which depends on Nodal, such as the notochord, anterior mesoderm and definitive endoderm. This deployment is consistent with recent findings by[8] Dias et al. (2025), demonstrating that gastruloids treated with Activin A or low dose of CHIR, exhibited features of anterior NODAL/EOMES-driven development. However, while those gastruloids lacked posterior WNT-dependent derivatives, the transcriptomic profile of RACL-aggregoids does not indicate a loss of posterior identity. This suggests that early spatial compartmentalization within the aggregoids, likely driven by partial cover of the 2i-derived population with DKK1-expressing ExEnd-like cells in the aggregoids, may support localized, endogenously established, morphogen gradients. In this scenario, such gradients would enable the coexistence of anterior and posterior programs by regionally modulating WNT activity, thereby supporting a more comprehensive axial patterning.

Despite their reproducibility and wide applicability, the commonly posteriorized identity and underrepresentation of endodermal, axial mesoderm, and cardiovascular derivatives remain major limitations of canonical gastruloids, restricting both their developmental potential and their applicability for studying certain aspects of embryonic development. Some of these lineages can be partially recovered through the addition of exogenous growth factors — for instance, in cardiogenic conditions[36] or in Activin A-treated gastruloids modeling early primitive streak formation[8]. However, such interventions act uniformly across the entire aggregate, whereas an embryo establishes its pattern through a finely tuned interplay of spatial gradients, localized inhibitors, and feedback mechanisms. In contrast, ExEnd-aggregoids reproducibly induce robust anteroposterior (AP) regionalization and lineage diversity without the need for additional exogenous factors beyond CHIR, providing an expanded physiologically relevant model for generating and studying aspects of mouse embryonic development.

Materials and Methods

Mouse Pluripotent Stem Cell culture and naive conversion

The following mouse embryonic stem cell (mESC) lines were used in this study: R1 (SCRC-1011, ATCC), Flk1-eGFP (Jakobsson et al., 2010), Rosa26-BFP/Flk1-GFP (unpublished; provided by Anna Bigas' laboratory), and Bra-GFP/Sox17-RFP (Pour et al., 2022). mESCs were cultured on 6-well culture plates covered with 0,1% Gelatine solution (ES-006-B, Merck) in serum + LIF medium consisting of KnockOut DMEM (10829018, Gibco) supplemented with 15% fetal bovine serum (FBS; 10309433, HyClone), non-essential amino acids (NEAA; 11140050, Gibco), GlutaMAX (35050061, Gibco), 0.1 mM β-mercaptoethanol (M3148, Merck), and mouse leukemia inhibitory factor (LIF; ESG1107, Merck).

For naïve state conversion, mESCs were plated on wells coated with 0.01% ornithine (P3655, Merck) and cultured in 2i/N2B27+LIF medium. 2i/N2B27+LIF composition was: 1:1 mixture of DMEM/F12 (11330032 or 11039021, Gibco) and Neurobasal medium (21103049 or 12348017, Gibco), supplemented with N2 (17502048, Gibco), B27 (17504044, Gibco), GlutaMAX (35050061, Gibco), NEAA (11140050, Gibco), sodium pyruvate (11360039, Gibco), β-mercaptoethanol (M3148, Merck), bovine serum albumin (BSA; A8412, Merck), LIF (ESG1107, Merck), CHIR99021 (3 μM; 4423, Tocris), and PD0325901 (1 μM; 4192, Tocris). Cells were maintained in naïve conditions for 5–7 days prior to downstream applications.

All cell cultures were incubated at 37 °C in 100% humidity with 5% CO_2. Both standard and 2i/N2B27+LIF media were refreshed daily. Cells were passaged every other day using Accutase (A6964, Merck). All cell cultures were routinely tested for Mycoplasma using BioNordika LonzaT MycoalertT Mycoplasma Detection Kit (LT07-318) every month.

Differentiation of ExEnd-like cell types

Mouse embryonic stem cells (mESCs) cultured in 2i/N2B27+LIF medium for 5–7 days were harvested using Accutase (A6964, Merck) or TrypLE Express (12604013, Gibco) and replated at a density of 1×10^5 cells/cm² on 0.1% gelatin-coated 6-well plates. Cells were seeded in RPMI 1640 medium (11875093, Gibco) supplemented with 1×B-27 Supplement Minus Insulin (A1895601, Gibco). After first 24 hours, the medium was replaced with RACL medium, consisting of RPMI 1640 (11875093, Gibco), 1×B-27 Supplement Minus Insulin (A1895601, Gibco), 100 ng/mL ActivinA (Peprotech, 120-14P), 3 μM CHIR99021 (4423, Tocris Bioscience), and mouse LIF (ESG1107, Merck). The cultures were passaged once between days 3 and 4 of the protocol. On days 6–8 of RACL treatment, cells were either used for co-aggregation experiments in RACL conditions or subjected to further differentiation protocols.

To initiate further differentiation, RACL medium was replaced with NACL medium for 24 hours. NACL medium consisted of N2B27 basal medium composed of 1:1 mix of DMEM/F12

(11320033, Gibco) and Neurobasal medium (21103049, Gibco) supplemented with N2 Supplement (17502048, Gibco), B-27 Supplement (17504044, Gibco), GlutaMAX (35050061, Gibco), non-essential amino acids (11140050, Gibco), sodium pyruvate (11360039, Gibco), β-mercaptoethanol (M3148, Merck), and BSA (A8412, Merck), supplemented with 100 ng/mL Activin A (Peprotech, 120-14P), 3 μM CHIR99021 (4423, Tocris Bioscience), and mouse LIF (ESG1107, Merck).

For differentiation into visceral endoderm-like (VE-like) cells, NACL medium was further supplemented with 50 ng/mL recombinant murine BMP4 (Peprotech, 315-27), referred to as NACL+B. To generate anterior visceral endoderm-like (AVE-like) cells, cells previously maintained in NACL were treated with XAL medium, consisting of 10 μM XAV939 (S1180,

Selleck Chemicals), 100 ng/mL Activin A (120-14P, Peprotech), and mouse LIF (ESG1107, Merck). Both NACL+B and XAL treatments were applied for three days.

Generation of gastruloids and aggregoids

Gastruloids were generated following previously established protocols[2,3]. Briefly, 300 mouse embryonic stem cells (mESCs) were plated in 40 µL of NDiff 227 medium (Y40002, Takara) or home-made N2B27 medium (composition described above) into each well of a 96-well U-bottom cell-repellent microplate (Greiner Bio-One, Cat# 650970). Naïve mESCs grown for 5-7 days in 2i/N2B27+LIF conditions were used as starting material. At 48 hours post-aggregation, 150 µL of NDiff or N2B27 medium supplemented with 3 µg/mL of CHIR99021 was added to each well. Media change was performed every 24 hours by removing 150 µL and replacing it with fresh NDiff or N2B27 medium. Gastruloids were harvested for analysis at 48, 72, 96, 120, 144, and 168 hours.

To generate aggregoids, 300 naïve mESCs were mixed with 100 additional cells (RACL, NACL+B, or XAL) at the aggregation step. The cell mixture was plated under the same conditions as described above for gastruloid formation. All subsequent steps, including CHIR99021 treatment and daily media changes, were performed as described for the standard gastruloid protocol. For “Minus Chir” experiments, Ndiff 227 (Takara, Y40002) of home-made N2B27 was added to the wells at 48h.

RNA extraction and RT-qPCR

Total RNA was extracted from either cell cultures or pooled gastruloids using the RNeasy Micro Kit (Qiagen, 74004), following the manufacturer’s instructions. For each extraction, 48–96 gastruloids or 300,000–1,000,000 cells were collected and stored at –80 °C until processing. RNA concentration was quantified using a Nanodrop spectrophotometer (ThermoFisher Scientific).

Complementary DNA (cDNA) was synthesized using the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific, 4368814). Quantitative real-time PCR (RT-qPCR)

was performed using a ViiA 7 thermocycler (ThermoFisher Scientific) with TaqMan Gene Expression Master Mix and gene-specific TaqMan probes (ThermoFisher Scientific). Primer details are listed in Table X.

Each qPCR reaction contained 1 µL of cDNA (diluted to 20 ng/µL) in a total volume of 10 µL, conducted in 386-well plates (Applied Biosystems, 4309849). Gene expression levels were normalized to the housekeeping gene Gapdh (probe ID: Mm01180221_g1, ThermoFisher Scientific, 4331182), and relative expression was calculated using the 2^–ΔΔCT method. The number of technical replications was 3, the number of biological replications was a minimum 3. 2i/N2B27 cells were used as a reference sample for normalization.

Bulk RNA-sequencing

Total RNA was extracted from totally 12 mouse cell samples: naive mESCs in 2i/N2B27 conditions and differentiated extraembryonic endoderm-like cell types: RACL, NACLB, and XAL. 2ug of total RNA in 100ul of RNAse-free water was used for library preparation. RNA quality was assessed using an Agilent Bioanalyzer. RNA-seq libraries were prepared using a strand-specific, poly(A)-selected protocol, TruSeq Stranded mRNA Library Prep (20020595, Illumina), and adapters TruSeq RNA UD Indexes (Cat.Nr. 2002237, Illumina). Sequencing was performed on an Illumina NovaSeq X Plus system using 1.5B flow cells, generating paired-end 150 bp reads. Libraries were multiplexed and sequenced across two lanes, and FASTQ files from the two lanes were merged prior to downstream analysis. Sequencing

instrumentation and software: NovaSeqXPlus System Suite v1.2.2.48004, BCL Convert v4.1.23, NSC scripts v1.0.

BBDuk (BBMap toolkit) was used to remove/trim low-quality reads, adapter sequences, and PhiX used as a spike-in for sequencing (10_bbduk folder). Clean and adapter removed reads were mapped to the human genome/transcriptome using hisat2 (20_hisat2 folder). FeatureCounts was used to count the reads mapping to the genes (30_featureCounts folder). Trimmed reads were aligned to the mouse genome using HiSat2 v2.1.0 with the RNA-strandness RF flag to account for strand-specific library preparation. The reference genome and annotation were from Ensembl release 113: Genome FASTA: Mus_musculus.GRCm39.dna.toplevel.fa; Gene annotation GTF: Mus_musculus.GRCm39.113.gtf. Aligned files were sorted and indexed using Samtools v1.2 (with HTSlib v1.2.1). Gene-level quantification was performed using FeatureCounts v1.4.6-p1 with the parameters: FeatureCounts outputs were organized in the 30_featureCounts folder, with supporting scripts and logs stored in corresponding *.sh and *.log files.

The gene counts from the featureCounts output were loaded into R (v4.5.1) and combined across all samples. A preliminary quality check was conducted to identify any technical outliers or anomalies. Subsequently, we filtered out genes with fewer than 50 total counts across the samples, ensuring they were expressed in at least two of the eight samples. The quality-controlled count matrix was loaded into DESeq2 (v1.48.1) using the DESeqDataSetFromMatrix function. Then, we applied the Variance Stabilizing Transformation (VST) to normalize the counts for PCA and hierarchical clustering, and to identify the leading genes. Differentially expressed genes were identified across all condition combinations; for example, 2i vs RACL, 2i vs NACL, etc. Volcano and MA plots were generated for each comparison. Subsequently, we conducted gene over-representation and pathway analyses using clusterProfiler (v4.16.0) for the up- and down-regulated genes.

Single cell suspension preparation and fixation.

Gastruloids corresponding to four conditions: standard gastruloid, RACL aggregoids, NACL+B aggregoids, XAL aggregoids, - were collected at 48h, 72h, 96h, and 120h. For each condition: 384 gastruloids were pooled at 48h, 288 gastruloids at 72h, 192 gastruloids at 96h and 96 gastruloids at 120h. For preparation of single cell suspension, the gastruloids were dissociated according to published protocol (Bolondi et al., 2021) with our modifications. The objects were collected and transferred to a Petri dish with 200 µl of TrypLE (12604013, Gibco), incubated at 37°C with pipetting every 5 min until full dissociation to a single-cell suspension. Then TryLE was quenched with 800 µl of ice-cold PBS + 0,5% BSA (A8412, Merck). After filtering through a cell strainer (27215, Stemcell Technologies), the cell suspension was centrifuged at 4°C, 300 rcf for 5 min. After discarding the supernatant, the cell pellet was washed with 1 ml of PBS + 0,04% BSA and centrifuged at 4°C, 300 rcf for 5 min. Washing was repeated one more time and finally the cell pellet was dissociated in 1 ml of PBS + 0,04% BSA. Cell viability was assessed by Trypan Blue staining. Amount of dead cells was less than 5% for all stages and conditions. Then cell suspension was fixed using Chromium Next GEM Single Cell Fixed RNA Sample Preparation Kit, 16 rxns (1000414; 10X Genomics) and stored at -80C or +4C until library preparation. Samples with not les than 300 000 cells after fixation and washing were used for library construction.

Samples were multiplexed, and cDNA libraries construction was performed from 10 000 cells per sample according to the Chromium Fixed RNA Kit (10x-1000496,10X Genomix) user guide. Quality control and concentration measurements of cDNA libraries prior sequencing were made using Agilent BioAnalyzer high sensitivity DNA kit on Agilent 2100 Bioanalizer

(Agilent Technologies). Single-cell RNA sequencing was done in Norwegian Sequencing Centre on NovaSeq 1 platform 1 lane 25B 300 cycles.

The sequenced data was aligned against the murine mm10 reference genome using the 10x Genomic Cell Ranger (v8.0.1) pipeline, de-multiplexing the samples in silico accordingly. The samples were then aggregated in the same pipeline, without normalization and batch-correction, or used non-aggregated where applicable.

The cell barcode matrices generated by the CellRanger pipeline (v8.0.1) were loaded into R (v4.5.1) using the Seurat library (v5.3.0). The first-pass quality control required cells to have at least 200 features and no more than 7500, and UMI counts to be between 400 and 40000. The mitochondrial gene content was set to below 25%. The mitochondrial gene content cut-off was determined by a data-driven method. Then, a simple Seurat pipeline was run to identify clusters, primarily based on the number of UMI and feature counts. Next, second-pass QC was performed by removing clusters mainly driven by low UMI/feature counts. We verified that removed clusters with low UMI/feature counts were solely due to low counts and not driven by biologically relevant signals, using Seurat's FindMarker function. Next, cells were scored for doublets using scDblFinder (v1.22.0). Cells with scDblFinder_score > 0.7 were excluded from further analyses.

The quality-controlled cells were integrated based on conditions (medium used) and time using Seurat's RPCA-based integration method. We used 5000 integration features for the process. Cells were annotated by transferring labels from the Mouse gastruloid reference atlas. Some cell types were merged into a standard label to reflect the biological type.

Microscopy and live imaging.

Automated live imaging of cell cultures and gastruloids with fluorescent reporters was made using EVOS M7000 (Thermo Fisher Scientific), IncuCyte S3 (Sartorius) and Nikon Crest Spinning disk (X-light V3 CREST). Time lapse images of gastruloids and RACL aggregoids were acquired using a Andor Dragonfly Spinning Disk confocal microscope (Dragonfly) with Zyla sCMOS camera, 25 micron pinhole disk and Nikon 20X/0.70 or 10X/0.40 air objectives. Images were taken every 15 minutes for a total duration of 24 hours. The microscope was equiped with a weather chamber from Okolab, which maintained 5% CO_2 and appropriate humidity levels throughout the imaging period. The Perfect Focus System (PFS) was utilized to ensure the focus of the sample was maintained during the entire imaging process.

Immunofluorescent staining and microscopy of fixed samples.

Gastruloids and cell culture were fixed in 4% PFA for 1h at room temperature on a shaking platform then washed three times with PBS and stored at +4°C in PBS. For immunostaining the specimen were first incubated in blotting solution: PBS (14190, Gibco), 10% BSA (422361V, VWR Life Science), 0,5% Triton x-100 (M143, VWR Life Science), 0,02 % SDS (L3771, Sigma) for 1h at the room temperature, then incubated with first antibodies diluted in the blotting solution for overnight at +4°C. After the incubation with primary antibodies the gastruloids were washed with blotting solution 3 times for 30 min at the room temperature and incubated with secondary antibodies diluted in blotting solution for 3h at the room temperature. After the incubation with secondary antibodies the gastruloids were washed with blotting solution 3 times for 30 min. The following primary antibodies were used: mouse anti-DKK1 (1:100, Santa Cruz, sc-374574), goat anti-Brachyury (1:200, R&D Systems, AF2085), rat anti-Eomesodermin (1:100, ThermoFisher, 14-4875-82), rabbit anti-FoxJ1 (1:200, Abcam, ab235445), mouse anti-Dab2 (1:200, Santa Cruz, sc-136964), rabbit anti-FoxF1 (1:200, Abcam, ab308633), rabbit anti-FoxA2 (1:200, Merck-Millipore, 07-633), rabbit anti-Otx2

(1:200, ThermoFisher, 13497-1-AP), mouse anti-Sox2 (1:200, R&D Systems, MAB2018), rabbit anti-FoxC1 (1:200, Abcam, ab223850), mouse anti-Cardiac Troponin T (1:200, ThermoFisher,MA5-12960), mouse anti-Tuj1 (1:200, Abcam, ab78078), goat anti-E-cadherin (1:200, R&D Systems, AF748), rat anti-PDGFRα (1:200, ThermoFisher, 14-1401-82), mouse anti-GATA4 (1:200, Santa Cruz, sc-25310), rabbit anti-HNF4α (1:200, ThermoFisher, MA5-32295), rabbit anti-Lefty (1:100, ThermoFisher, BS-11236R). For cell membrane staining we used Wheat Germ Agglitinin (WGA) Alexa Fluor 647 conjugated (1:250, ThermoFisher, W32466). For actin staining we used Phalloidin Alexa Fluor 488 conjugated (1:200, ThermoFisher, A12379).

The following secondary antibodies were used: donkey anti-rabbit Cy3 (1:500, Jackson ImmunoResearch, 711-165-152), donkey anti-rabbit AlexaFluor 647 (1:500, Jackson ImmunoResearch, 711-605-152), donkey anti-mouse Cy3 (1:500, Jackson ImmunoResearch, 715-165-150), donkey anti-mouse AlexaFluor 488 (1:500, Jackson ImmunoResearch, 715-545-150), donkey anti-mouse AlexaFluor 647 (1:500, Jackson ImmunoResearch, 715-605-150), donkey anti-goat Cy3 (1:500, Jackson ImmunoResearch, 705-165-003), donkey anti-goat AlexaFluor 488 (1:500, Jackson ImmunoResearch, 705-545-147), donkey anti-goat AlexaFluor 647 (1:500, ThermoFisher, A32849), donkey anti-rat Cy3 (1:500, Jackson ImmunoResearch, 712-165-153), donkey anti-rat AlexaFluor 488 (1:500, ThermoFisher, A48269).

All the incubations and washing steps were performed at the shaking conditions. After the staining the gastruloids were mounted in a RapiClear solution (152001, SunJin Lab) between slide and cover slip. The samples were visualized with X Light V3 CREST confocal spinning disc microscope (Nikon).

Images were processed using Fiji (ImageJ), an open-source image analysis software[53].

Acknowledgements

The project has received funding from the University of Oslo, the University Hospital of Oslo, the Research Council of Norway through its Centers of Excellence scheme grant Grant ID 262613, EU's H2020 Marie Skłodowska-Curie Actions / COFUND Grant ID 801133 and the European Innovation Council (EIC) (Horizon Europe) Pathfinder Challenge "Supervised Morphogenesis in Gastruloids" Grant ID 801133. We thank Anna Bigas and Susanne van den Brink (Hospital del Mar Medical Research Institute, Barcelona, Spain) for providing Rosa-BFP/Flk1-GFP mESC, Alexander Medvinsky (The University of Edinbugh, Edinburgh, UK) for providing Flk1-GFP mESC and Iftach Nachman (Tel-Aviv University, Israel) for providing Bra-GFP/Sox17-RFP mESC. We would like to thank Xian Hu(Edna) From the NorMIC Imaging Platform at the Department of Biosciences, University of Oslo, for providing assistances and access to the spinning disk microscope for imaging of gastruloids.

Declaration of interests

The authors declare no competing interests.

Author information

Affiliations:

Hybrid Technology Hub - Centre of Excellence, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway:
Natalia Smirnova, Sergei Ponomartcev, Eric Maxmilian Lycke, Espen Melum, Stefan Krauss

Department of Immunology and Transfusion Medicine, Oslo University Hospital, Oslo, Norway:

Natalia Smirnova, Sergei Ponomartcev, Stefan Krauss

The Experimental Liver Research Group, Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway: Espen Melum
Stembryogenesis Lab, Max Planck Institute of Molecular Cell Biology and Genetics, 01307, Dresden, Germany: Jesse Veenvliet

Cluster of Excellence Physics of Life, TU Dresden, 01062 Dresden, Germany: Jesse V Veenvliet

Center for Systems Biology Dresden, 01307 Dresden, Germany: Jesse V Veenvliet

Liyakat Ali Bioinformatics: Tharvesh Liakat

Contributions

N.S. and S.P. contributed equally. N.S. S.P. and S.K. conceived and designed the experiments; N.S. and S. P. carried out the experiments, N.S. and S. P., J. V.V., T.L and E.M.L. analyzed and elaborated data, T.L. executed the scRNA-seq analysis; J.V.V. and E.M.L. provided critical comments to the manuscript; N.S. S.P. and S.K wrote the manuscript; S.K. provided funds.

Ethical statement

All mouse stem cell lines used in this study were obtained from established repositories. No new animals were used, and therefore ethical approval was not required under the Norwegian Regulation on the Use of Animals in Research.

References

1. van den Brink, S.C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.K., Nowotschin, S., Turner, D.A., and Martinez Arias, A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. Development *141*, 4231–4242. 10.1242/dev.113001.
2. Baillie-Johnson, P., van den Brink, S.C., Balayo, T., Turner, D.A., and Martinez Arias, A. (2015). Generation of Aggregates of Mouse Embryonic Stem Cells that Show Symmetry Breaking, Polarization and Emergent Collective Behaviour In Vitro. J Vis Exp. 10.3791/53252.
3. Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.C., Lutolf, M.P., Duboule, D., and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature *562*, 272–276. 10.1038/s41586-018-0578-0.
4. Bardot, E.S., and Hadjantonakis, A.K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate. Mech Develop *163*. ARTN 10361710.1016/j.mod.2020.103617.

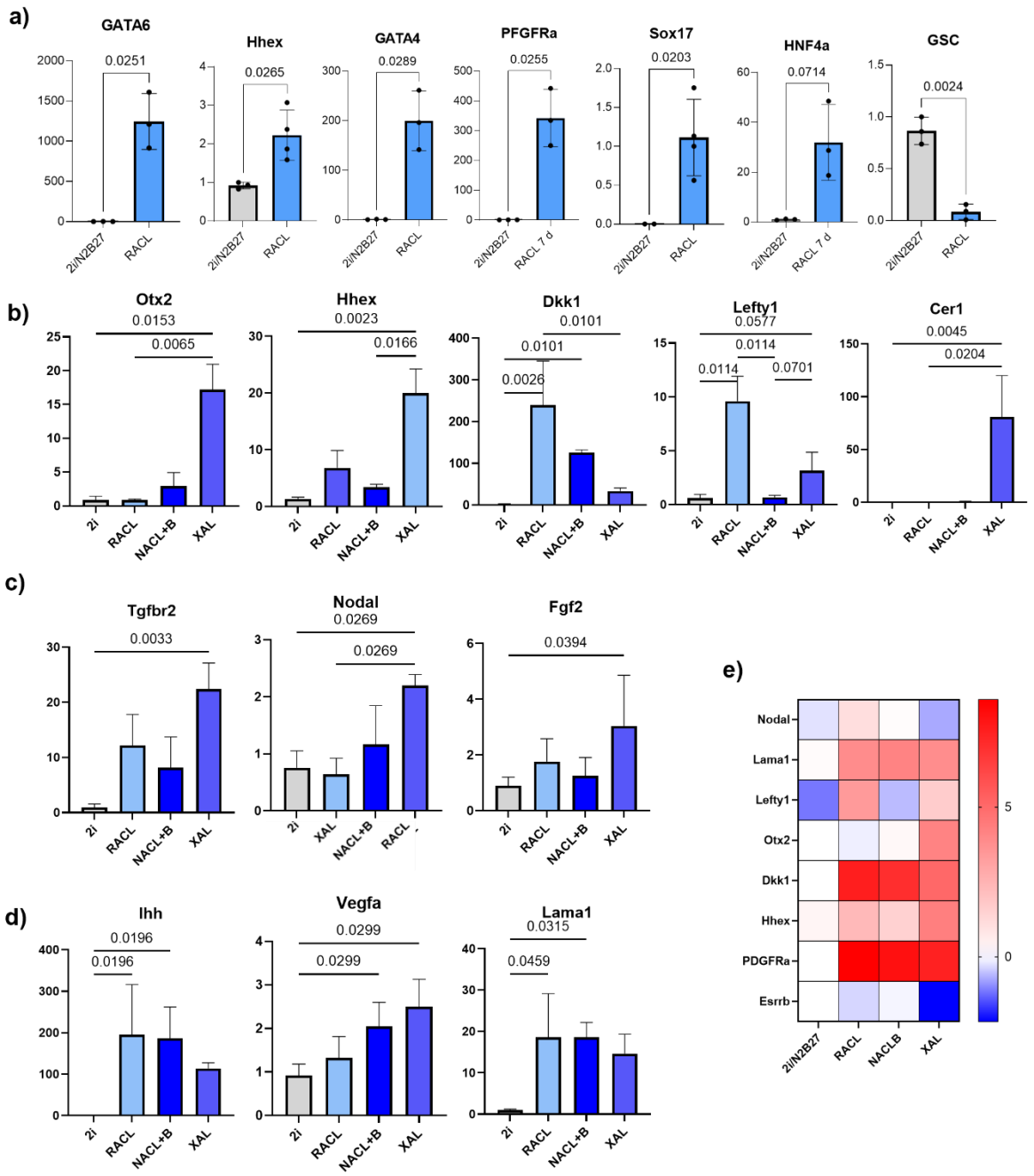
5. Morgani, S.M., and Hadjantonakis, A.K. (2020). Signaling regulation during gastrulation: Insights from mouse embryos and in vitro systems. Curr Top Dev Biol *137*, 391–431. 10.1016/bs.ctdb.2019.11.011.
6. Haantjes, R.R., Strik, J., de Visser, J., Postma, M., van Amerongen, R., and van Boxtel, A.L. (2025). Towards an integrated view and understanding of embryonic signalling during murine gastrulation. Cells Dev, 204028. 10.1016/j.cdev.2025.204028.
7. Turner, D.A., Girgin, M., Alonso-Crisostomo, L., Trivedi, V., Baillie-Johnson, P., Glodowski, C.R., Hayward, P.C., Collignon, J., Gustavsen, C., Serup, P., et al. (2017). Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids. Development *144*, 3894–3906. 10.1242/dev.150391.
8. Dias, A., Pascual-Mas, P., Robertson, G., Torregrosa-Cortés, G., Stelloo, S., Casaní-Galdón, P., Babin, S., Romaniuk, Y., Mayran, A., Wehmeyer, A.E., et al. (2025). Opposing Nodal and Wnt signalling activities govern the emergence of the mammalian body plan. bioRxiv, 2025.2001.2011.632562. 10.1101/2025.01.11.632562.
9. Wehmeyer, A.E., Schmitt, J.K., Eggersdorfer, F., Zissel, L., Schröder, C.M., Tekman, M., Dias, A., Schüle, K.M., Probst, S., Martinez-Arias, A., et al. (2025). Competing regulatory modules control the transition between mammalian gastrulation modes. bioRxiv, 2025.2005.2007.652670. 10.1101/2025.05.07.652670.
10. Nowotschin, S., and Hadjantonakis, A.K. (2020). Guts and gastrulation: Emergence and convergence of endoderm in the mouse embryo. Curr Top Dev Biol *136*, 429–454. 10.1016/bs.ctdb.2019.11.012.
11. Kumar, A., Lualdi, M., Lyozin, G.T., Sharma, P., Loncarek, J., Fu, X.Y., and Kuehn, M.R. (2015). Nodal signaling from the visceral endoderm is required to maintain Nodal gene expression in the epiblast and drive DVE/AVE migration. Dev Biol *400*, 1–9. 10.1016/j.ydbio.2014.12.016.
12. Yoon, Y., Huang, T., Tortelote, G.G., Wakamiya, M., Hadjantonakis, A.K., Behringer, R.R., and Rivera-Perez, J.A. (2015). Extra-embryonic Wnt3 regulates the establishment of the primitive streak in mice. Dev Biol *403*, 80–88. 10.1016/j.ydbio.2015.04.008.
13. Tortelote, G.G., Hernandez-Hernandez, J.M., Quaresma, A.J., Nickerson, J.A., Imbalzano, A.N., and Rivera-Perez, J.A. (2013). Wnt3 function in the epiblast is required for the maintenance but not the initiation of gastrulation in mice. Dev Biol *374*, 164–173. 10.1016/j.ydbio.2012.10.013.
14. Anderson, K.G.V., Hamilton, W.B., Roske, F.V., Azad, A., Knudsen, T.E., Canham, M.A., Forrester, L.M., and Brickman, J.M. (2017). Insulin fine-tunes self-renewal pathways governing naive pluripotency and extra-embryonic endoderm. Nat Cell Biol *19*, 1164–1177. 10.1038/ncb3617.
15. Perera, M., Nissen, S.B., Proks, M., Pozzi, S., Monteiro, R.S., Trusina, A., and Brickman, J.M. (2022). Transcriptional heterogeneity and cell cycle regulation as central determinants of Primitive Endoderm priming. Elife *11*. 10.7554/eLife.78967.
16. Schumacher, S., Fernkorn, M., Marten, M., Chen, R., Kim, Y.S., Bedzhov, I., and Schroter, C. (2024). Tissue-intrinsic beta-catenin signals antagonize Nodal-driven anterior visceral endoderm differentiation. Nat Commun *15*, 5055. 10.1038/s41467-024-49380-0.
17. Srinivasan, R., Nady, N., Arora, N., Hsieh, L.J., Swigut, T., Narlikar, G.J., Wossidlo, M., and Wysocka, J. (2020). Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage. Sci Adv *6*, eaaz9115. 10.1126/sciadv.aaz9115.
18. Weberling, A., Siriwardena, D., Penfold, C., Christodoulou, N., Boroviak, T.E., and Zernicka-Goetz, M. (2025). Primitive to visceral endoderm maturation is essential for mouse epiblast survival beyond implantation. iScience *28*, 111671. 10.1016/j.isci.2024.111671.

19. Artus, J., Douvaras, P., Piliszek, A., Isern, J., Baron, M.H., and Hadjantonakis, A.K. (2012). BMP4 signaling directs primitive endoderm-derived XEN cells to an extraembryonic visceral endoderm identity. Dev Biol *361*, 245–262. 10.1016/j.ydbio.2011.10.015.
20. Thowfeequ, S., Fiorentino, J., Hu, D., Solovey, M., Ruane, S., Whitehead, M., Zhou, F., Godwin, J., Mateo-Otero, Y., Vanhaesebroeck, B., et al. (2024). An integrated approach identifies the molecular underpinnings of murine anterior visceral endoderm migration. Dev Cell *59*, 2347–2363 e2349. 10.1016/j.devcel.2024.05.014.
21. Kretzschmar, J., Goodwin, K., and McDole, K. (2025). Organizer activity in the mouse embryo. Cells Dev, 204001. 10.1016/j.cdev.2025.204001.
22. Hoshino, H., Shioi, G., and Aizawa, S. (2015). AVE protein expression and visceral endoderm cell behavior during anterior-posterior axis formation in mouse embryos: Asymmetry in OTX2 and DKK1 expression. Dev Biol *402*, 175–191. 10.1016/j.ydbio.2015.03.023.
23. Pour, M., Kumar, A.S., Farag, N., Bolondi, A., Kretzmer, H., Walther, M., Wittler, L., Meissner, A., and Nachman, I. (2022). Emergence and patterning dynamics of mouse-definitive endoderm. iScience *25*, 103556. https://doi.org/10.1016/j.isci.2021.103556.
24. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490–495. 10.1038/s41586-019-0933-9.
25. Ragusa, D., Suen, C.-W., Torregrosa-Cortés, G., Pastorino, F., Johns, A., Cicirò, Y., Dijkhuis, L., van den Brink, S., Cilli, M., Byrne, C., et al. (2025). Dissecting infant leukemia developmental origins with a hemogenic gastruloid model. eLife Sciences Publications, Ltd.
26. Schule, K.M., Weckerle, J., Probst, S., Wehmeyer, A.E., Zissel, L., Schroder, C.M., Tekman, M., Kim, G.J., Schlagl, I.M., Sagar, and Arnold, S.J. (2023). Eomes restricts Brachyury functions at the onset of mouse gastrulation. Dev Cell *58*, 1627–1642 e1627. 10.1016/j.devcel.2023.07.023.
27. Yamanaka, Y., Tamplin, O.J., Beckers, A., Gossler, A., and Rossant, J. (2007). Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. Dev Cell *13*, 884–896. 10.1016/j.devcel.2007.10.016.
28. Solovieva, T., Lu, H.C., Moverley, A., Plachta, N., and Stern, C.D. (2022). The embryonic node behaves as an instructive stem cell niche for axial elongation. Proc Natl Acad Sci U S A *119*. 10.1073/pnas.2108935119.
29. Balmer, S., Nowotschin, S., and Hadjantonakis, A.K. (2016). Notochord Morphogenesis in Mice: Current Understanding & Open Questions. Dev Dynam *245*, 547–557. 10.1002/Dvdy.24392.
30. Krauss, S., Concordet, J.P., and Ingham, P.W. (1993). A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. Cell *75*, 1431–1444. 10.1016/0092-8674(93)90628-4.
31. Stemple, D.L. (2005). Structure and function of the notochord: an essential organ for chordate development. Development *132*, 2503–2512. 10.1242/dev.01812.
32. Johnson, R.L., Laufer, E., Riddle, R.D., and Tabin, C. (1994). Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. Cell *79*, 1165–1173. 10.1016/0092-8674(94)90008-6.
33. Fan, C.M., and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. Cell *79*, 1175–1186. 10.1016/0092-8674(94)90009-4.
34. Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and Brachyury

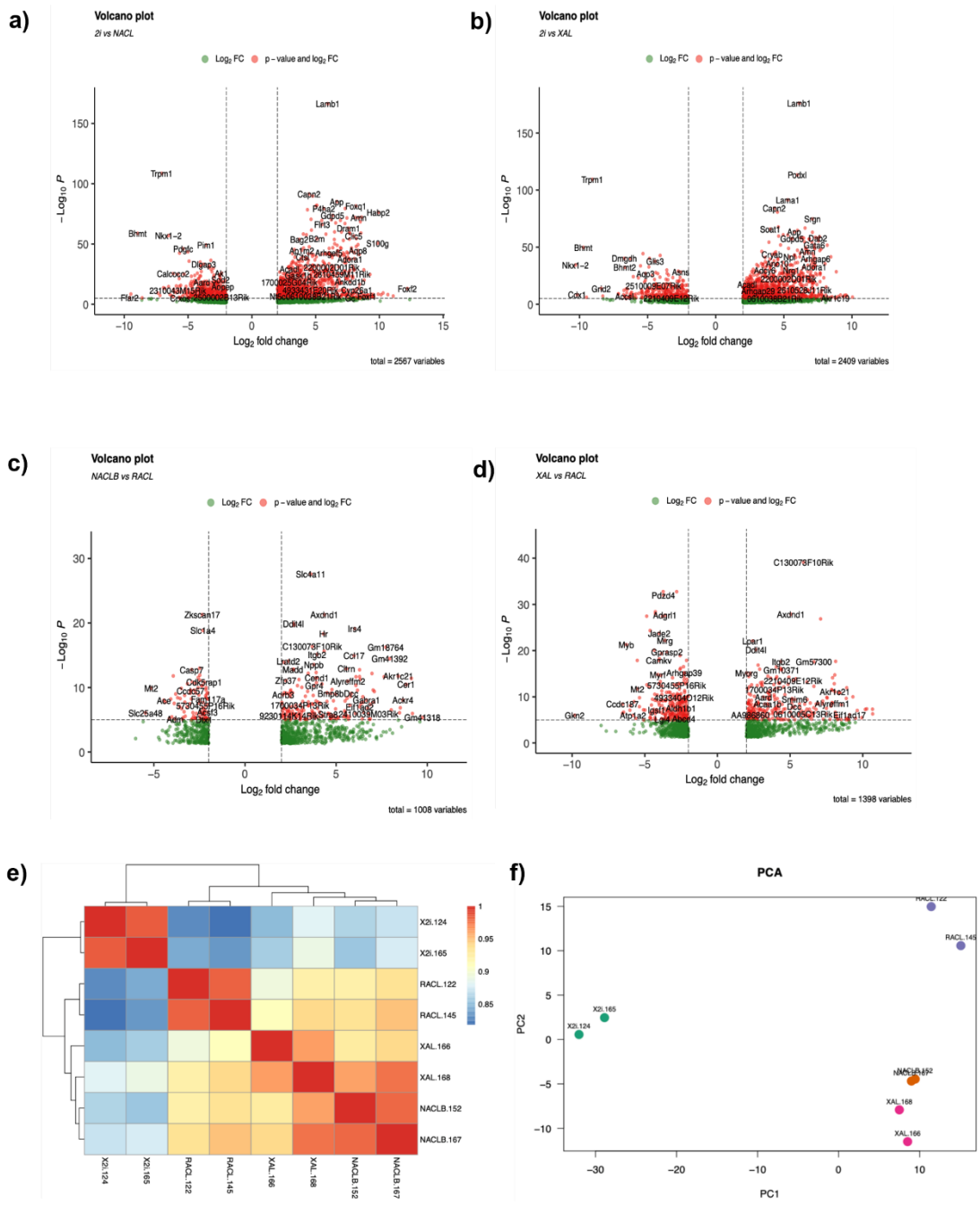
Control the Fate Choice of Neuro-Mesodermal Progenitors. Dev Cell *42*, 514–526 e517. 10.1016/j.devcel.2017.07.021.

35. Binagui-Casas, A., Granés, A., Ceccarelli, A., Karagianni, E., Ramajo, D.L., Portero, R., French, M., Annoh, J., Wong, F.C.K., Lowell, S., et al. (2025). Brachyury expression levels predict lineage potential and axis-forming ability of in vitro derived neuromesodermal progenitors. bioRxiv, 2025.2005.2015.654369. 10.1101/2025.05.15.654369.
36. Rossi, G., Broguiere, N., Miyamoto, M., Boni, A., Guiet, R., Girgin, M., Kelly, R.G., Kwon, C., and Lutolf, M.P. (2021). Capturing Cardiogenesis in Gastruloids. Cell Stem Cell *28*, 230–240 e236. 10.1016/j.stem.2020.10.013.
37. Jakobsson, L., Franco, C.A., Bentley, K., Collins, R.T., Ponsioen, B., Aspalter, I.M., Rosewell, I., Busse, M., Thurston, G., Medvinsky, A., et al. (2010). Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. Nature Cell Biology *12*, 943–953. 10.1038/ncb2103.
38. Ferrer-Vaquer, A., Piliszek, A., Tian, G., Aho, R.J., Dufort, D., and Hadjantonakis, A.K. (2010). A sensitive and bright single-cell resolution live imaging reporter of Wnt/ss-catenin signaling in the mouse. BMC Dev Biol *10*, 121. 10.1186/1471-213X-10-121.
39. Turner, D.A., and Martinez Arias, A. (2024). Three-dimensional stem cell models of mammalian gastrulation. Bioessays *46*, e2400123. 10.1002/bies.202400123.
40. Berenger-Currias, N.M., Mircea, M., Adegeest, E., van den Berg, P.R., Feliksik, M., Hochane, M., Idema, T., Tans, S.J., and Semrau, S. (2022). A gastruloid model of the interaction between embryonic and extra-embryonic cell types. J Tissue Eng *13*, 20417314221103042. 10.1177/20417314221103042.
41. Sozen, B., Amadei, G., Cox, A., Wang, R., Na, E., Czukiewska, S., Chappell, L., Voet, T., Michel, G., Jing, N., et al. (2018). Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. Nat Cell Biol *20*, 979–989. 10.1038/s41556-018-0147-7.
42. Amadei, G., Lau, K.Y.C., De Jonghe, J., Gantner, C.W., Sozen, B., Chan, C., Zhu, M., Kyprianou, C., Hollfelder, F., and Zernicka-Goetz, M. (2021). Inducible Stem-Cell-Derived Embryos Capture Mouse Morphogenetic Events. Dev Cell *56*, 366–+. 10.1016/j.devcel.2020.12.004.
43. Tarazi, S., Aguilera-Castrejon, A., Joubran, C., Ghanem, N., Ashouokhi, S., Roncato, F., Wildschutz, E., Haddad, M., Oldak, B., Gomez-Cesar, E., et al. (2022). Post-gastrulation synthetic embryos generated ex utero from mouse naive ESCs. Cell *185*, 3290–3306 e3225. 10.1016/j.cell.2022.07.028.
44. Dupont, C., Schaffers, O.J.M., Tan, B.F., Merzouk, S., Bindels, E.M., Zwijsen, A., Huylebroeck, D., and Gribnau, J. (2023). Efficient generation of ETX embryoids that recapitulate the entire window of murine egg cylinder development. Sci Adv *9*, eadd2913. 10.1126/sciadv.add2913.
45. Nowotschin, S., Hadjantonakis, A.K., and Campbell, K. (2019). The endoderm: a divergent cell lineage with many commonalities. Development *146*. 10.1242/dev.150920.
46. Levine, A.J., and Brivanlou, A.H. (2007). Proposal of a model of mammalian neural induction. Developmental Biology *308*, 247–256. https://doi.org/10.1016/j.ydbio.2007.05.036.
47. Andoniadou, C.L., and Martinez-Barbera, J.P. (2013). Developmental mechanisms directing early anterior forebrain specification in vertebrates. Cell Mol Life Sci *70*, 3739–3752. 10.1007/s00018-013-1269-5.
48. Yamamoto, M., Saijoh, Y., Perea-Gomez, A., Shawlot, W., Behringer, R.R., Ang, S.L., Hamada, H., and Meno, C. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. Nature *428*, 387–392. 10.1038/nature02418.

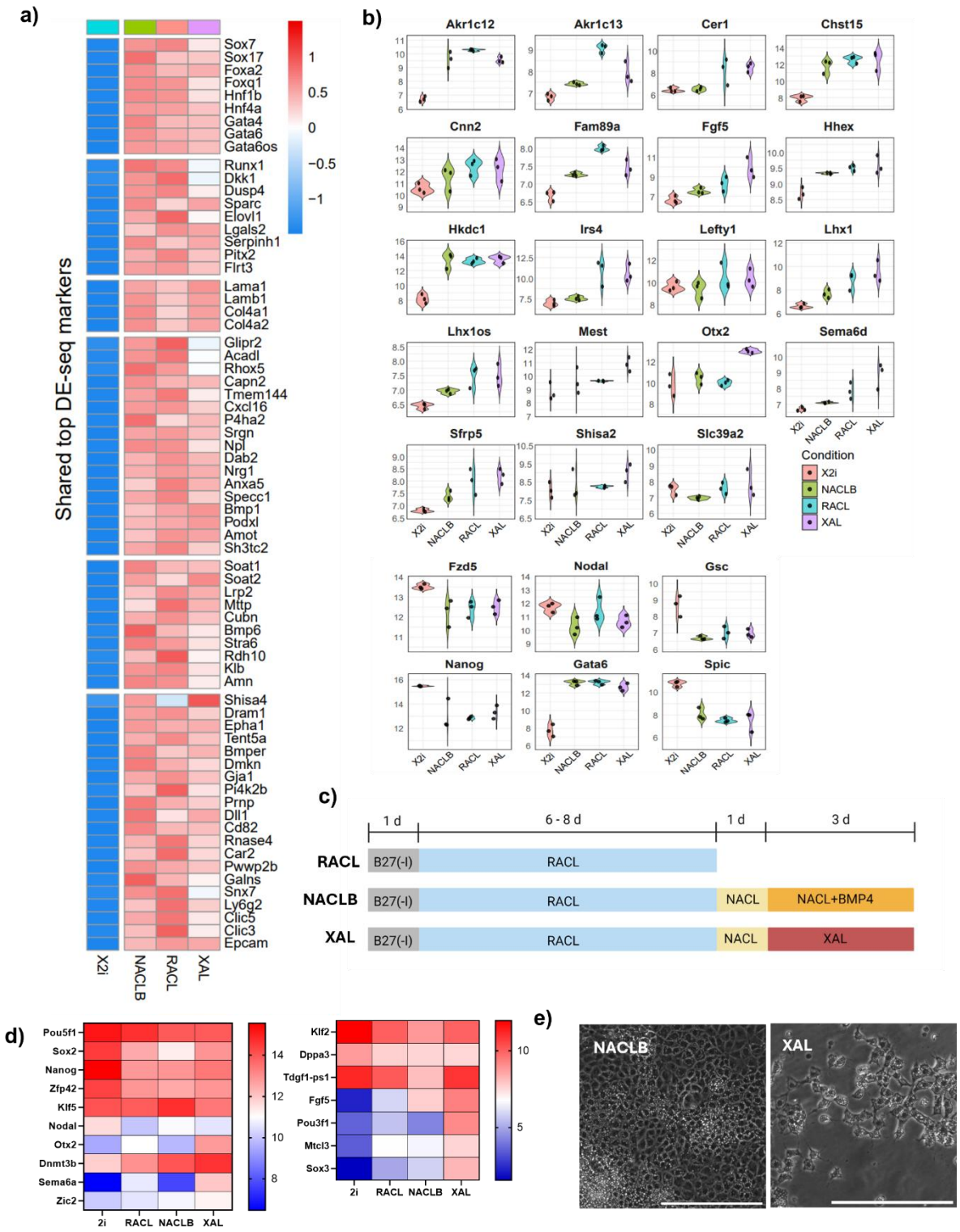
49. Yang, Y.-P., and Klingensmith, J. (2006). Roles of organizer factors and BMP antagonism in mammalian forebrain establishment. Developmental Biology *296*, 458–475. https://doi.org/10.1016/j.ydbio.2006.06.014.
50. Tsiairis, C.D., and McMahon, A.P. (2009). An Hh-dependent pathway in lateral plate mesoderm enables the generation of left/right asymmetry. Curr Biol *19*, 1912–1917. 10.1016/j.cub.2009.09.057.
51. Cohen, E.D., Miller, M.F., Wang, Z., Moon, R.T., and Morrisey, E.E. (2012). Wnt5a and Wnt11 are essential for second heart field progenitor development. Development *139*, 1931–1940. 10.1242/dev.069377.
52. Wymeersch, F.J., Wilson, V., and Tsakiridis, A. (2021). Understanding axial progenitor biology in vivo and in vitro. Development *148*. 10.1242/dev.180612.
53. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676–682. 10.1038/nmeth.2019.



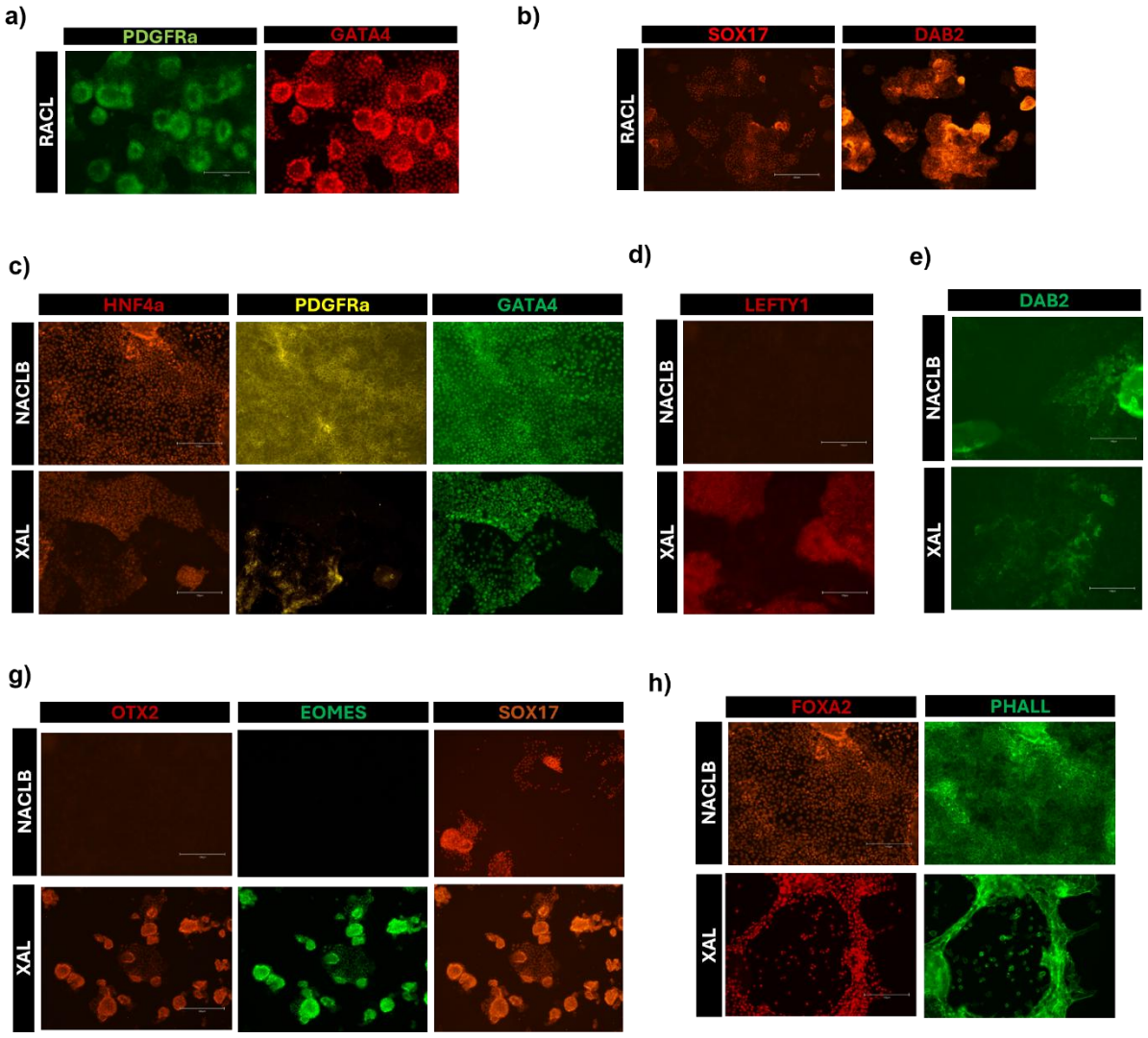
Supplementary Figure 1. Derivation of extraembryonic endoderm-like cells. a) *Relative analysis showing changes in the level of PrE marker gene expression during RACL conversion, normalized to expression levels in 2i cells (non-parametric t-test used). The downregulation of Gsc demonstarated that RACL cells did not exhibit DE identity.* ***b–d)*** *qPCR profiles of 2i, RACL, NACLB, and XAL cells presented on a* $\log_2$ *scale; statistical comparisons performed using Kruskal-Wallis analysis. Cer1 expression was normalized to RACL levels, as its expression in 2i cells was undetectable.* ***e)*** *Heatmap demonstrating the average relative gene expression measured by qPCR and normalized to the 2i sample (n=3, N=3).*



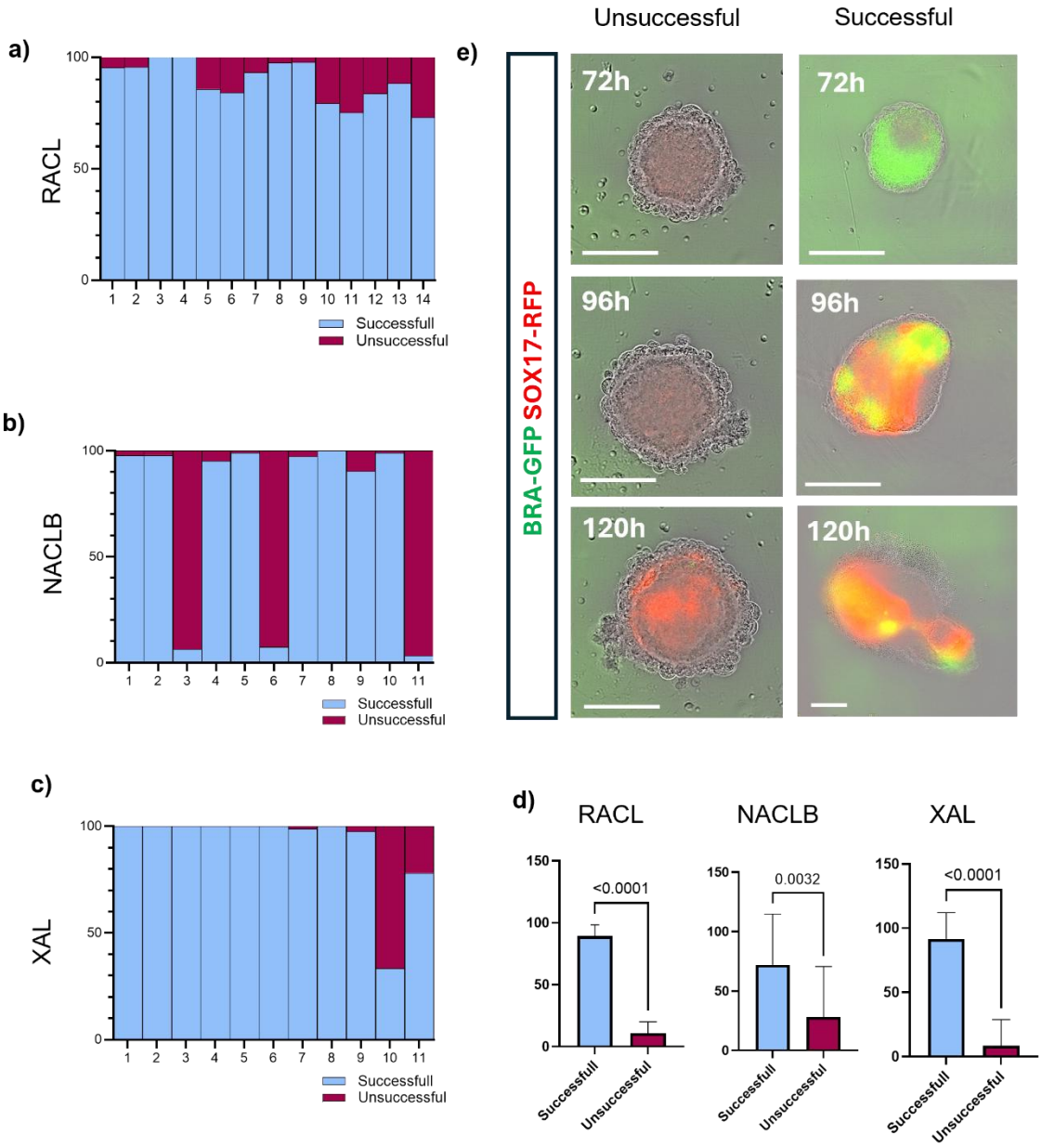
Supplementary figure 2. Obtaining and characterization of ExEnd-like cell types: a)-d) *Volcano plots demonstrating the differences in gene expression between 2i and NACLB, 2i and XAL, NACLB and RACL, XAL and RACL respectively.* ***e)*** *Correlation-based hierarchical clustering of 2i, RACL, NACLB and XAL cells based on bulkRNA sequencing, (n=2). f) Principal component analysis of 2i, RACL, NACLB and XAL cells expression profiles.*



Supplementary 3. Obtaining and characterization of ExEnd-like cell types: a) *Averaged gene expression heatmap of selected markers based on bulk RNA sequencing (n=3).* ***b)*** *Violin plots representing AVE-specific marker gene expression levels according to bulk RNA seq.* ***c)*** *Schematic outline of the protocol used for the derivation of ExEnd-like lineages.* ***d)*** *Heatmap of naïve and primed marker genes expression levels based on reads numbers, bulk RNA sequencing, log2 scale.* ***e)****Phase contrast images of NACLB and XAL cells demonstrating the appearance of lipid droplets and shape.*

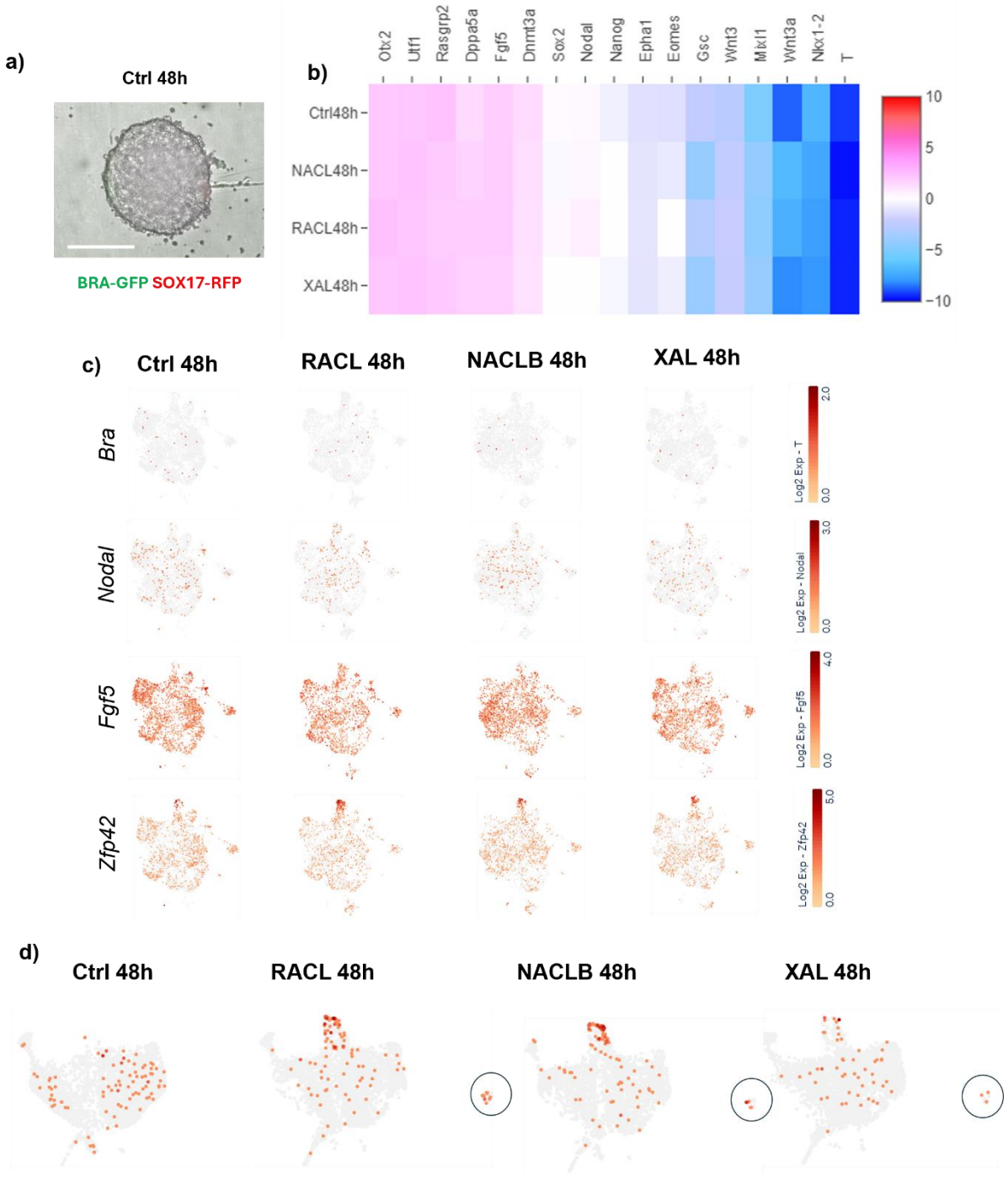


Supplementary 4. Derivation and characterization of ExEnd-like cell types: a) *representative immunofluorescence staining of a 2D culture of RACL cells stained for PDGFRa and GATA4, nuclei stained with DAPI, scale bar = 150 µm;* ***b)*** *Representative immunofluorescence staining of a 2D culture of RACL cells stained for DAPI and DAB2, SOX17-RFP signal from the reporter, scale bar = 300 µm;* ***c)*** *Representative immunofluorescence staining of NACLB and XAL cells in 2D culture with HNF4, PDGFRa and GATA4, scale bar = 150 µm;* ***d)*** *Representative immunofluorescence staining of NACLB and XAL cells in 2D culture with LEFTY1, scale bar = 150 µm;* ***e)*** *Representative immunofluorescence staining of NACL and XAL cells in 2D culture with DAB2, scale bar = 150 µm;* ***f)*** *Representative immunofluorescence staining of NACL and XAL cells in 2D culture with OTX2 and EOMES; the SOX17 signal is from SOX17-RFP reporter, scale bar = 300 µm;* ***g)*** *Representative immunofluorescence staining against FOXA and F-ACTIN (Phalloidin) of NACL and XAL, scale bar = 150 µm.*

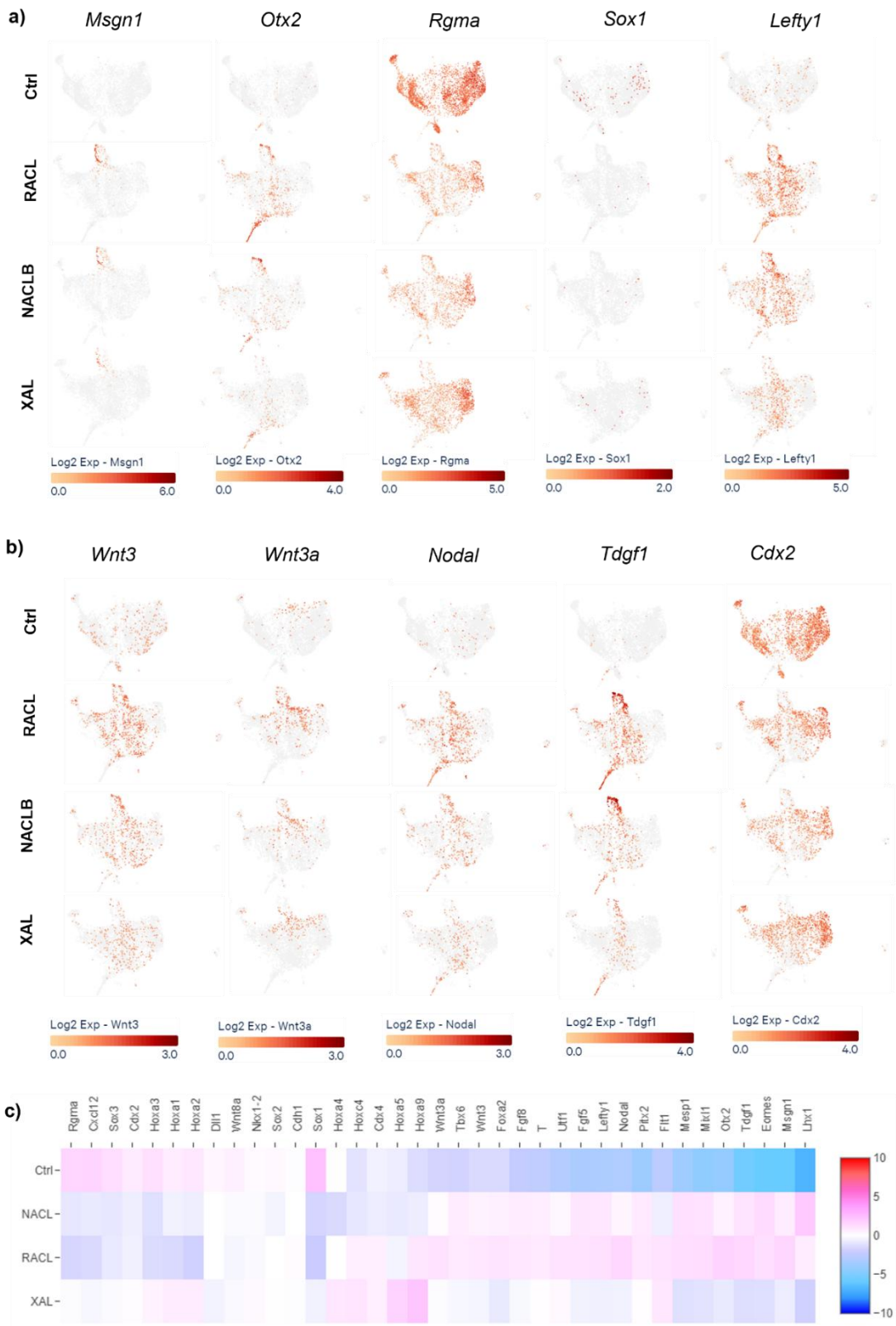


Supplementary 5. Co-aggregation

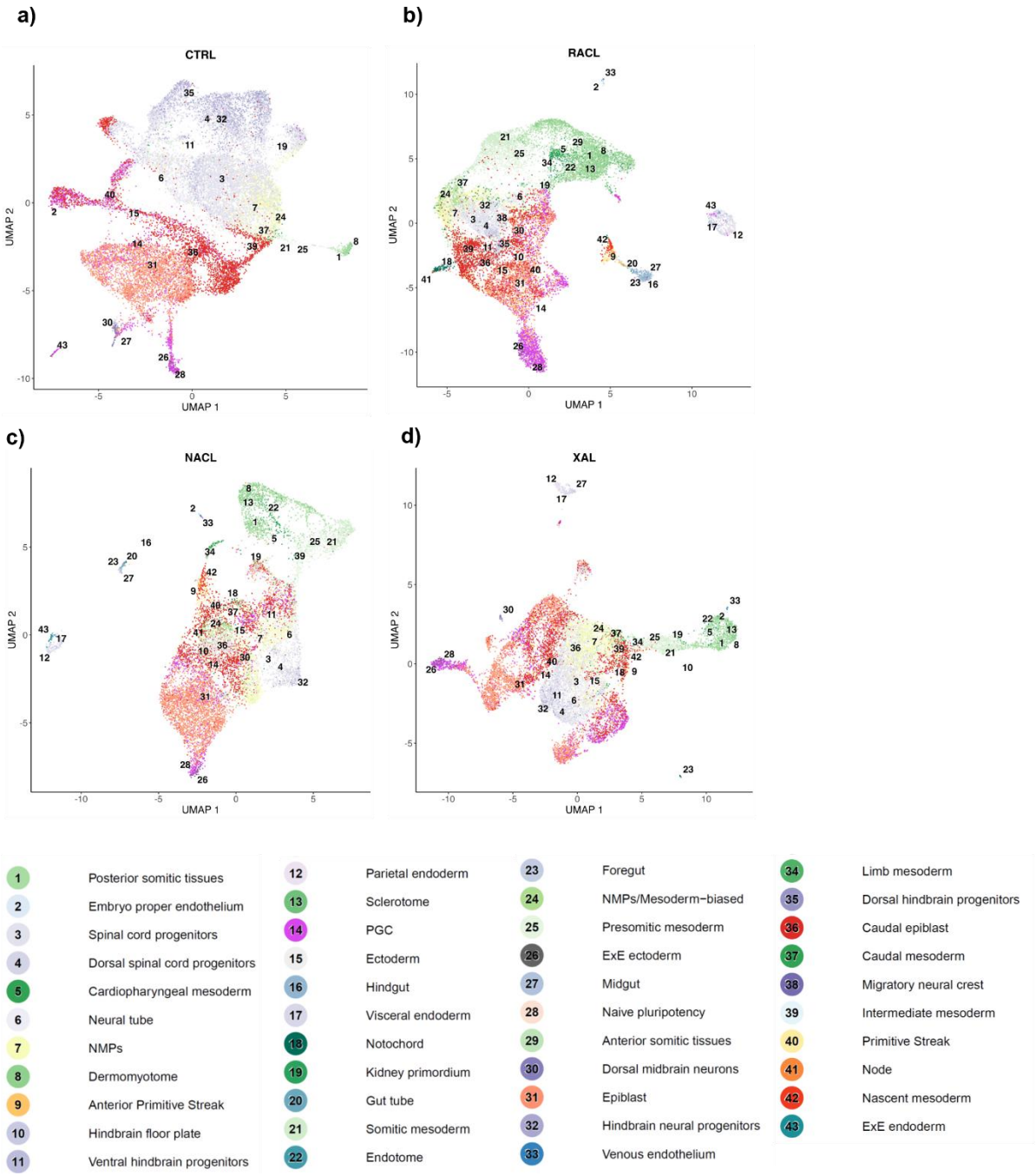
a-c) Bar plots showing the proportions of "successful" and "unsuccessful" structures generated upon co-aggregation of 2i cells with RACL, NACLB, and XAL cells, respectively presented for individual experimental replicates d) Successful structures formation rate presented as mean values ± SEM. Statistical significance was assessed using a non-parametric t-test (Mann–Whitney U test). e) Representative images of "unsuccessful" and "successful " structures. Unsuccessful structures were fully covered with RACL cells. At 72h and 96h no BRA-GFP expression was observed, at 120h cells the aggregate expresses SOX17-RFP showing that despite inhibited BRA expression, cells rather differentiate toward endoderm than stay non-differentiated. Scale bar = 200 µm.



Supplementary 6. Co-aggregation with ExEnd-like cells and the aggregoid development at 48h: a) *Brightfield image of control gastruloid before CHIR pulse (48h). BRA.GFP SOX17-RFP reporter line. Scale bar = 200 µm.* ***b)*** *Heatmap showing the expression levels of selected marker genes in 48h gastruloid, scRNAseq.* ***c)*** *UMAP projection of T/Bra, Nodal, Fgf5 and Zfp42 – expressing cells at 48h in control, RACL, NACLB and XAL conditions, scRNAseq;* ***d)****UMAP projection showing the expression of Dkk1; VE and PE clusers are marked with circle*

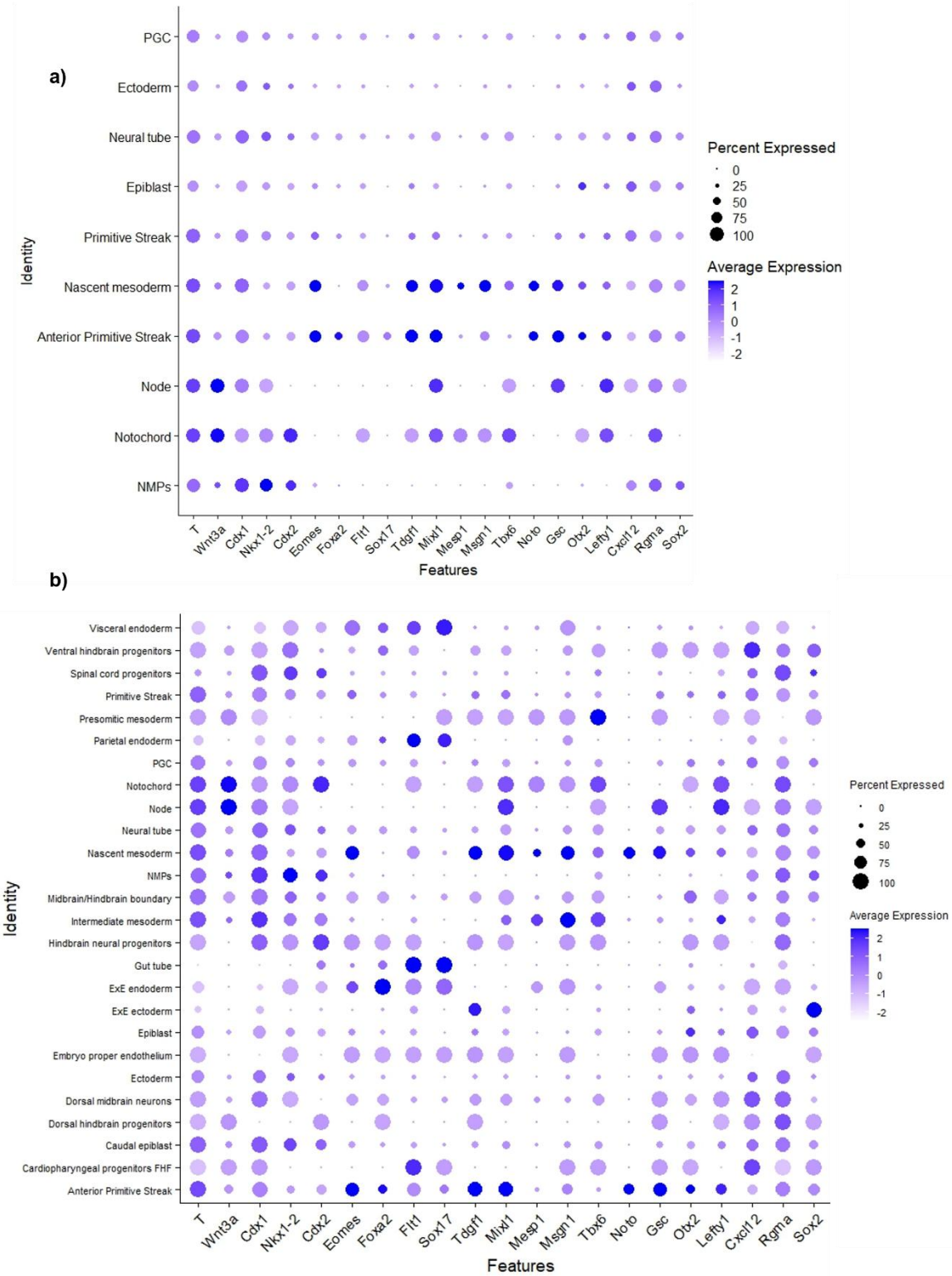


Supplementary 7. Aggregoid expression profile at 72h: a),b) *scRNAseq UMAP projection of selected marker genes at 72h in control, RACL, NACLB and XAL gastruloids;* ***c)*** *Heatmap showing differential gene expression of selected markers ay 72h across four conditions: control, RACL, NACLB and XAL gastruloids.*

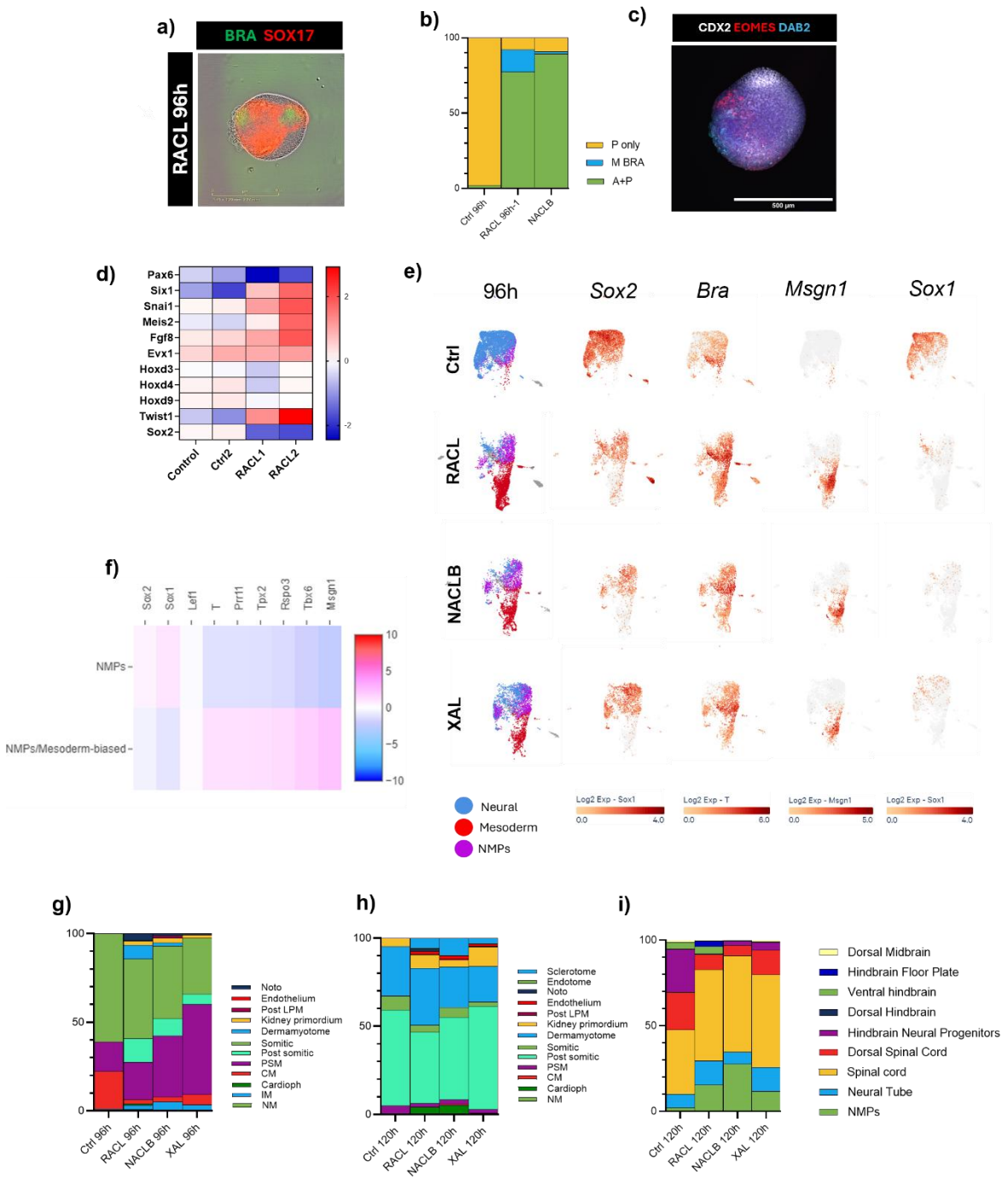


Supplementary 8 a-d) *Integrated UMAP representing predicted cell in control gastruloids (a), RACL (b), NACLB (c) and XAL(d) aggregoids.*

‘



Supplementary 9. Expression profile of RACL-aggregoid: *Dot plot demonstrating expression of selected markers across the predicted id in RACL-aggregoids at 72h.* ***b)*** *Dot plot demonstrating expression of selected markers across the predicted id in RACL-aggregoids at 96h.*



Supplementary 10 a)*Phase-contrast and double-reporter BRA-GFP SOX17-RFP image demonstrating the anterior BRA-GFP domain at 96h, RACL condition. Scale bar = 400 µm.* ***b)*** *Bar plot showing the proportions of gastruloids/aggregoids at 96h having only posterior BRA domain (P only), anterior in addition to posterior (A+P) and also BRA domain appearing in the central part of the structure (M).* ***c)****Confocal image of RACL-aggregoid at 96h, immunostaining for CDX2, EOMES and DAB2****. d)****Heatmap presenting results of qPCR (n=2, N=3) for panel of genes associated with development and patterning of mesoderm and neuroectoderm,* ***e)****UMAP projection of NMPs, meural and mesodermal clusters at 96h and markers of mesodermal and neuroectodermal lineages.* ***g-h)*** *Bar plots representing proportions of mesodermal cell types in control gastruloids, RACL-, NACLB- and XAL-aggregates at 96 and 120h respectively. i)Bar plot presenting proportions of neuroectodermal cell types in control, RACL, NACLB and XAL conditions at 120h.*

Supplementary Movie 1 *Time-lapse fluorescence footage of RACL-aggregoid development from 72h to 96h. Images taken every 15 minutes. BRA-GFP: SOX17-RFP lineage was used for the gastruloid part (2i cells), BFP-ROSA26 cell line was used for RACL.*