MULTI-AXIAL SELF-ORGANISATION PROPERTIES OF MOUSE EMBRYONIC STEM CELLS INTO GASTRULOIDS

Leonardo Beccari[1,*], Naomi Moris[2,*], Mehmet Girgin[3,*], David A. Turner[2], Peter Baillie-Johnson[2,4], Anne-Catherine Cossy[3], Matthias P. Lutolf[3], Denis Duboule[1,3,¶] and Alfonso Martinez Arias[2,¶].

[1]Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland.

[2]Department of Genetics, University of Cambridge, Cambridge, UK.

[3]School of Life Sciences, Federal Institute of Technology EPFL, Lausanne, Switzerland.

[4]Current address: Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, UK.

*Shared first authorship

[¶]Shared last authorship

Correspondence to:

Alfonso Martinez-Arias: ama11@hermes.cam.ac.uk

Denis Duboule:Denis.Duboule@epfl.ch

The emergence of multiple axes is an essential element in the establishment of the mammalian body plan. This process takes place shortly after implantation of the embryo within the uterus and relies on the activity of Gene Regulatory Networks (GRNs) that coordinate transcription in space and time. While genetic approaches have revealed important aspects of these processes[1], a mechanistic understanding is hampered by the poor experimental accessibility of early post-implantation stages. Here we show that small aggregates of murine Embryonic Stem cells (ESC) stimulated to undergo gastrulation-like events and elongation in vitro, are capable of organising a post-occipital pattern of neural, mesodermal and endodermal derivatives that mimic the embryonic spatial and temporal gene expression. The establishment of the three major body axes in such 'gastruloids' [2,3] suggests that the mechanisms involved are interdependent. Specifically, gastruloids display the hallmarks of axial gene regulatory systems as exemplified by the implementation of Hox collinear transcriptional patterns along an extending anterior-posterior axis. These results reveal an unanticipated self-organising capacity for aggregated ESC and suggest that gastruloids may be used as a complementary system to study early developmental events in the mammalian embryo.

Recent work on stem cell derived organoids has revealed a surprising autonomy in the development of particular tissues and organs[4,5]. When *ca.* 250 ESCs are aggregated, given a pulse of the *Wnt* agonist CHIR99021 (Chi) between 48 and 72h of culture, and returned to N2B27 medium (Fig. 1a), a pole of *T/Brachyury* (*Bra*) expression, resembling the elongating embryonic tail bud emerges reproducibly[6] (Fig. 1b, Extended data Fig. 1). The aggregates keep elongating up to 120h after aggregation (AA), when they display a 'rostral' cell-dense region and a polar extension towards a 'caudal' extremity reaching up to 500µm in size (Fig. 1b). Shaking the culture allows to reach 850-1000 micron in length at 168hrs AA (Fig. 1c,d). At these late stages, a *Gata6*-positive domain is detected at the opposite side of a *Bra* and *Cdx2* expressing region, likely corresponding to the cardiac crescent, which delimits the embryonic post-occipital region[7] (Fig. 1b-d, Extended data Fig.1, Supplementary Movies 1 and 2). In contrast, *Sox1*/*Sox2* positive cells localised centrally, with the exception of the rostral-most portion (Fig. 1c, d).

To characterize the transcriptional programmes of these gastruloids, we carried out RNAseq on duplicated pools and compared their profiles with those of developing mouse embryos from E6.5 to E9.5. Since gastruloids display hallmarks of post-occipital embryos[6] (Fig.1b-d) we excluded the anterior portion of E7.5-E9.5 embryos (Fig. 1e, top). Principal Component Analysis (PCA) showed reproducibility between samples and a clear clustering along PC1 corresponding to the temporal

order of samples (Fig. 1e), while embryo samples segregated from gastruloids in PC2 only. The main (top 100) clustering determinants of gastruloid samples included several pluripotency-related genes, epiblast markers and genes involved in gastrulation, as well as *Hox* genes and other transcription factors such as *Cdx1/2*, *Meis1/2*, *Meox1*, *Bra*, and *Gata4* (Figure 2a). These genes are normally expressed in post-occipital structures of the developing mouse embryo. 25 out of 100 of these PCA determinants were identified independently in both gastruloids and embryos temporal series, (Figure 2a, red-labelled genes) supporting the idea that gastruloids and embryos elongate by implementing similar transcriptional programs. The analysis of specific genes associated with particular developmental landmarks confirmed this point (Fig. 2b, Extended data Fig. 2b). For instance genes associated with gastrulation like *Mixl1, Eomes, Gsc* or *Chrd* were transiently and orderly transcribed at around 48h AA (Fig. 2b and Extended data Fig. 2), suggesting that at this stage the gastruloids transcriptome resembles that of mouse epiblast at the onset of gastrulation. By 72h AA, we observed an increase in the complexity of gene expression profiles, with the appearance of markers for different embryonic lineages including mesendoderm and neuroectoderm and the transcription of *Hox* gene clusters (Extended data Fig. 2 and Fig. 2a, b; see below). Genes associated either with extraembryonic structures or with anterior neural plate derivatives were not (or poorly) expressed in gastruloids (Fig. 2b, Supplementary information Files 1 and 2),

PCA analysis using single gastruloids revealed a robust clustering for any developmental stage assessed and with the pooled RNAseq datasets (Extended data Fig. 2c) showing that the population of gastruloids was rather homogenous at the time points analysed and hence that the pooled RNAseq datasets reflected the transcription status of single gastruloids. Gastruloids transcriptome analyses revealed mRNAs normally associated with neural, endodermal and mesodermal derivatives, including paraxial, cardiac, intermediate and hematopoietic progenitors as well as neural crest (e.g.[8,9]) (Fig. 2b; Extended data Figs. 2b and 3). We also observed an antero-posterior pattern of differentiation along these lineages, reminiscent to what occurs in the embryo. For example, the sequential expression of *Bra*, *Msgn1*, *Meox1*/*Tcf15* recapitulates the spatio-temporal differentiation pattern of paraxial mesoderm (Extended data Fig. 3a, b). In neural tissue, while *Hes5* and *Dll1* were strongly expressed during gastruloid extension, the density of terminally differentiating *Phox2a*/*Mnx1* positive cells formed an apparent anterior to posterior gradient, almost completely absent from the posterior-most aspect (Extended data Fig. 3c, d). These ordered patterns of gene expression nevertheless did not correlate with any precise morphogenesis. Neural markers (*Sox1, Sox2*, *Lnfg*) were expressed in a continuous domain, yet without forming a proper neural tube-like structure (Extended data Figs. 3d and 4a-c; Fig. 3) even though sporadic tubular structures were

scored along this domain (Extended data Fig. 4a, white arrowheads, c, right panel). Also, clumps of cells positive for either SOX1 and OLIG2, SOX1 and PAX3 or SOX1 and PAX7, indicative of dorsal and ventral neural tube progenitors, were observed (Extended data Fig. 4d), yet without a clear segregation along the dorso-ventral extension of the SOX1 domain. Similarly, *Tcf15* expressing cells did not condense into somites (Extended data Fig. 3b).

The analysis of different endodermal markers revealed temporal dynamics also reminiscent of the embryonic situation[10] (Extended data Figs. 3e, f and 4e-f). *Gsc* and *Cdx2* transcripts, markers of definitive endoderm[11,12] were upregulated soon (72h AA) after Chiron induction, followed by *Cer1* (96-120h AA) and by *Sorcs2*, *Pax9* or *Shh* subsequently (120h-144h AA). All endoderm expressed genes assayed (*Sox17*, *Sorcs2*, *Nedd9*, *Pyy*, *Cdx2* and *Shh*) were active in the ventral-like domain of gastruloids (Fig. 3, Extended data Figs. 3f, 4e-f), resembling the embryo situation. Of note, *Cdx2* transcripts were confined to the posterior most gastruloid endoderm (Fig. 1d, Extended data Fig. 4f), in agreement with this gene specifying the hindgut domain. In a majority of cases, gut-endoderm progenitors appeared as a continuous tubular structure (Extended data Fig. 4a, e-f; red arrowheads), often spanning the entire antero-posterior extension, reminiscent of an embryonic digestive tract.

This unanticipated level of organization and capacity to self-organize an integrated axial system reminiscent of the embryo was further explored by assessing the expression of genes associated with the developing embryonic axes (Fig. 3). *Wnt3a* and *Cyp26a1* transcripts were scored at the caudal extremity of gastruloids similar to *Bra* and *Cdx2* (Fig. 1c, d; Extended data Figs. 3a, d and 4a, d, Extended data Fig. 5a), complementary to the localization of *Raldh2* mRNAs further supporting the existence of an antero-posterior axis. On the other hand, *Lnfg, Sox1 and Sox2* were transcribed in a central and dorsal domain at 144h AA (Figs. 1b, 3b and Extended data Figs. 3a and 4a), complementary to the ventrally located intestinal tract markers (Figs. 3b, Extended data Figs. 3d and 4c, d, Extended data Fig. 5b). Additional signs of multi-axial organisation were provided by the expression of mesoderm specific genes *Osr1*, *Pecam*, *Meox1* and *Pax2* in a medio-lateral symmetry flanking the centrally located *Sox2* positive domain (Fig. 3c). Double staining of *Sox2* and *Meox1* (Fig. 3c, Extended data Fig. 5c) and cross-sections (Extended data Fig. 4b) confirmed the medio-lateral and dorso-ventral distribution of neural and mesodermal progenitors.

Nodal expression was found confined to a small and compact region on the ventral most posterior aspect at 120h AA (Extended data Fig. 6 and 7). These cells displayed high levels of E-cadherin and a dense phalloidin staining (Extended data Fig. 6a, b) suggestive of a node-like identity, a hypothesis supported by the presence of *Nodal* mRNAs in a domain comparable to that of

Goosecoid, Bra and *Chordin* at 96h AA (Extended data Fig. 6c,d). Time course analysis of *Chrd* and *Nodal* indicated that such putative node-like cells were detected at 96h and persisted until 144h at least (Extended data Fig. 6d). *Nodal* mRNAs in these cells nevertheless rapidly decreased and were almost undetectable at 144h AA. Despite these evidences for a node-like structure, we did not observe any notochord derivatives, which normally originate from the node, raising questions as to whether this putative structure may excert an organizing activity reminiscent to that of its embryonic counterpart. The downregulation of *Nodal* in presumptive node-like cells at 120h AA coincided with patches of *Nodal*-expressing cells along the posterior half of extending gastruloids, often distributed in an asymmetric manner (Fig. 3d, Extended data Fig 6d, e). This pattern was maintained at 144h AA (extended data Fig. 6d) but was not observed with *Meox1*, which was predominantly expressed on both sides (Fig. 3d, e). Accordingly, the *Nodal* target gene *Cerberus* was also expressed asymmetrically at both 120h AA and 144h AA (Extended data Fig. 6f). Altogether these data suggest that besides the establishment of a medio-lateral axis, gastruloids may implement the start of a left-right asymmetry,

The formation and patterning of post-occipital embryonic territories is tightly linked to the sequential activation of the 39 *Hox* gene, which are clustered at four distinct genomic loci in mammals. As *Hox* genes appeared differentially regulated in the RNAseq time-course (Fig. 2a, b, Extended data Fig. 2a), we assessed whether their sequential activation in time and space [13] was recapitulated too. A pooled PCA analysis considering exclusively *Hox* genes transcripts, revealed robust clustering along the time axis (81% variance) and a close correspondence with the dynamic activation of these genes *in embryo* (Fig. 4a, Extended data Fig. 8a-c). The variability in *Hox* RNAs content amongst gastruloids was then evaluated using ten individual specimens from three different stages (Extended data Fig. 9a). Gastruloids from the same time-point tightly clustered together solely based on their *Hox* transcripts. Transcript profiles over *Hox* clusters revealed signs of collinear activation, the hallmark of this gene family[14]. In E6.5 embryo, some *Hoxa* and *Hoxd* genes are expressed before gastrulation in extraembryonic tissues[15] (Extended data Fig. 8a). From E7.8 to E9.5, *Hox* genes start to be transcribed in an order which reflect their 3' to 5' position within each cluster (Extended data Fig. 8a, b). The RNAseq profiling revealed an activation dynamic comparable to that observed *in embryo* (Fig. 2a and Extended data Fig. 8c). For instance while *Hoxa* RNAs were not detected until 48h AA, *Hoxa1* to *Hoxa3* expressions were robust at 72h, followed by sustained transcription of *Hoxa5*, *Hoxa7* and *Hoxa9* at 96h to 120h. *Hoxa10* and *Hoxa11* RNAs appeared at 144h AA, at the same time *Hoxa1*, *Hoxa2* and *Hoxa3* transcripts started to disappear (Fig. 4b, Extended data Fig. 8c). Similar dynamics were observed for *Hoxd* genes (Extended data Fig. 8c), which were activated in a sequence starting from 72h AA until 168h AA (Extended data Fig. 8c-e).

The early transcription of 5' *Hoxa*/*Hoxd* genes (Extended data Fig. 8a-b) was not observed in gastruloids (Extended data Fig. 9c, d and Fig. 4b), in agreement with the absence of extraembryonic derivatives.

Comparable profiles were also scored when single organoids were examined (Extended data Fig. 9a, b), again revealing the high reproducibility of this activation process. In the embryo this temporal activation is paralleled by a collinear distribution of transcripts in space[14]. Likewise, *Hoxa4*/*Hoxd4* displayed an AP boundary close to the anterior aspect of the gastruloid, whereas *Hoxa9/Hoxd9*, *Hoxa11/Hoxd11* and *Hoxd13* showed successively more posterior boundaries (Fig. 4c, Extended data Fig. 9c). Notably, *Hoxd13* transcripts appeared in a population of cells located centrally at the posterior extremity, resembling its normal expression in the embryonic cloacal area (Fig. 4c). *Hoxa13* expression was also detected at 168h AA in the posterior aspect, yet rarely (1/20), in agreement with the low transcript levels detected in the pooled RNAseq analysis (Extended data Fig. 9c). Double staining for *Hoxd4* and either *Sox2* or *Meox1* revealed expression of *Hox* genes in both neural and mesodermal derivatives (Fig. 4d, Extended data Fig. 5d, e). The implementation in space and time of the *Hox* gene network confirmed the surprisingly high level of organisation in the processing of gene regulatory networks, in particular without any extraembryonic component (see ref.[16]).

We tested the ability of several induced pluripotent stem cells (iPSC) lines to produce gastruloids (Extended data Fig. 10) and a similar elongation process was observed in one of them. iPSCs can thus generate gastruloids. However, these gastruloids showed a reduced elongation rate, particularly between 48h and 96h (extended data Fig. 10a, b). The expression dynamics of *Bra* in these iPSC gastruloids was nevertheless similar to their ESC counterparts (Extended data Fig. 11c, d). The neural markers *Sox1* and *Sox2* as well as *Cdx2* were also expressed as in ESC-derived gastruloids (Extended data Fig. 10d, compare with Fig. 1b, c). Furthermore, iPSC gastruloids implemented *Hoxd* temporal and spatial collinear expressions, though with a delay in the expression onset and a spatial collinearity not as clearly organized as in ESC-derived gastruloids (Extended data Fig. 9d, e).

When compared to single tissue organoids (e.g.[17,18]), gastruloids exhibit an integrated structure, which seems to specify all major embryonic axes in a coordinated manner, thus complementing recent reports where stem cells were used to recapitulate morphological and transcriptional events of early blastocyst, yet not of subsequent embryonic stages[19,16]. However, the activation of tissue specific patterns of gene expression in our gastruloids was not paralleled by a clear internal organization of the corresponding embryonic tissue layers. This observation suggests

that a direct causal relationship between transcriptional programs and early morphogenesis ought to be considered with caution. One potential reason for the low level of tissue organization in gastruloids may be the absence of mechanical interactions and constraints that characterize the developing embryonic context[20]. The remarkable autonomy in the patterns of gene expression reported here highlights the potential of gastruloids in the study of complex regulatory circuits, particularly during early post-implantation development and the emergence of body axes.

Acknowledgements

We thank members of the Duboule, Lutolf and Martinez-Arias laboratories for sharing material and discussions, J. Deschamps for DNA clones and D. Trono for the *Oct4::GFP* iPS cell line. We also thank the histology platform and the Gene Expression Core Facility (EPFL) as well as the genomics platform in Geneva University. This work was supported by funds from the BBSRC (No. BB/M023370/1 and BB/P003184/1 to AMA), an NC3Rs David Sainsbury Fellowship (No. NC/P001467/1 to D.A.T.), an Engineering and Physical Sciences Research Council (EPSRC) Studentship (to P.B.-J.), a Company of Biologists Development Travelling Fellowship (DEVTF-151210 to P.B.-J.), a Newnham College Constance Work Junior Research Fellowship (to NM), the Ecole Polytechnique Fédérale de Lausanne (D.D. and M.L.), the University of Geneva (D.D.), the Swiss National Research Fund (No. 310030B_138662 to D.D.) and the European Research Council grants ERC MOFDH (No. 250316 to A.M.A.), Regul*Hox* (No 588029 to D.D.) and STEMCHIP (No 311422 to M.L.).

Authors contributions:

Conceived the study: AMA, DD, LB, MG, ML
Conducted experiments:, ACC, DT, LB, MG, MM, PBJ
Analyzed and interpret datasets: DD, NM, LB, MG, PBJ.
Wrote the Manuscript: AMA, DD, LB.
Corrected/Amended the MS: MG, ML, NM.

Legends to Figures

Figure 1: Elongation of gastruloids. a. Schematic of the culture protocol. 200 to 300 ESCs were allowed to aggregate. The *Wnt* agonist CHIR99201 (Chi) was added between 48h and 72h after aggregation. Organoids were kept in suspension until 120h (grey rectangle) and transferred into shaking cultures until 168h. **b**. Three dimensional renderings and confocal sections of gastruloids at different times showing the elongation and expression of BRA, SOX2 and *Gata6*$^{H2B\text{-}Venus}$ (green). **c-d**. Three-dimensional rendering (**c**, left panel) and confocal sections (**c**, central and right panels and **d,** zoom of tail region) of gastruloids at 168h, showing the localization of CDX2, SOX2, SOX1 and BRA proteins. Scale bar: 150 μm. For each time point analyzed, the results reported were scored in at least 80% of the cases ($n \geq 20$). **e.** PCA analysis of RNAseq datasets using time-pooled gastruloids from 24h to 168h (2 replicates per time-point) and pooled mouse embryos at E6,5 (3 replicates), E7.8 (3 replicates), E8.5 (12-14 somites, 2 replicates) and E9.5 (ca. 24 somites, 2 replicates). For E7.8 embryos, the posterior half was used. For E8.5 and E9.5, the post-occipital embryonic domain was used. The dissected portion is colored in pale green. All autosomal genes were considered for this analysis. Principal Component 1 (PC1) shows a strong temporal component while PC2 discriminates between gastruloids or embryonic samples.

Figure 2. Temporal patterns of gene expression in gastruloids. a. PCA of either pooled gastruloids during temporal progression from 24h to 168h (left), or murine embryos from E6.5 to E9.5 (right). The 100 top contributing genes to the first two principal components are overlaid, with those observed in both gastruloids and embryonic datasets shown in red text. **b**. Heatmap of scaled expression of genes associated with development of different embryonic structures in pooled gastruloids and embryos over time.

Figure 3: Multi-axial organization of gastruloids. a-c. Gene expression in gastruloids at 144h AA showing their axial organization. **a**. *Wnt3a* and *Cyp26a1* expression (arrowhead) at the posterior end, where *Raldh2* is not transcribed (empty arrowhead). Double FISH staining of *Meox1* and *Cyp26a1* (**a**, right-most panel) showing antero-posterior segregation of mesodermal precursors. **b**. Dorso-ventral (D-V) axis revealed by the ventral expression of *Shh* and *Krt18*, and of *Lnfg* dorsally (empty arrowheads). Double FISH staining of *Sox2* and *Shh* confirmed a dorso-ventral segregation, with *Shh* expressed exclusively in endoderm precursors **(b**, righ panel). **C.** Medio-lateral (M-L) axis of symmetry (dotted line) revealed by the bilateral expression of *Meox1* and *Pax2,* complementary to the central distribution of *Sox2* transcripts (empty arrowheads). For each gene, the proportion of gastruloids displaying the reported pattern is shown. Scale bar: 100 μm. **d**. 3D renderings of confocal

stacks of 120h gastruloids containing a *Nodal*^{YFP} reporter gene, stained for SOX2 (white) and BRA (red) proteins and imaged from the dorsal (left) and ventral (right) orientations; insets details of the posterior region. Reporter gene expression within the *Bra* expressing domain on the ventral surface is suggestive of a node-like structure (middle panel; Extended Figure 6). Additional expression of *Nodal* as a bilaterally asymmetric cluster of cells (white open arrow) is reminiscent of the asymmetric *Nodal* expression in the embryo (middle panel). Right panel shows a posterior view of the 3D rendering. **e**. Bar graph showing the frequency distribution of asymmetric and symmetric expression of *Nodal* or *Meox1* in 120h gastruloids (see Extended data Table 1 for more details). ### (observed versus expected frequency, based on the embryonic gene expression pattern): p<0,0001; **(observed versus expected frequency in asymmetric *Nodal* expression based on the frequency of *Meox1* asymmetry in gastruloids).

Figure 4: **Collinear *Hox* gene expression in gastruloids. a.** PCA plot solely based on *Hox* transcripts datasets extracted from pooled gastruloid and embryonic data across time points. Replicate batches of organoids primarily cluster according to their age at collection. **b.** Transcript profiles over the *HoxA* cluster, using time-sequenced pooled gastruloids. A progressive wave of transcription through *Hoxa* genes is observed between the 72h and 168h time-points. **c**. *In situ* hybridizations of 168h gastruloids using probes for various *Hoxd* genes. Expression becomes spatially restricted along the A-P axis along with the respective position of the genes in the cluster. For each gene, the proportion of gastruloids displaying the reported expression pattern is shown in the bottom right corner of the image, expressed as a fraction of the total number of gastruloids analyzed. Scale bar: 100μm. **d**. Double FISH staining of *Hoxd4* with *Sox2* or *Meox1* (respectively marking the neural and mesodermal precursors) showed that *Hoxd4* expression colocalized with both markers, suggesting that gastruloid implement both neural and mesodermal *Hoxd* gene expression. Scale bar: 200μm.

Materials and Methods

ES/iPS cells and gastruloid cultures: The culture conditions and a detailed protocol for ES/iPS cells culturing and gastruloid production can be found in Supplementary material and method and in the associated Protocol Exchange article.

Animal experimentation: wild-type CD1 mouse embryos were used for RNAseq experiments. All experiments were performed in agreement with the Swiss law on animal protection (LPA) under license number GE 81/14 (to D. Duboule).

RNA extraction and RNAseq libraries preparation and sequencing can be found in supplementary information.

Libraries and qPCR analysis: Purified RNA from iPS cell derived gastruloids was retrotranscribed using the Promega GoScript retrotranscription kit. Quantitative PCR analysis of mRNA levels for different *Hoxd* genes, *Bra* and the housekeeping gene *Hmbs* was performed using the Syber select master mix for CFX (Thermofisher) kit according to manufacturer instruction and specific primers[6,21]. The Biorad CFX96 thermocycler was used. At least two technical (PCR) replicates and two biological replicates were analyzed per time-point after aggregation.

Probe cloning, *in vitro* transcription and *in situ* hybridization. Refer to Supplementary material and methods and to Extended data Tables 3 and 4.

Immunostaining and confocal microscopy. Refer to Supplementary material and methods and to Extended data Table 5.

Data availability statement. All RNAseq datasets produced in this study are publicly available in the Gene Expression Omnibus (GEO) database under #GSE10622. All the scripts used for the analyzis of the RNAseq data are freely available upon request.

Extended data figure legends

Extended data Figure 1. a-c. Gastruloids produced using *Gata6*$^{H2B\text{-}Venus}$ mESCs treated with a pulse of the GSK3 inhibitor Chiron between 48h and 72h AA and fixed either at 48hh (**a**), 72 (b), 96h (**c**) or 120h (**d**) and imaged by confocal microscopy. BRA and SOX2 proteins are stained in red and white, respectively. VENUS signal (green) reports *Gata6* expression and Hoechst (Blue) marks the nuclei. Gastruloids corresponding to the 3D renderings shown in **Fig. 1a**. Each fluorescent channel is displayed to the right of each merged image. *Gata6* or *Gata6* and SOX2 signals were undetectable in **a** and **b**, respectively, and therefore not shown. Three z-sections are shown for each gastruloid. The bright field outline of each gastruloid is indicated by the dashed lines. Scale-bars as indicated.

Extended data Figure 2. a. Heat map showing the temporal evolution of 97 out of the 250 most variable genes throughout embryonic development from E6.5 to E9.5 (left) and their corresponding expression over the gastruloid time-course, from 24h to 168h (right). Expression levels are highlighted by color scale from blue to red (bottom left). Genes were clustered according to their

expression behaviour in the embryo and enriched GO term categories were identified for each cluster by using the Gorilla and Revigo tools (see supplementary material and methods and Supplementary file 1). Finally, a functional classification of each cluster was established based on the identified GO term categories and literature-based evidences. **b.** Expression of markers for different embryonic tissues through the gastruloid time-course. The two replicates of each time point are represented by triangles and circles. The black dotted line in each plot represents the average behaviour of the different genes displayed in the plot. **c**. PCA analysis of RNAseq datasets from either pooled or individual gastruloids using the top 1000 most highly expressed genes. Despite different strategies used for pooled *versus* individual gastruloids RNAseq (accounting for the sample segregation across PC1), the clustering of pooled and single-gastruloid datasets illustrates both the homogeneity of gastruloid cultures and the representativeness of pooled samples to single gastruloid samples.

Extended data Figure 3: Gastruloids display spatio-temporal organization in the expression profiles of neural, mesodermal and endodermal marker genes. a-f. The expression profiles of several genes normally expressed in the embryonic neural, mesodermal and endodermal domains was analyzed by plotting the RNAseq data in pooled gastruloids in heatmaps of scaled gene expression (**a**, **c**, **e**) and/or by WISH (**b**, **d**, **f**). **a-b**. Expression of different neural markers was detected in our RNAseq. Genes like *Lnfg* or *Irx3* were detected forming continuous and homogenous domains located in the central and dorsal portion of the gastruloids, reminiscent of their expression domains in the embryo (**b**, upper panels). Instead, genes involved in *notch* signaling in neural progenitors (*Hes5*, *Dll1*) and in the terminal differentiation of neural precursor (*Phox2a*, *Mnx1*) displayed a salt and pepper expression pattern, consistent with the lack of an organized neural tube structure (Extended data Figs. 4-5). However, the latter mRNAs also displayed a graded distribution along the anterior to posterior extension of the gastruloid axis and were absent from its posterior half (empty red arrowheads**). c-d.** Genes normally expressed in different types of mesoderm precursors in the embryo (e.g. *Tcf15* in paraxial somatic mesoderm, *Osr1* in intermediate mesoderm, *Bra* in notochord and PSM and *Pecam* in lateral plate mesoderm) were expressed in reproducible and spatially restricted domains within the gastruloids. **e-f**. Endoderm specific genes were also expressed in gastruloids. Particularly, genes expressed in the embryonic digestive tract were consistently found on the ventral side of gastruloids. For each gene, the proportion of gastruloids displaying the reported expression pattern is shown in the upper right corner of the image, expressed as a fraction of the total number. Scale bar: 100μm.

Extended data Figure 4. a. Gastruloids formed from *Sox1*^{GFP} (green); B*ra*^{mCherry} (red) mESCs were fixed at 168h AA and stained for SOX2. White arrowheads indicate tubular SOX2/SOX1 positive neural structures. Red arrowheads point to the presumptive digestive tube. **b**. Gastruloids at 144h were WISHed for *Sox2* and *Meox1* antisense probes, cryo-sectioned in 8 μm thick transversal cross-sections and counter-stained with Nuclear Fast Red. *Sox2* positive cells localized predominately in a compact dorsal domain, whereas *Meox1* signals was found in two bilateral domains. The domain of expression of each gene is delimited with white dashed lines. **c**. Haematoxylin-Eosin staining of transversal paraffin sections of different gastruloids at 120h AA, showing the cell type diversity and several degrees of tissue organization. **d**. Gastruloids formed from *Sox1*^{GFP} (green); *Bra*^{mCherry} (red) mESCs were fixed and stained at 168h for OLIG2 (top-panel, white), PAX3 (mid-panel, red) and PAX7 (lower-panel, red). Scale bars as indicated. **c, d**. Gastruloids formed from *Sox1*^{GFP} (green); *Bra*^{mCherry} (red) mESCs collected at 168h AA and stained for SOX17 (magenta in **c**) or CDX2 (magenta in **d**). Scale bars as indicated.

Extended data Figure 5. a-e. Double FISH staining of gastruloids from *Sox1*^{GFP} (green); B*ra*^{mCherry} (red) mESCs at 144h with *Meox1* and *Cyp26a1* (**a**), *Sox2* and *Shh* (**b**), *Sox2* and *Meox1* (**c**), *Meox1* and *Hoxd4* (**d**) or *Sox2* and *Hoxd4* (**e**). Scale bar: 200μm

Extended data Figure 6. a-b. Gastruloids formed from *Nodal*^{YFP} mESCs were fixed at 120h AA. They were stained for CDX2, YFP (*Nodal*^{YFP}) and E-Cadherin **(a**, top panel), CDX2, YFP (*Nodal*^{YFP}; green) and Phalloidin (**a**, bottom panel) or CDX2, YFP and E-CADHERIN (both with an Alexa-488 secondary antibody) and SOX2 (**b**). Maximum intensity projection of a representative gastruloid in **b,** with the node-like structure highlighted. Hoechst marks the nuclei (greyscale in **a**, blue in **b**). **c-d**. *In situ* hybridizations of gastruloids at different time-points AA. Asterisk in **d** mark the presumptive node-like cells. White arrowheads point towards *Nodal* expressing cells distributed asymmetrically, on the lateral side of the gastruloid. **e**. Three dimensional renderings of confocal stacks of 120h gastruloids containing a *Nodal*^{YFP} reporter gene (green) and stained for SOX2 (white) and BRA (red) proteins. SOX2 signal identifies dorsal cells. Left and right panels show the same gastruloid, imaged from two different polar directions i.e. top (dorsal) and bottom (ventral) or 'left' and 'right' depending on the orientation of the gastruloid. Insets in specific panels show a cross-section through the gastruloid at the indicated *z* plane. White arrowheads indicate the region of biased *Nodal* expression. Empty white arrowheads point to the node-like cells marked by the *Nodal*^{YFP} reporter gene. See also **Fig. 4d** for and Extended data Table 1. **f**. *In situ* hybridizations of gastruloids 120h (left) and 144h AA (right). The midline of the gastruloid is marked by a dashed white line. White

arrowheads point towards *Nodal* expressing cells distributed asymetrically on the lateral side of the gastruloid. In **c**, **d**, and **f**, the proportion of gastruloids displaying the reported expression pattern is shown in the bottom left corner of each image, expressed as a fraction of the total number.

Extended data Figure 7. a, b. Dorsal (**a**) and ventral (**b**) sections of the same representative gastruloid shown in the 3D renderings in **Fig. 4d**, fixed at 72h AA and stained at 120h for $Nodal^{YFP}$ (green), BRA (red) and SOX2 (white). Hoechst was used to mark the nuclei. Scale bar indicates 100μm.

Extended data Figure 8. **a**. Heatmap of unscaled gene expression in E6.5-E9.5 mouse embryos, showing *Hox* gene transcript levels over time. **b**. RNAseq mapping showing *Hoxa* and *Hoxd* gene expression in these embryos. After a first wave of transcription of 5' *Hoxa* and *Hoxd* genes, likely reflecting their activation in extraembryonic tissues, the *HoxA* and *HoxD* clusters were progressively transcribed from E7.8 until E9.5 when expression of *Hox13* paralogs was detected. **c**. Heatmap of unscaled gene expression in pooled gastruloids, showing *Hox* gene transcript levels over time. **d.** RNAseq mapping showing *Hoxd* gene expression in pooled gastruloids at different time points. *Hoxd* genes sub-groups are progressively activated starting at 72h until 168h AA, when expression of *Hoxd13* starts to be detected (**e**), thus resembling the temporal activation described *in vivo* (**a, b**). **e**. Whole mount *in situ* hybridization of gastruloids collected at different time points and showing the detectable initiation of different *Hoxd* genes expression. Each panel report the earliest stage where transcripts of the corresponding gene were detected (black arrowhead). Expression of *Hoxd4* was already strong at 96h AA indicating that its transcripts are rapidly upregulated compared to *Hoxd9* which is faintly expressed at this stage. Scale bar: 100μm.

Extended data Figure 9. a. Principal component analysis (PCA) based on *Hox* transcripts datasets only, extracted from individually sequenced gastruloids across time points (10 individual organoids per time point). The analysis was carried out using the Log_2 transformed FPKM+1 value of all 39 *Hox* genes. Replicate batches of organoids primarily cluster according to their age at collection. The clustering revealed the low sample-to-sample variation. Instead, replicates were clearly separated through the temporal parameter, representing 93.6% of total sample variation. **b**. Comparison of *Hoxa* (top panel) and *Hoxd* (bottom panel) gene expression profiles amongst individual gastruloids confirmed the low inter-sample variation among time-points, illustrated with the 120h condition. **c**. Whole mount *in situ* hybridization of 168h AA gastruloids showing the expression of different *Hoxa*

paralogs. The proportion of gastruloids displaying the reported expression pattern is shown in the upper right corner of the image, expressed as a fraction of the total number. Scale bar: 100μm.

Extended data Figure 10. a. Dot plot representing the progression in the measured longitudinal extension of gastruloids produced either from ES or from iPS cells. **b**. Light microscopy images showing representative examples of gastruloids at the different time-points analysed in (**a**). Zoom: 10x. Note that iPS derived gastruloid display delay in their longitudinal extension rate and at 120h AA, they are significantly smaller than their ESC-derived counterparts. For this analysis, gastruloids were produced starting from the same number of cells (800 cells per well). **c**. Dot plots representing the mRNA levels of *Bra,* showing comparable dynamics of this gene in both types of gastruloids. **d**. Confocal images showing the expression of *Oct4*, SOX2 and BRA (upper panel) or of *Oct4*, SOX1 and CDX2 (bottom panel) in 120h gastruloids derived from the iPS cell line Oct4::Gfp (IpSL40N). iPS-derived gastruloids were fixed at 120h AA and stained for SOX2-BRA (left) and CDX2-SOX1 (right). *Oct4::GFP* signal is shown in greys. Scale bars: 200μm. **e**. Dot plots representing the mRNA levels of *Hoxd* genes in ES or iPS cell-derived gastruloids collected at different time-points AA. Both gastruloids sequentially activated *Hoxd* gene expression. However, their temporal activation seemed to be delayed in iPS gastruloids (especially that of the most 3' *Hoxd* paralogs). **e**. Whole mount *in situ* hybridization of 144h AA gastruloids, showing the expression of different *Hoxd* paralogs. Even though iPS derived gastruloids reproduced the antero-posterior *Hoxd* collinear expression, *Hoxd9* expression domain often extended more anteriorly when compared to ESs cell derived gastruloids (see Fig. 5e), occupying approximately the same domain than *Hoxd4*. Also patches of *Hoxd* negative cells were often found within the *Hoxd4*/*Hoxd9* expression domain (white asterisks). Scale bar: 100μm.

Supplemental Movies 1 and 2 (corresponding to Fig. 1a 96h and 120h AA). a. Gastruloids were produced from $Gata6^{H2B\text{-}Venus}$ mESCs treated with Chiron between 48h and 72h AA, fixed and stained at either 96h AA (**movie 1**) or 120h (**movie 2**) for *Gata6* (green), BRA (red) and SOX2 (white) and imaged by confocal microscopy as indicated in Materials and Methods. Each fluorescent channel is highlighted in turn for clarity. Hoechst was used to stain the nuclei.

Supplemental Movies 3 and 4 (for Fig. 3d). Gastruloids were produced from $Nodal^{YFP}$ mESCs and treated with Chiron between 48h and 72h AA, fixed and stained at 120h for YFP (green), BRA (red) and SOX2 (white). Views from the dorsal (**movie 3**) and ventral (**movie 4**) aspects are shown. The Node-like region and bilateral expression of *Nodal* is highlighted in the ventral region (**movie 4**).

References

1. Rossant, J. & Joyner, A. L. Towards a molecular-genetic analysis of mammalian development. *Trends Genet* **5,** 277–83 (1989).
2. Simunovic, M. & Brivanlou, A. H. Embryoids, organoids and gastruloids: new approaches to understanding embryogenesis. *Development* **144,** 976–985 (2017).
3. van den Brink, S. C. *et al.* Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141,** 4231–42 (2014).
4. Sasai, Y., Eiraku, M. & Suga, H. In vitro organogenesis in three dimensions: self-organising stem cells. *Development* **139,** 4111–21 (2012).
5. Turner, D. A., Baillie-Johnson, P. & Martinez Arias, A. Organoids and the genetically encoded self-assembly of embryonic stem cells. *Bioessays* **38,** 181–91 (2016).
6. Turner, D. A. *et al.* Anteroposterior polarity and elongation in the absence of extraembryonic tissues and spatially localised signalling in Gastruloids, mammalian embryonic organoids. *Development* (2017). doi:10.1242/dev.150391
7. Zhao, R. *et al.* Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev Biol* **317,** 614–9 (2008).
8. Kurokawa, D. *et al.* Regulation of Otx2 expression and its functions in mouse forebrain and midbrain. *Development* **131,** 3319–31 (2004).
9. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–63 (2005).
10. Wang, P. *et al.* A molecular signature for purified definitive endoderm guides differentiation and isolation of endoderm from mouse and human embryonic stem cells. *Stem Cells Dev.* **21,** 2273–2287 (2012).

11. Beck, F., Erler, T., Russell, A. & James, R. Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn* **204,** 219–27 (1995).

12. Blum, M. *et al.* Gastrulation in the mouse: the role of the homeobox gene goosecoid. *Cell* **69,** 1097–1106 (1992).

13. Deschamps, J. & van Nes, J. Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* **132,** 2931–42 (2005).

14. Deschamps, J. & Duboule, D. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev* **31,** 1406–1416 (2017).

15. Scotti, M. & Kmita, M. Recruitment of 5' Hoxa genes in the allantois is essential for proper extra-embryonic function in placental mammals. *Development* **139,** 731–9 (2012).

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

17. Kadoshima, T. *et al.* Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc Natl Acad Sci U A* **110,** 20284–9 (2013).

18. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501,** 373–9 (2013).

19. Rivron, N. C. *et al.* Blastocyst-like structures generated solely from stem cells. *Nature* **557,** 106–111 (2018).

20. Heisenberg, C. P. & Bellaiche, Y. Forces in tissue morphogenesis and patterning. *Cell* **153,** 948–62 (2013).

21. Guerreiro, I. *et al.* Reorganisation of Hoxd regulatory landscapes during the evolution of a snake-like body plan. *Elife* **5,** (2016).

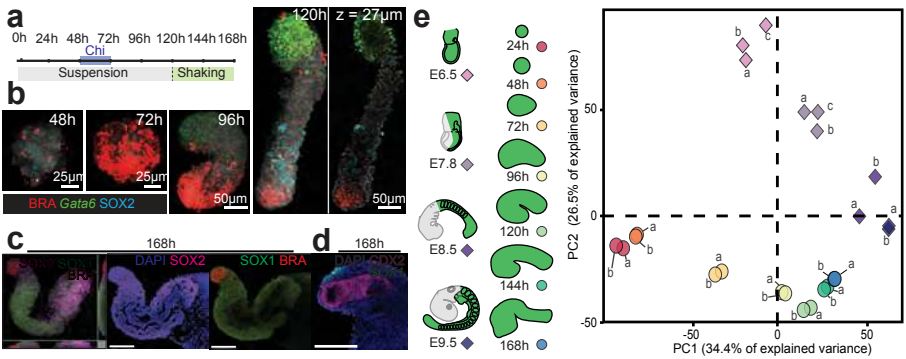
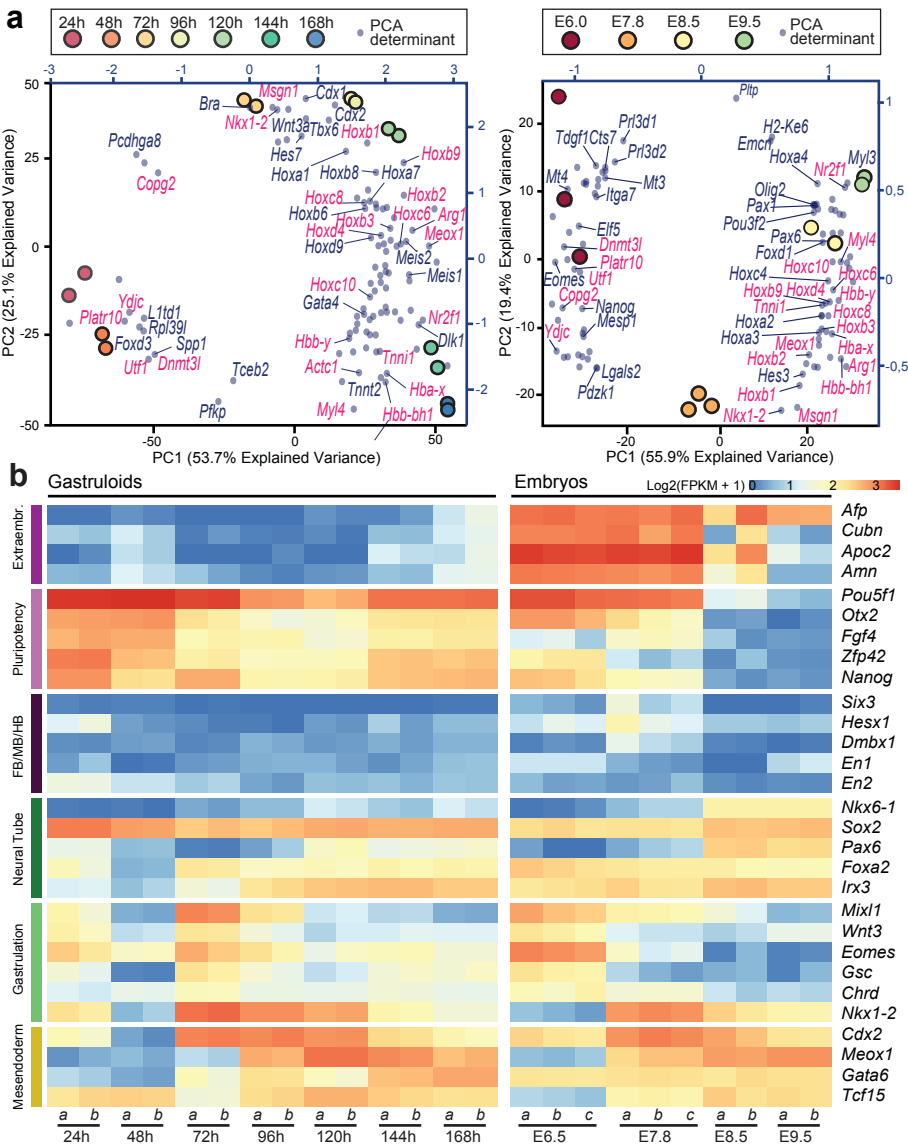

a
0h 24h 48h 72h 96h 120h 144h 168h
Chi
Suspension
Shaking
120h z = 27µm
50µm
b
48h
72h
96h
25µm
25µm
50µm
BRA Gata6 SOX2
c
168h
DAPI SOX2
SOX1 BRA
d
168h
DAPI CDX2
e
E6.5
E7.8
E8.5
E9.5
24h
48h
72h
96h
120h
144h
168h
PC2 (26.5% of explained variance)
PC1 (34.4% of explained variance)
50
0
-50

Figure 1


a
24h 48h 72h 96h 120h 144h 168h
PCA determinant
PC1 (53.7% Explained Variance)
PC2 (25.1% Explained Variance)
Msgn1
Cdx1
Bra
Cdx2
Nkx1-2
Wnt3a
Tbx6
Hoxb1
Pcdhga8
Hes7
Hoxb9
Hoxa1
Hoxb8
Hoxa7
Copg2
Hoxc8
Hoxb2
Hoxb6
Hoxb3
Hoxc6
Arg1
Hoxd4
Meox1
Hoxd9
Meis2
Meis1
Hoxc10
Gata4
Ydjc
Platr10
L1td1
Rpl39l
Foxd3
Spp1
Utf1
Dnmt3l
Nr2f1
Dlk1
Hbb-y
Tceb2
Actc1
Tnni1
Tnnt2
Hba-x
Pfkp
Myl4
Hbb-bh1
E6.0 E7.8 E8.5 E9.5
PCA determinant
PC1 (55.9% Explained Variance)
PC2 (19.4% Explained Variance)
Pltp
Prl3d1
Tdgf1
Cts7
Prl3d2
H2-Ke6
Emcn
Hoxa4
Nr2f1
Myl3
Mt4
Mt3
Itga7
Olig2
Pax1
Pou3f2
Elf5
Pax6
Myl4
Dnmt3l
Foxd1
Platr10
Hoxc10
Hoxc4
Hoxc6
Eomes
Utf1
Hoxb9
Hoxd4
Hbb-y
Copg2
Nanog
Tnni1
Mesp1
Hoxc8
Hoxa2
Hoxb3
Ydjc
Hoxa3
Meox1
Hba-x
Hoxb2
Arg1
Lgals2
Hes3
Hbb-bh1
Pdzk1
Hoxb1
Nkx1-2
Msgn1

b
Gastruloids
Embryos
Log2(FPKM + 1)
Extraembr.
Pluripotency
FB/MB/HB
Neural Tube
Gastrulation
Mesendoderm
Afp
Cubn
Apoc2
Amn
Pou5f1
Otx2
Fgf4
Zfp42
Nanog
Six3
Hesx1
Dmbx1
En1
En2
Nkx6-1
Sox2
Pax6
Foxa2
Irx3
Mixl1
Wnt3
Eomes
Gsc
Chrd
Nkx1-2
Cdx2
Meox1
Gata6
Tcf15
a b
24h
48h
72h
96h
120h
144h
168h
a b c
E6.5
E7.8
E8.5
E9.5

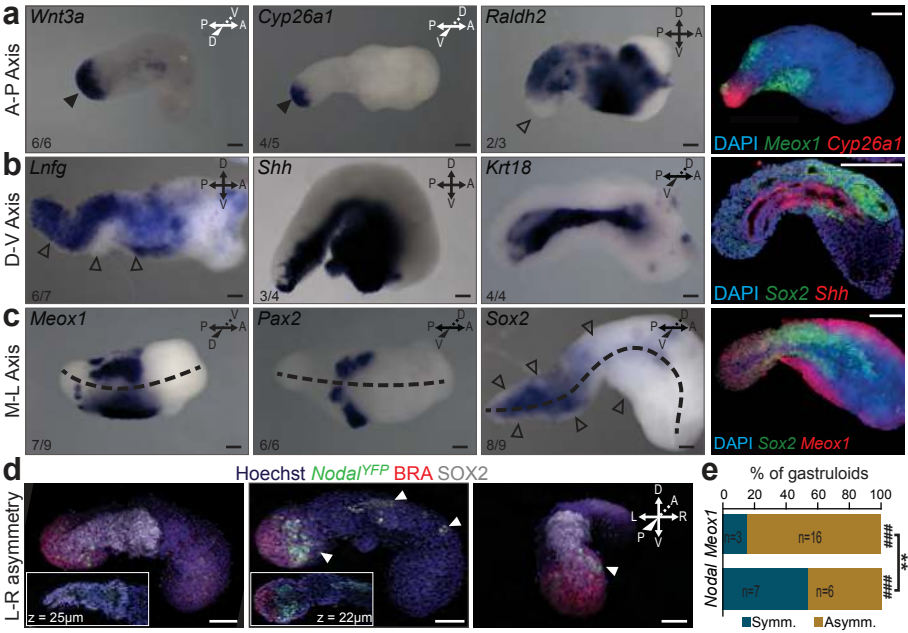

a
A-P Axis
Wnt3a
Cyp26a1
Raldh2
6/6
4/5
2/3
DAPI Meox1 Cyp26a1
b
D-V Axis
Lnfg
Shh
Krt18
6/7
3/4
4/4
DAPI Sox2 Shh
c
M-L Axis
Meox1
Pax2
Sox2
7/9
6/6
8/9
DAPI Sox2 Meox1
d
L-R asymmetry
Hoechst NodalYFP BRA SOX2
z = 25µm
z = 22µm
e
% of gastruloids
0 20 40 60 80 100
Nodal Meox1
n=3
n=16
n=7
n=6
###
**
###
Symm.
Asymm.

Figure 3

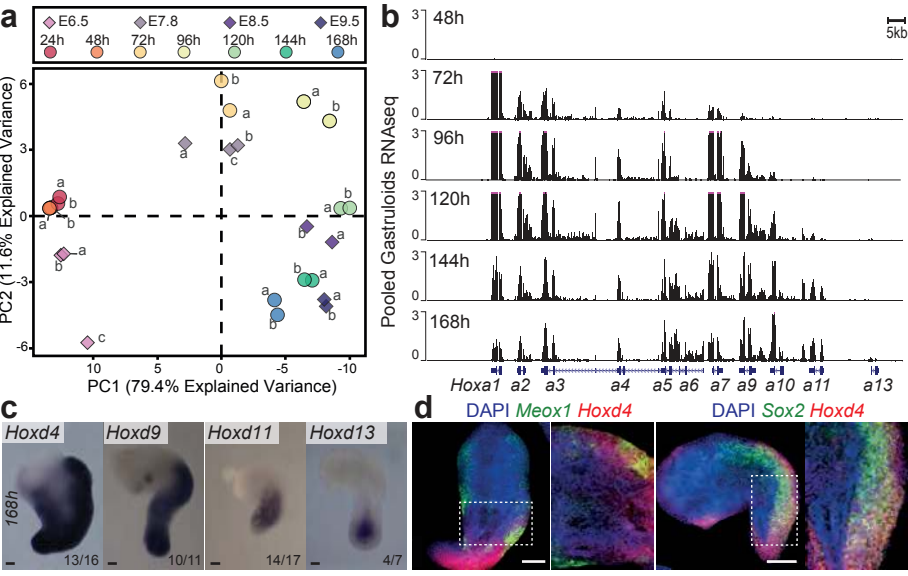
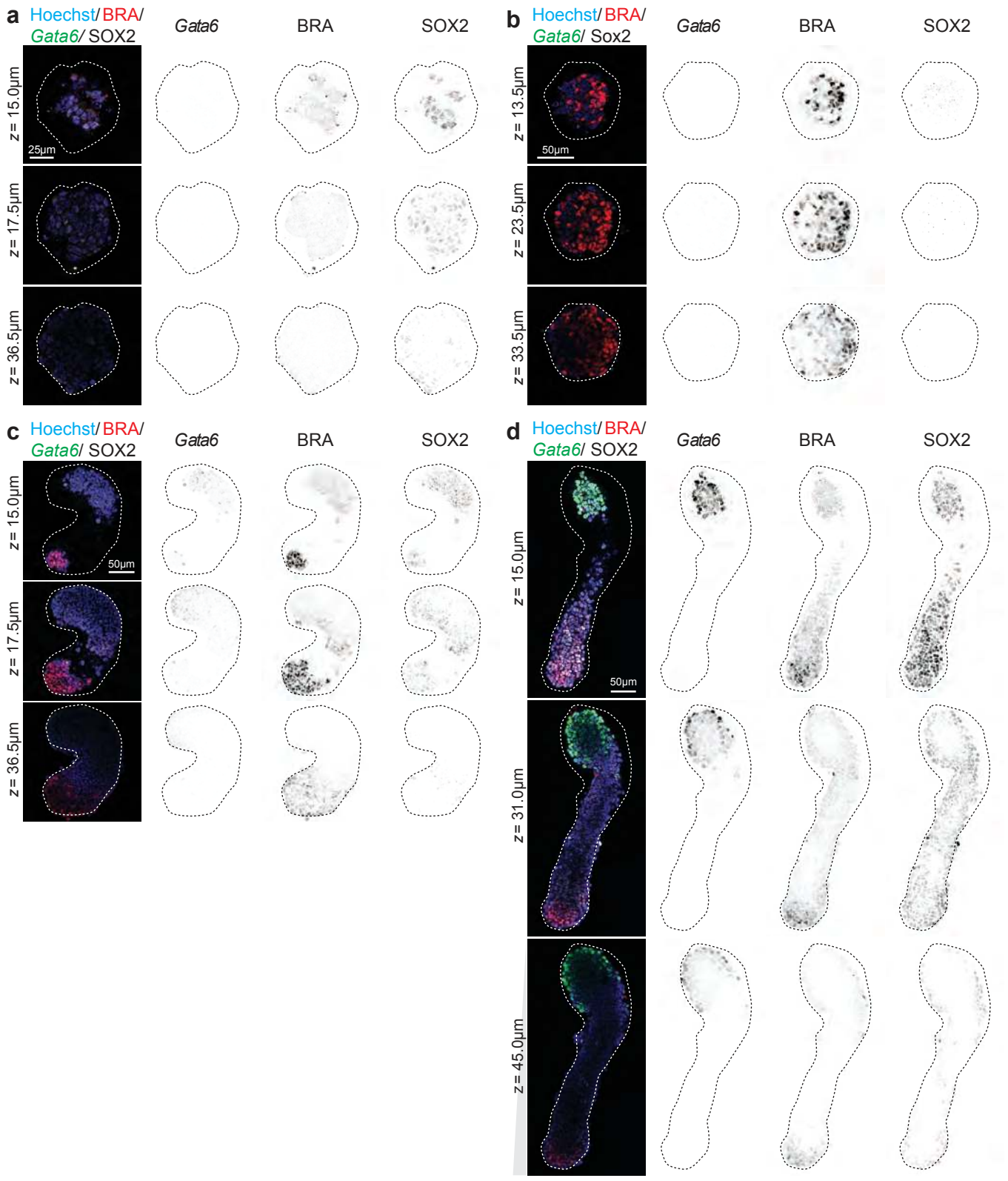
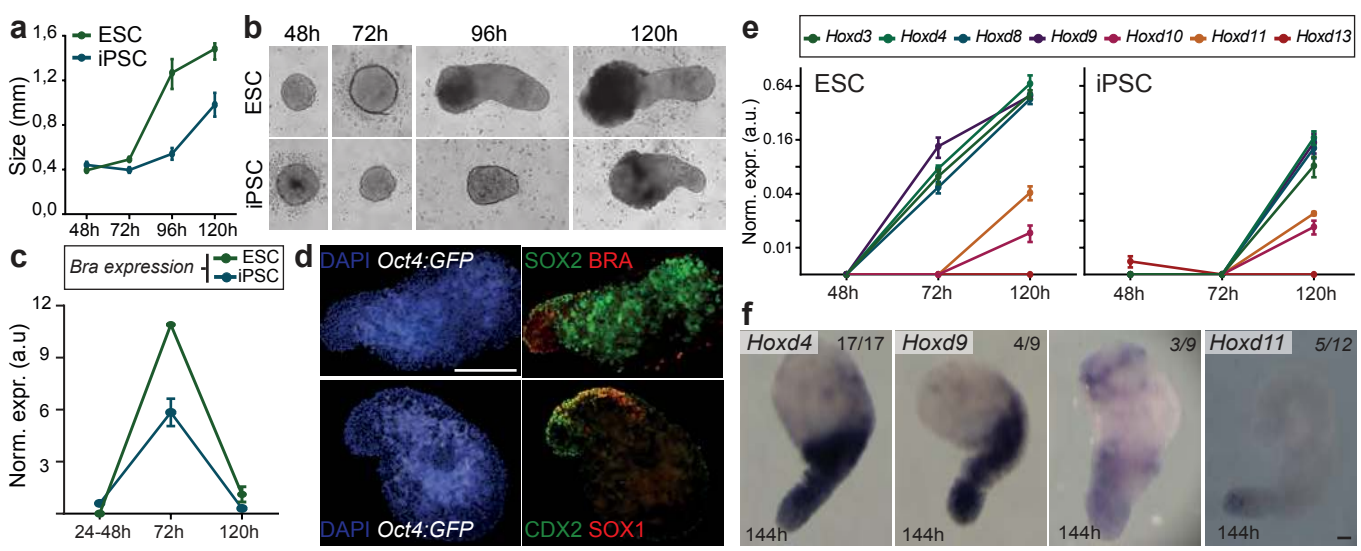

a
E6.5
E7.8
E8.5
E9.5
24h
48h
72h
96h
120h
144h
168h
PC2 (11.6% Explained Variance)
PC1 (79.4% Explained Variance)
b
48h
72h
96h
120h
144h
168h
5kb
Pooled Gastruloids RNAseq
Hoxa1 a2 a3 a4 a5 a6 a7 a9 a10 a11 a13
c
Hoxd4
Hoxd9
Hoxd11
Hoxd13
168h
13/16
10/11
14/17
4/7
d
DAPI Meox1 Hoxd4
DAPI Sox2 Hoxd4

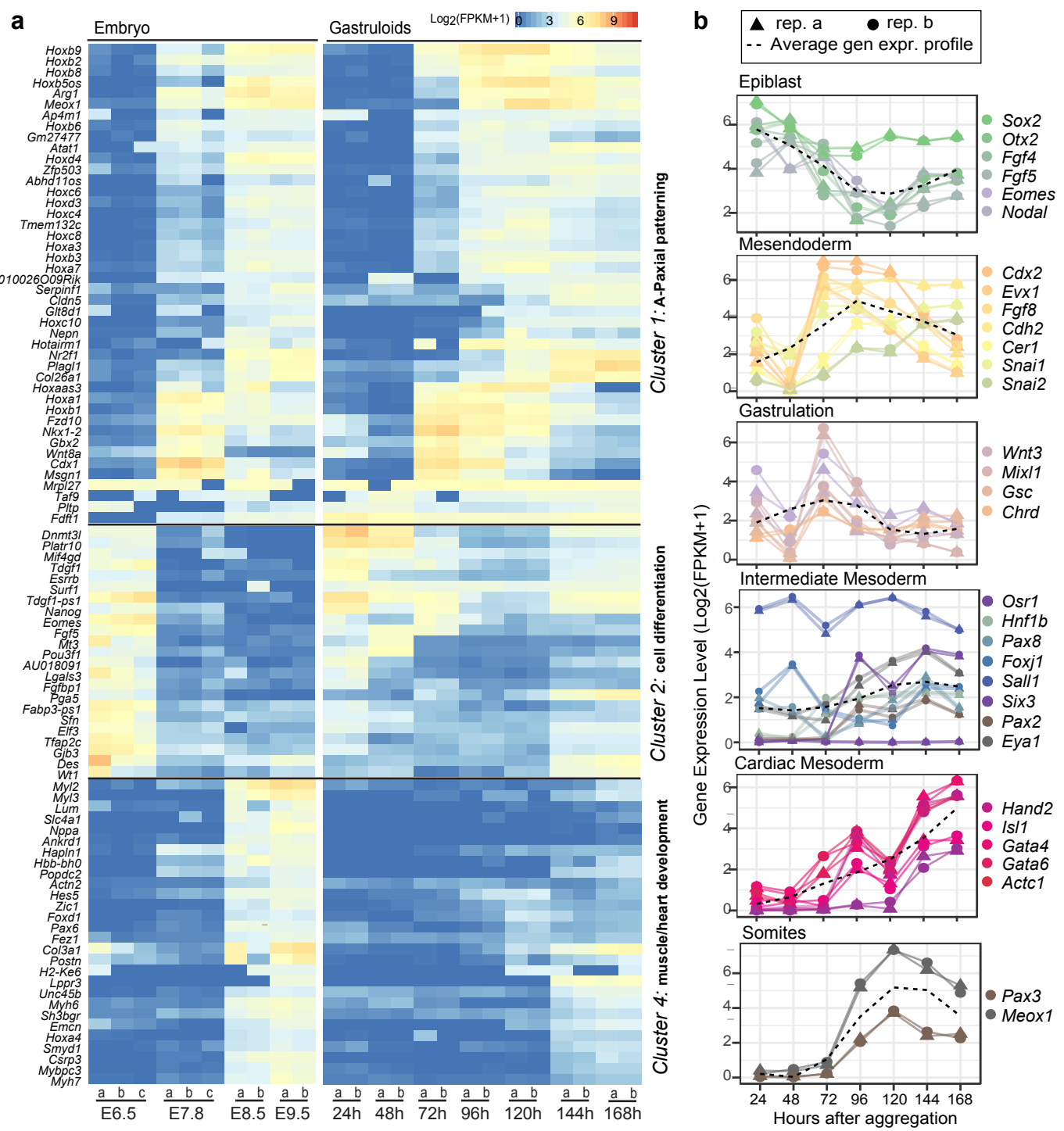
Figure 4



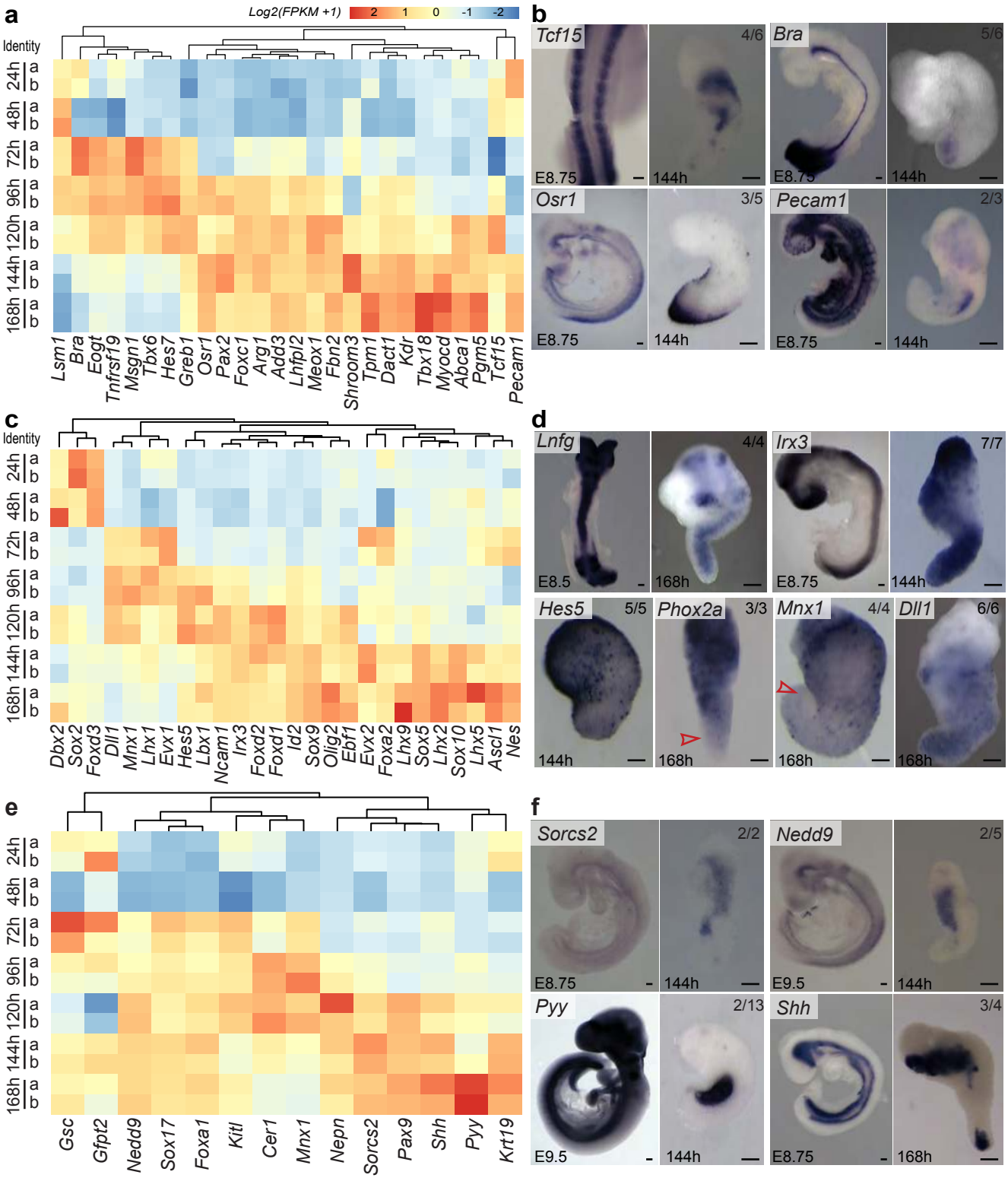
Extended data Figure 1



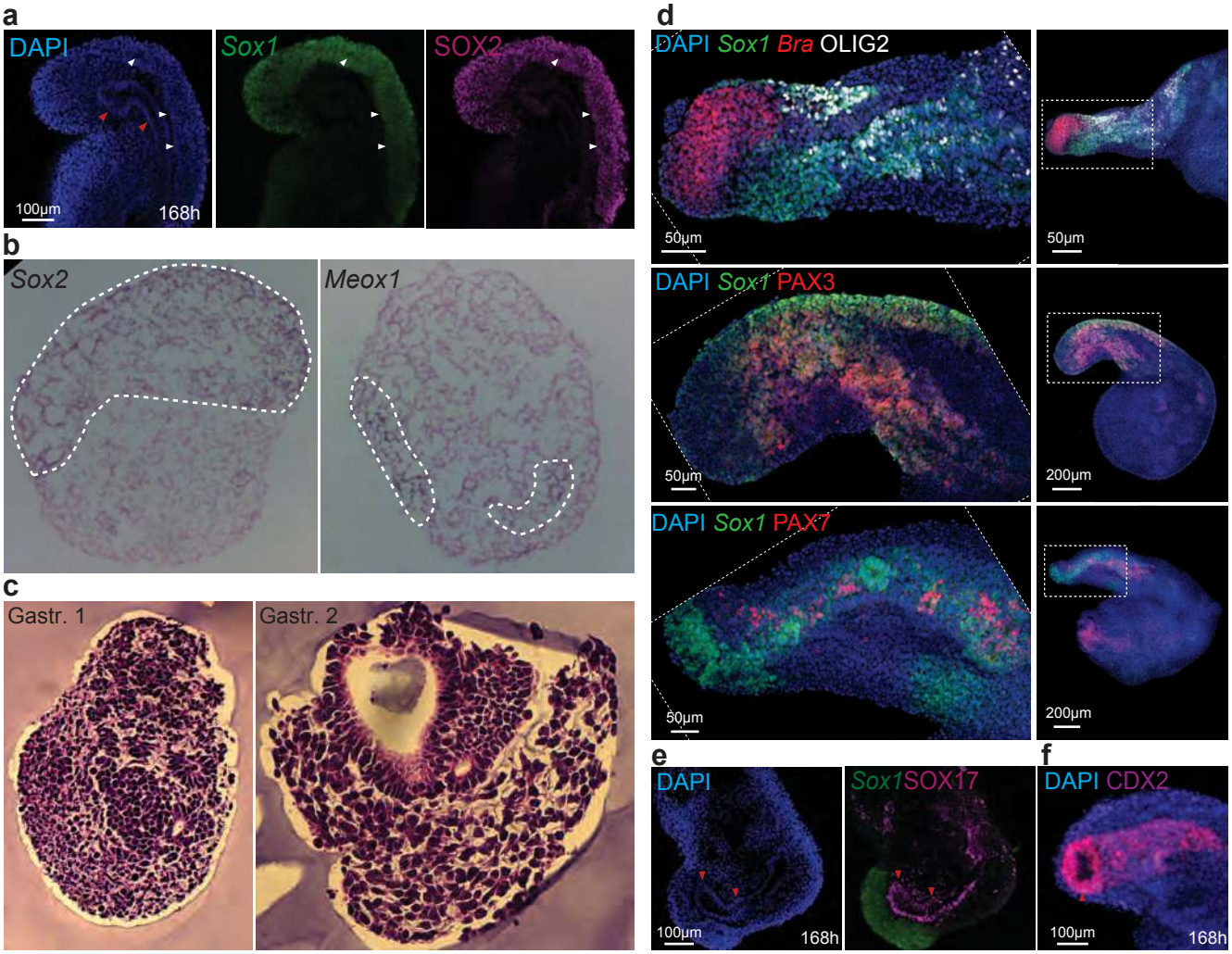
Extended data Figure 10



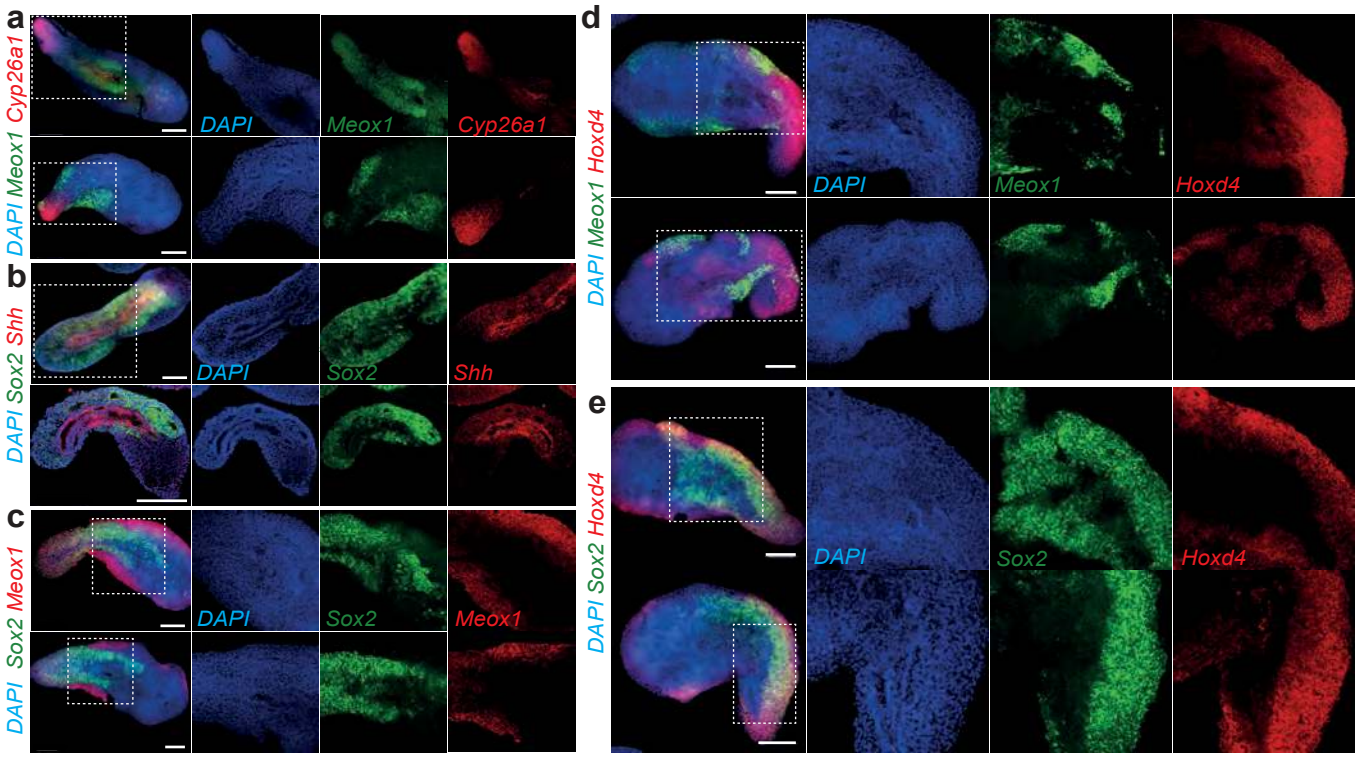
Extended data Figure 2



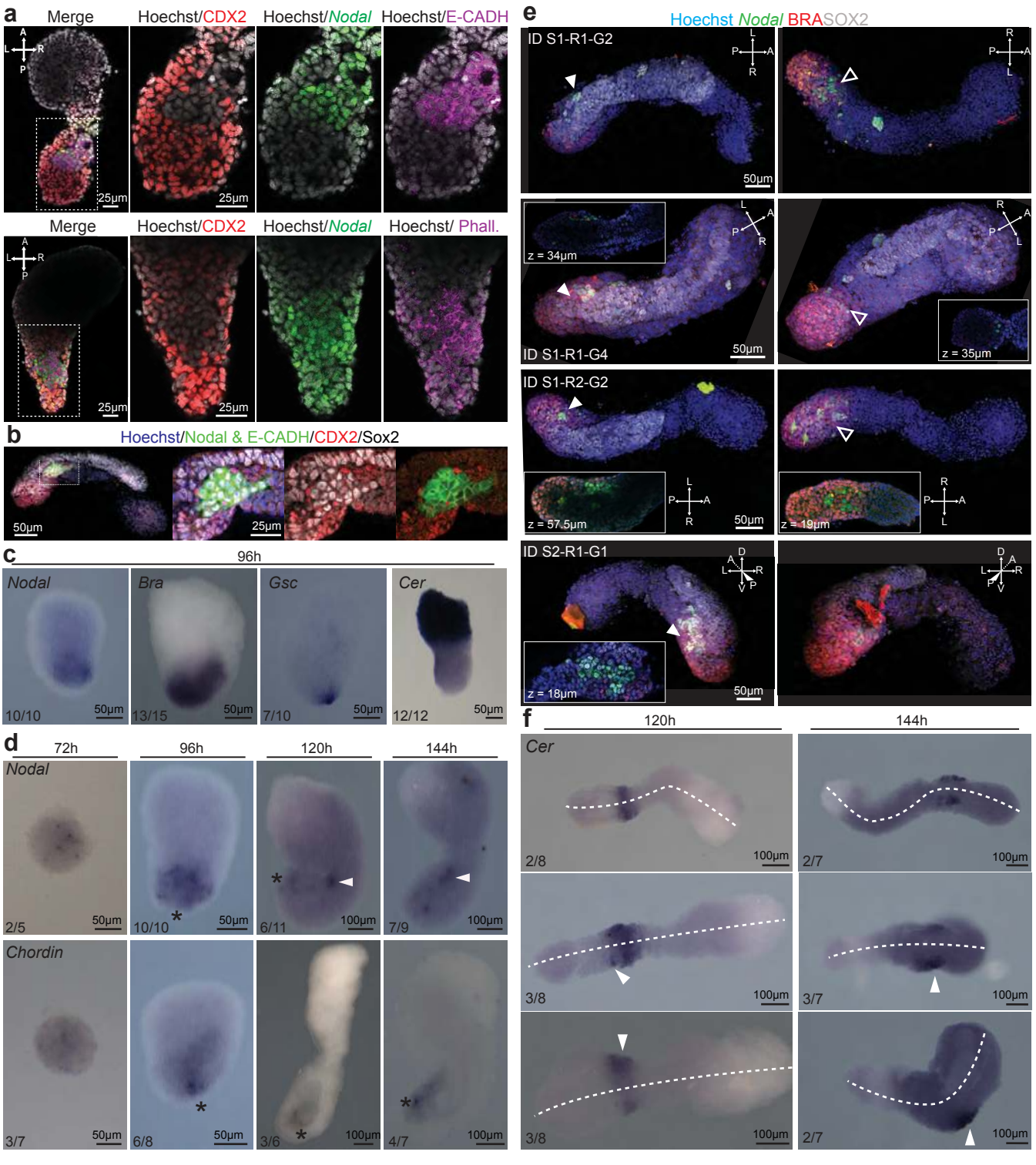
Extended data Figure 3



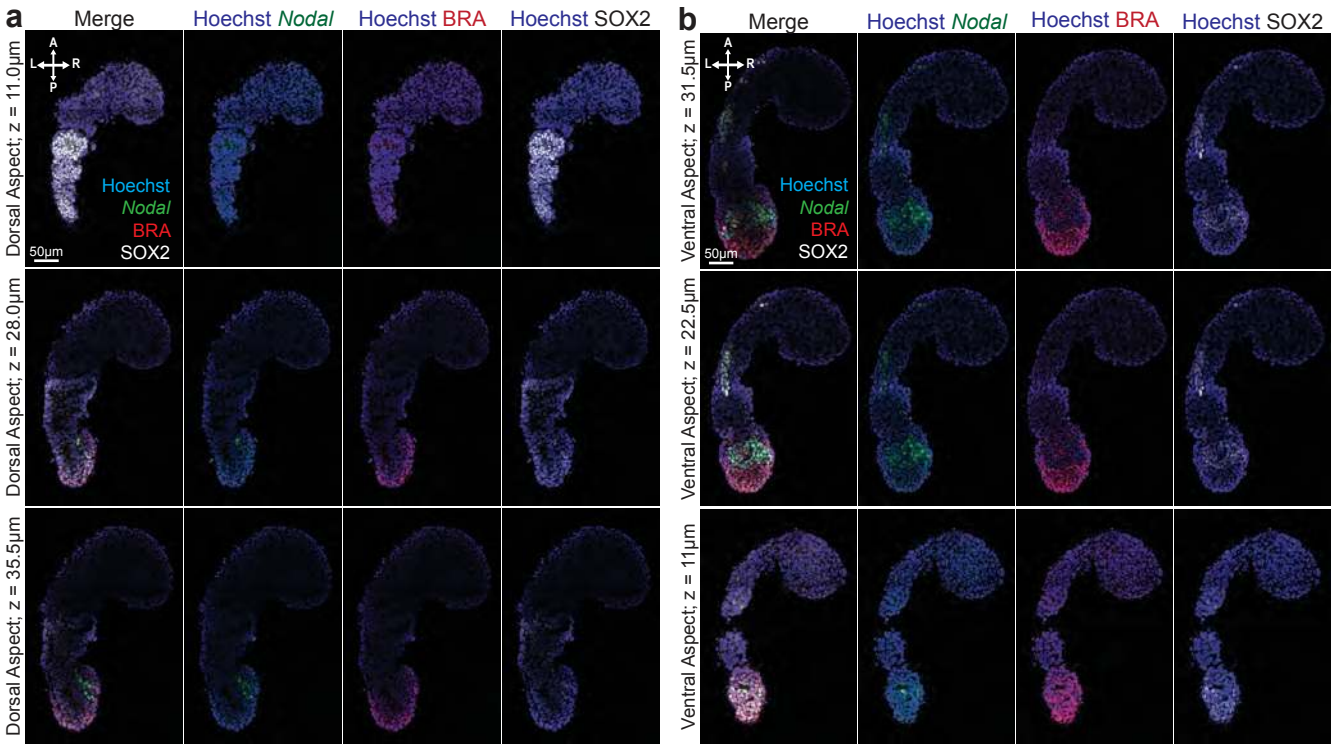
Extended data Figure 4



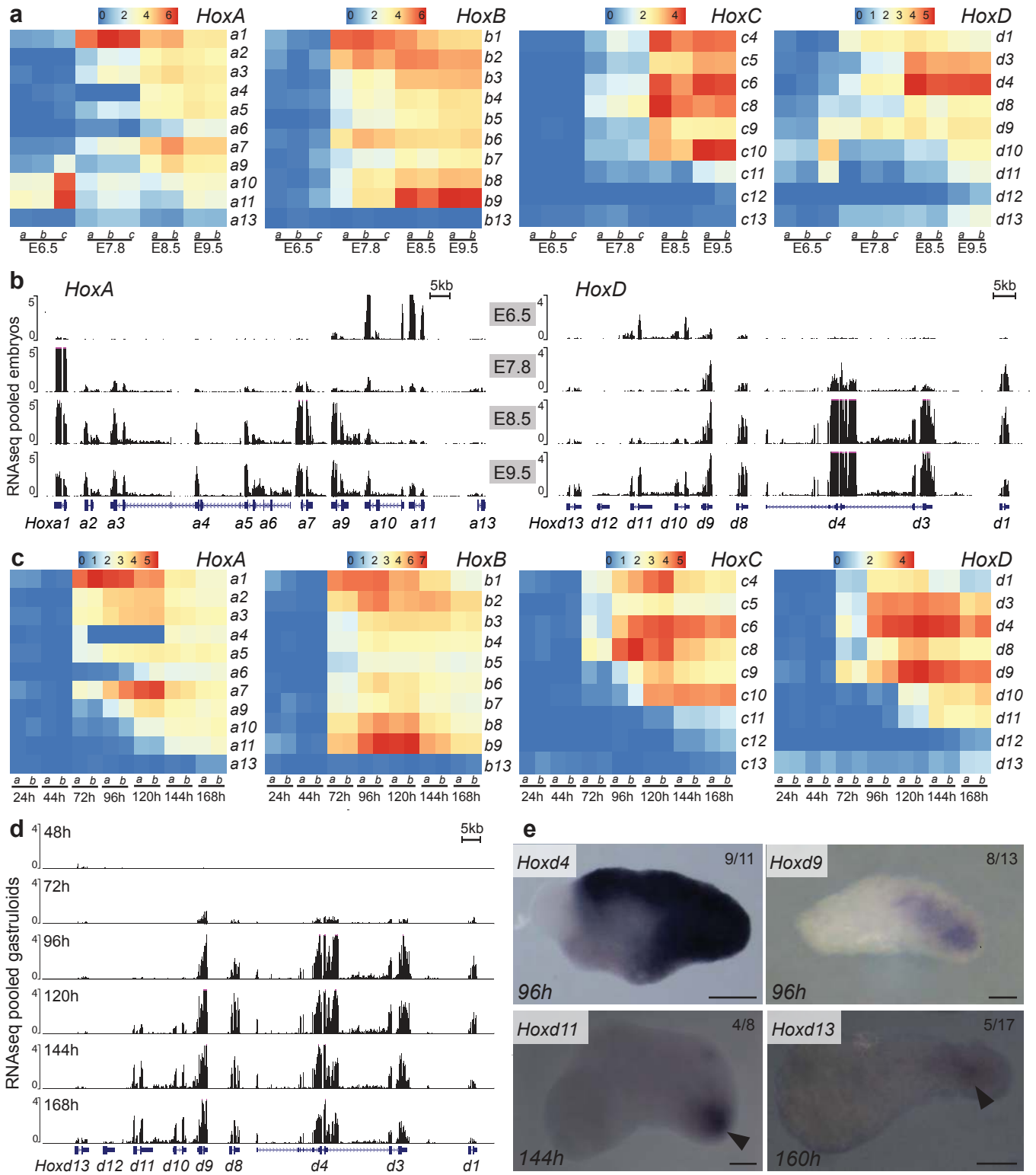
Extended data Figure 5



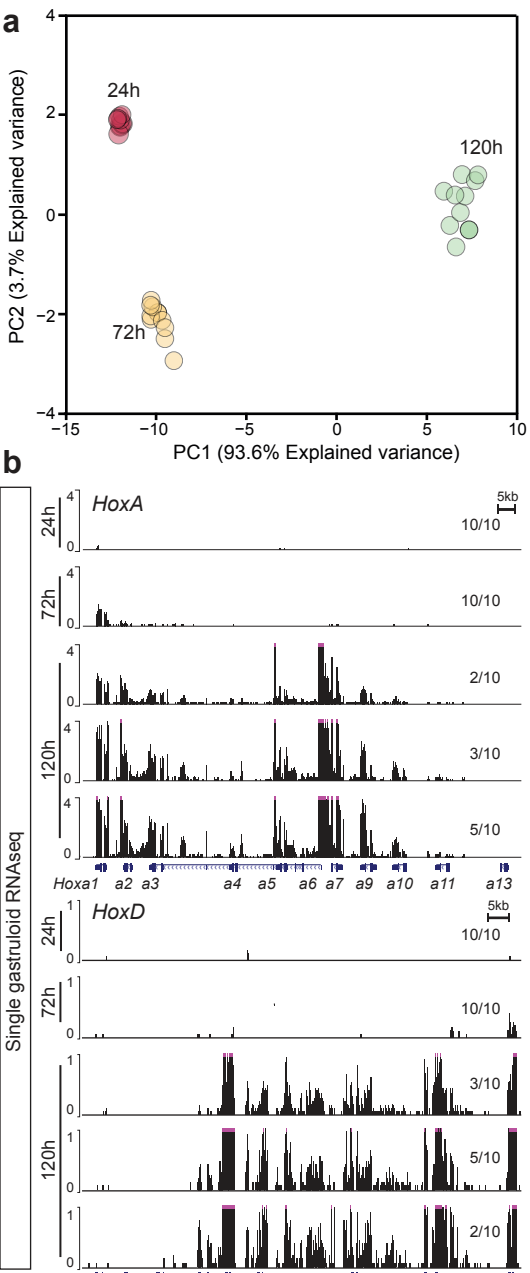
Extended data Figure 6



Extended data Figure 7



Extended data Figure 8



Extended data Figure 9