Generation and molecular characterization of mouse embryonic stem cells derived Trunk-Like Structures

Inaugural-Dissertation

to obtain the academic degree

Doctor of Philosophy in Natural Science - PhD

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Adriano Bolondi

Berlin, 2022

The work was carried out between 01/2017 and 12/2021 under the supervision of Prof. Dr. Alexander Meissner at the Max Planck Institute for Molecular Genetics - Department of Genome Regulation in Berlin.

Reviewer 1:

Prof. Dr. Alexander Meissner

Max Planck Institute for Molecular Genetics - Department of Genome Regulation

Department of Biology, Chemistry, Pharmacy of Freie Universität Berlin

Harvard University, Department of Stem Cell and Regenerative Biology

Broad Institute of MIT and Harvard

Reviewer 2:

Prof. Dr. Sigmar Stricker

Department of Biology, Chemistry, Pharmacy of Freie Universität Berlin

Date of defense: 13th June 2022

ACKNOWLEDGEMENTS

First of all I would like to thank Alex for having given me the chance to work in his lab. You gave for the freedom and independence to pursue my research goals, to explore, and most importantly to make mistakes. I would not have gone so far without your trust and support.

An important mention of course to all present and past members of the Meissner Lab, especially to the hardcore group of incredible friends that shared this journey with me from the very beginning (Christina, Sandro, Nina, Raha, Abhi, Sandy, Helene, Chuck). You are very precious to me.

A special thanks also to the "*TLS team*" for sharing with me all the steps of this fantastic adventure. Jesse, Leah, Ipek, Helene, Dennis, Bernhard you thought me a lot and I am extremely grateful to have been working in close contact with you.

Another thanks goes to my past and present mentors and collaborators worldwide (Jesse, Zack, Michelle, John, Sigmar, Denes, Aydan, Davide, Marcella, Fei, Evan, Leo, Giorgio, Diletta, Steffi, Philipp, Dario). I have learned immensely from our discussions and am really looking forward to more exciting Science together in the future.

The research done during my PhD was made possible thanks to the continuous and untiring work of so many people at the MPIMG including Cordula, Birgit, Maria, Jen, Sabine, Petra, Thorsten, Rene, Norbert, Lars, Uta, Claudi, Thomas, Paul, David, Patricia. Thank you all for the amazing support.

A special thanks also goes to my family for constantly supporting my decisions and for having given me all the possible love anyone could dream. Nonna Lora, who thought me to respect and include everyone, to always offer help to others and to greatly appreciate what we have without complaining of what we miss. My mum, who grew me and my brother with immense love and altruistically pushed us to pursue our passions and desires every day of our lives. My dad, who I am sure would be extremely proud of my achievements. My brother, who was and continue to be a role model for me and for my growth. All the rest of my family (Zia Rossella, Zio Furio, Zia Elisabetta, Zio Pierluigi, Fabbri, Giovi, Tanit, Livia, Marco, Martin, Olimpia, Natalie) for having given me a warm and stimulating environment to grow in.

Another big thanks to the *"Impennati del tavolo 8"* (Ludo, Coglia, Ika, Vale, Checco, Eli, Jimmy, Ventu, Caro, Tommy, Ciocco, Mora, Martuccia, Angelino, Puccio, Silve, Gabri, Kiko, Zim). You have been my best friends since I am four years old, we have been doing so many things and have gone through so many memorable moments together. I don't think I would actually exist without you.

Last in the list, but first in importance, the love of my life, Chiara. In the past years together I realized what it means to really be happy, to be in a deep connection with a person, and to love someone more than yourself. We have gone through so many beautiful and horrible things together, but these experiences made us stronger and even more bound. You pushed me to look for a PhD position abroad, and then followed me in this Berlin experience. Thank you so much for all you have done for me, for re-choosing me every day, and for blindly supporting my decisions. I love you.

Declaration of Independence:

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

“The interdependence of the chemical and mechanical data adds enormously to the difficulty, and attention will therefore be confined, so far as is possible, to cases where these can be separated.“

Alan Turing, The Chemical Basis of Morphogenesis, 1952

“The integrative powers of the embryo, at all of its levels, are however so pervasive that they never permit themselves to be overlooked.“

Jane Oppenheimer, Essays in the History of Embryology and Biology, 1967

SUMMARY

One of the greatest mysteries in developmental biology is how the orchestrated process of cellular differentiation guides a single cell to form a functional organism comprised of complex tissues and organs. A detailed understanding of these events in mammals is obscured by the inaccessibility of embryonic development *in utero*. In recent years, community efforts have led to the development of stem cell-based models of embryogenesis, able to mimic some cellular and molecular aspects of the early stages of development *in vitro*. Despite these advances, such models lack proper embryo-like morphogenesis necessary for tissue formation and organ development. To overcome this limitation, I optimized and characterized a protocol to precisely model the morphogenetic changes that occur during gastrulation and early organogenesis. Within only five days, the protocol produces "Trunk-Like Structures (TLS)," self-organized synthetic embryos originating from mouse pluripotent stem cell aggregates treated with chemical signaling molecules (gastruloids) and embedded in an extra-cellular-matrix (ECM) to mirror the *in utero* environment. TLS morphologically and molecularly resemble *in vivo* development, displaying a high level of trunk-tissue organization that includes the presence of somites (building blocks of future bones, muscles, and cartilage), a neural tube (future spinal cord), and a gut tube (future gastrointestinal tract). To test TLS similarities to the *in vivo* embryo, I integrated different single-cell technologies and confirmed the high complexity of these synthetic embryos. Specifically, I observed that TLS progress from their stem cell origin through cell type maturation, producing relevant subtypes of the developing trunk. Unlike *in vivo* development, the TLS platform allows rapid and tunable genetic and chemical perturbations with the benefits of uninterrupted and continuous observation. As a proof of principle, I utilized CRISPR/Cas9 based genetic ablation to recapitulate a well-studied mutant phenotype. Next, I influenced TLS developmental trajectories by targeted chemical modulation, resulting in the overproduction of somites, a phenotype never observed before *in vivo*. Finally, I investigated the molecular mechanism responsible for the enhanced tissue morphogenesis observed in TLS. I found that cells' ability to organize into complex tissues requires the presence of the extra-cellular-matrix, a specific feature we implemented in our protocol. Trunk-Like Structures provide a scalable, tractable, and reproducible platform to study normal and aberrant embryonic development *in vitro* at an unprecedented spatiotemporal resolution.

ZUSAMMENFASSUNG

Eines der größten Rätsel der Entwicklungsbiologie ist die Frage, wie der orchestrierte Prozess der zellulären Differenzierung einer einzelnen Zelle zur Bildung eines funktionellen Organismus mit komplexen Geweben und Organen führt. Ein detailliertes Verständnis dieser Vorgänge bei Säugetieren wird durch die Unzugänglichkeit der Embryonalentwicklung *in utero* erschwert. In den letzten Jahren haben Bemühungen der Wissenschaft zur Entwicklung stammzellbasierter Modelle der Embryogenese geführt, die einige zelluläre und molekulare Aspekte der frühen Entwicklungsstadien *in vitro* nachahmen können. Trotz dieser Fortschritte mangelt es diesen Modellen an einer angemessenen embryoähnlichen Morphogenese, die für die Gewebebildung und Organentwicklung erforderlich ist. Um diese Einschränkung zu überwinden, habe ich ein Protokoll optimiert und charakterisiert, das die morphogenetischen Veränderungen, die während der Gastrulation und der frühen Organogenese auftreten, präzise nachbildet. Innerhalb von nur fünf Tagen erzeugt das Protokoll "Trunk-Like Structures (TLS)", selbstorganisierte synthetische Embryonen, die aus pluripotenten Stammzellaggregaten der Maus stammen, die chemischen Signalmolekülen (Gastruloiden) ausgesetzt und in eine extrazelluläre Matrix (ECM) eingebettet wurden, um die Umgebung in utero nachzubilden. Die Entwicklung von TLS ähnelt morphologisch und molekularbiologisch der Entwicklung *in vivo* und zeigt ein hohes Maß an Stammgewebsorganisation, die das Vorhandensein von Somiten (Bausteine der zukünftigen Knochen, Muskeln und Knorpel), eines Neuralrohrs (zukünftiges Rückenmark) und eines Darmrohrs (zukünftiger Magen-Darm-Trakt) umfasst. Um die Ähnlichkeit der TLS mit dem *In-vivo*-Embryo zu testen, habe ich verschiedene Einzelzelltechnologien integriert und die hohe Komplexität dieser synthetischen Embryonen bestätigt. Insbesondere habe ich beobachtet, dass TLS von ihrem Stammzellursprung die Zelltypreifung durchlaufen und relevante Subtypen des sich entwickelnden Stammes hervorbringen. Im Gegensatz zur *In-vivo*-Entwicklung ermöglicht die TLS-Plattform schnelle und einstellbare genetische und chemische Störungen mit dem Vorteil einer ununterbrochenen und kontinuierlichen Beobachtung. Zum Beweis des Prinzips habe ich eine CRISPR/Cas9-basierte genetische Ablation eingesetzt, um einen gut untersuchten Mutantenphänotyp zu rekapitulieren. Als Nächstes habe ich die Entwicklungsbahnen der TLS durch gezielte chemische Modulation beeinflusst, was zu einer Überproduktion von Somiten führte - ein Phänotyp, der bisher noch nie *in vivo* beobachtet wurde. Schließlich untersuchte ich den molekularen Mechanismus, der für die bei TLS beobachtete verstärkte Gewebemorphogenese verantwortlich ist. Ich fand heraus, dass die Fähigkeit der Zellen, sich zu komplexen Geweben

zu organisieren, das Vorhandensein der extrazellulären Matrix voraussetzt, eine Besonderheit, die wir in unserem Protokoll implementiert haben. Trunk-Like Structures bieten eine skalierbare, nachvollziehbare und reproduzierbare Plattform zur Untersuchung der normalen und abweichenden Embryonalentwicklung *in vitro* mit einer noch nie dagewesenen räumlichen und zeitlichen Auflösung.

LIST OF PUBLICATIONS, CONTRIBUTIONS AND AWARDS

Publications

Veenvliet JV*, **Bolondi A***, Kretzmer H, Haut L, Scholze-Wittler M, Schifferl D, Koch F, Guignard L, Sampath Kumar A, Pustet M, Heimann S, Buschow R, Wittler L, Timmermann B, Meissner A, Herrmann BG. "Mouse Embryonic Stem Cells self-organize into Trunk-Like-Structures with Neural Tube and Somites" (2020). ***Science*** DOI:10.1126/science.aba4937
* Veenvliet and Bolondi contributed equally to this study.

Bolondi A, Haut L, Gassaloglu SI, Burton P, Kretzmer H, Buschow R, Meissner A, Herrmann BG, Veenvliet JV. "Generation of Mouse Pluripotent Stem Cell-derived Trunk-Like-Structures: An in vitro Model of Post-implantation Embryogenesis" (2021). ***BioProtocol*** DOI: 10.21769/BioProtoc.4042

Conferences and talks

Bolondi A. "*Simultaneous somitogenesis and neural tube formation in self-organized Trunk-Like-Structures*". Selected talk at the 13th Berlin late Summer Meeting – "Rising from the Ashes: Regeneration at the Single Cell Level" (2020).

Bolondi A. "*Simultaneous somitogenesis and neural tube formation in self-organized Trunk-Like-Structures*". Invited speaker at the Cell Circuits and Epigenomics Program Seminar Series – Broad Institute of Harvard and MIT (2020).

Awards

Bolondi A. "Paper of the year" at the BSRT-Charite'-DFG PhD student award in regenerative medicine (*second place*). "*Mouse Embryonic Stem Cells self-organize into Trunk-Like-Structures with Neural Tube and Somites*". (2020).

LIST OF ABBREVIATIONS

AA	Ascorbic acid
AP	Anterior-posterior
AVE	Anterior visceral endoderm
bFGF	basic Fibroblast growth factor
BMP	Bone morphogenetic protein
cAMP	Cyclic adenosine monophosphate
CARM1	Coactivator-associated arginine methyltransferase
CDX2	Caudal type homeobox 2
CER1	Cerebrus 1
CLE	Caudal lateral epiblast
DBX	Developing brain homeobox protein family
DKK1	Dickkopf WNT Signaling Pathway Inhibitor 1
DV	Dorso-ventral
DVE	Distal visceral endoderm
EBs	Embryoid bodies
ECM	Extracellular matrix
EMT	Epithelial-to-mesenchymal transition
EN2	Engrailed homeobox 2
EOMES	Eomesodermin
EPI	Epiblast
EPSCs	Expanded potential mESCs
ETS	ESC- and TSC-derived embryos
ETX	ESC-,TSC- and XEN-derived embryos
ExE	Extraembryonic ectoderm
FGF	Fibroblast growth factors
FGF4	Fibroblast growth factor 4

FGF8	Fibroblast growth factor 8
FGFR1	Fibroblast growth factor receptor 1
FOXA2	Forkhead box protein A2
FOXF1	Forkhead box F1
GAG	Glycosaminoglycans
GATA4	GATA binding protein 4
GATA6	GATA binding protein 6
GRN	Gene regulatory network
HAND1	Heart and neural crest derivatives expressed 1
HAND2	Heart and neural crest derivatives expressed 2
HES7	Transcription factor Hes7
HHEX	Hematopoietically-expressed homeobox protein
ICM	Inner cell mass
IM	Intermediate mesoderm
LEF1	Lymphoid enhancer-binding factor-1
LEFTY1	Left-right determination factor 1 precursor
LIM1	LIM homeobox gene
LNFG	Lunatic fringe
LPM	Lateral plate mesoderm
mESCs	Mouse embryonic stem cells
MESP2	Mesoderm Posterior BHLH Transcription Factor 2
MIXL1	Mix Paired-Like Homeobox 1
MYF5	Myogenic factor 5
MYOD	Myogenic differentiation 1
NKX	Nk homeobox proteins
NMPs	Neuromesodermal progenitors
NOG	Noggin
NOTCH1	Notch homolog 1
NP	Neural plate

NSB	Node-streak border
NT	Neural tube
OCT4	Octamer binding transcription factor 4
OTX2	Orthodenticle Homeobox 2
PAX1	Paired box protein 1
PAX2	Paired box protein 2
PAX3	Paired box protein 3
PAX6	Paired box protein 6
PAX7	Paired box protein 7
PAX8	Paired box protein 8
PAX9	Paired box protein 9
PrE	Primitive endoderm
PGCs	Primordial germ cells
PGCLCs	Primordial germ cells-like cells
PRRX1	Paired Related Homeobox 1
PS	Primitive streak
PSM	Pre-somitic mesoderm
RA	Retinoic Acid
RIPPLY2	Ripply Transcriptional Repressor 2
SALL4	Spalt like transcription factor 4
scRNA-seq	single-cell RNA-sequencing
SHH	Sonic hedgehog
SNAI1	Snail Family Transcriptional Repressor 1
SOX17	Sry-Box transcription factor 17
SOX2	Sry-Box transcription factor 2
TBX18	T-box transcription factor 18
TBX6	T-box transcription factor 6
TBXT or T	T-box transcription factor T
TCF1	Transcription factor 1

TE	Trophoectoderm
TEAD4	TEA domain transcription factor 4
TSC	Trophoblast stem cells
UNCX	UNC Homeobox
VE	Visceral endoderm
VEGF	Vascular endothelial growth factor 165
XEN	Extra-embryonic endoderm cells
β-CAT	β-Catenin

LIST OF FIGURES

1 INTRODUCTION

1.1 Mouse embryogenesis

1.1.1 Pre- and peri-implantation development

Mammalian embryonic development is a highly dynamic process involving cell movement and organization, tissue growth, and cellular specification to ensure a functional body plan formation. Embryogenesis starts at fertilization when the egg and the sperm fuse to form the zygote (1-cell stage). This totipotent cell, able to originate all other cells of the future organism, undergoes a series of asynchronous cell divisions during the developmental time before the embryo implants in the uterus, without increasing embryo size or volume (preimplantation development) (**Figure 1**). During preimplantation development, which in mice lasts three and a half days from fertilization to early blastocysts stage, the first cell specification event occurs, with few cells in the embryo differentiating towards the trophectoderm (TE) lineage, destined to form the placenta [1]. Genetic, epigenetic, and mechanical factors control the spatial and temporal specification of the TE. For example, TEA domain transcription factor 4 (TEAD4) and Caudal type homeobox 2 (CDX2) are transcription factors necessary for TE specification [1]. Moreover, recent studies have shown that the transient and asymmetric expression of a long non-coding RNA (*Linc-GET*), combined with the enzymatic activity of the Coactivator-associated arginine methyltransferase (CARM1), contributes to the TE cell fate initiation in the 2-cell stage embryo [2, 3]. Cellular polarization and movement are also crucial for the specification of the TE, as demonstrated by the temporally asymmetric deposition of actin in cells of the early embryo [4].

Cells that do not commit to TE will contribute to the future blastocyst's inner cell mass (ICM). Cells of the ICM are pluripotent, and they express a set of transcription factors that guarantee proliferation and oppose differentiation, including Nanog, Octamer binding transcription factor 4 (OCT4), Sry-Box transcription factor 2 (SOX2), and Spalt-like transcription factor 4 (SALL4) [5-8]. In the early blastocysts, and concomitant with embryo implantation (peri-implantation development), the second cellular differentiation events take place, with cells from the ICM differentiating into epiblast (EPI) and primitive endoderm (PrE). The PrE, or extraembryonic endoderm, will then contribute to the yolk sac, the parietal endoderm, and partially to the embryonic endoderm [9]. This second differentiation event necessitates the expression of transcription factors such as GATA binding protein 4 and 6

(GATA4 and GATA6), which antagonize the activation of pluripotency-associated genes [10-12]. The EPI will then give rise to the embryo proper, and therefore originate tissues of the future adult organism. The embryo implants then into the uterine wall, and can proceed with the differentiation and morphogenetic events that will shape the future body plan (post-implantation development) (**Figure 1**).

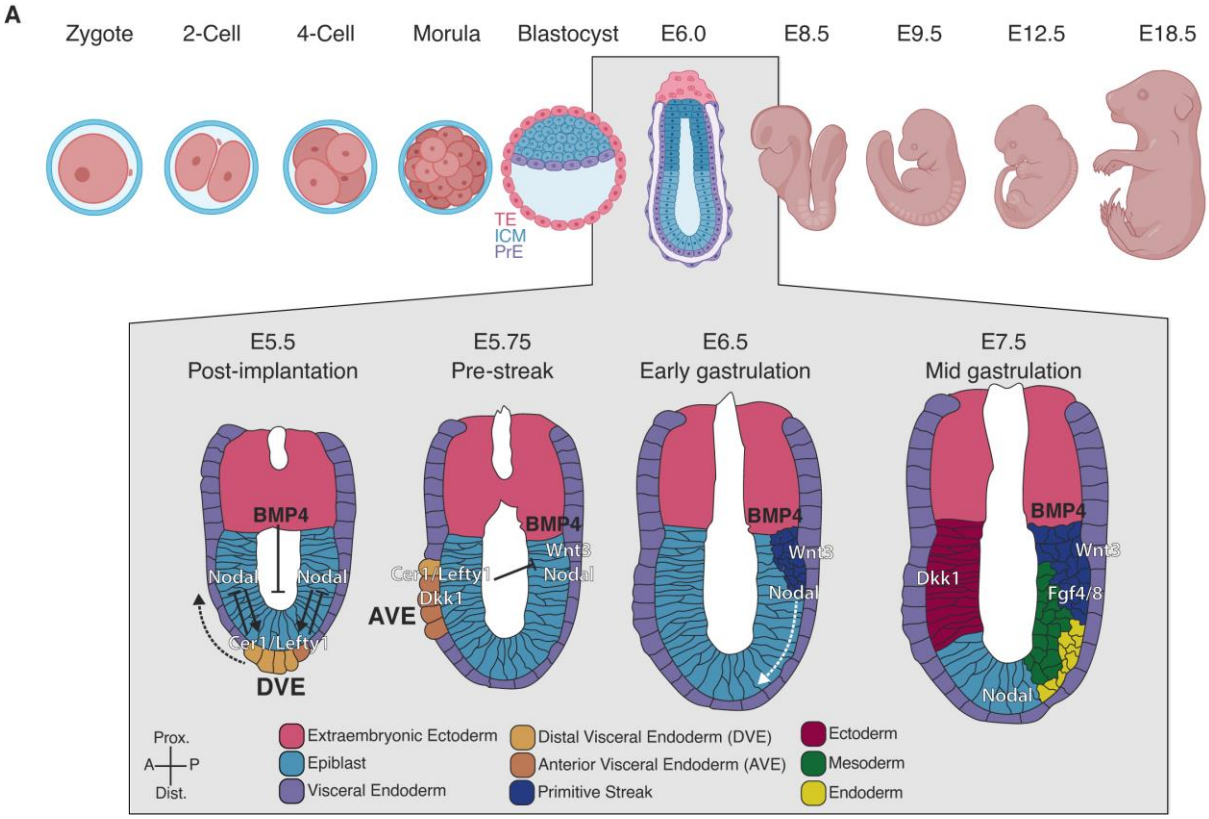

Figure 1. Mouse embryogenesis and gastrulation

(A) Schematic representation of mouse embryonic development from fertilization (zygote) to term (E18.5). The grey box depicts the signaling gradients and subsequent cellular responses after implantation, from E5.5 to E7.5. After implantation, at E5.5, different signaling gradients, including NODAL and BMPs, contribute to the specification and initial migration of the DVE from the distal to the anterior side of the epiblast. The migrated AVE secretes NODAL antagonists such as LEFTY1 and CER1, creating an anterior to posterior Nodal gradient in the epiblast. These signals induce *Wnt* activation in the most posterior portion of the epiblast and the subsequent specification of the primitive streak in the early phases of gastrulation. At mid gastrulation stage, around E7.5, the three germ layers (ectoderm, mesoderm, and endoderm) start emerging patterning. TE, trophectoderm; ICM, inner cell mass; PrE, primitive endoderm; E, embryonic day; Prox, proximal; Dist, distal; A, anterior; P, posterior. Figure realized using Adobe Illustrator and BioRender. Schematics in the grey box adapted from Bardot & Hadjantonakis, 2020.

1.1.2 Gastrulation

Gastrulation is the embryonic process that results in the differentiation of the pluripotent epiblast cells into the three germ layers (ectoderm, mesoderm, and endoderm) and the consequent formation of the three body axes. This event is crucial for the progression of embryonic development and, therefore, highly sensitive to genetic and mechanical perturbations. In mice, gastrulation spans almost three days of embryonic development, from E6.25 to E9.5, with various signaling molecules gradients playing fundamental roles [13] (**Figure 1**). Gastrulation can initiate only after the anterior-posterior (AP) axis is established, with distal visceral endoderm (DVE) cells migrating around the epiblast cone to mark the anterior side of the embryo. These cells specify from the visceral endoderm (VE) when exposed to low levels of NODAL and Bone morphogenetic protein (BMP) signaling around E5.5 [14-16]. Subsequently, cells in proximity of the migrated DVE forms the anterior visceral endoderm (AVE), an important signaling center for the AP axis formation. The AVE maintains high levels of Cerebrus 1 (CER1) and Left-right determination factor 1 precursor (LEFTY1), two NODAL antagonists, concomitant with the expression of Dickkopf WNT Signaling Pathway Inhibitor 1 (DKK1), to mark the anterior portion of the embryo (**Figure 1**) [17, 18]. Thus, around E6.0, a two-dimensional signaling gradient along the AP axis is established, with WNT and NODAL being restricted to the posterior portion of the embryo [17, 19]. The high levels of WNT in the posterior portion of the embryo drive gastrulation initiation with the formation of the primitive streak (PS) (**Figure 1**). The PS consists of a group of cells that, around E6.25, start delaminating from the posterior epiblast portion and undergo epithelial-to-mesenchymal transition (EMT). The EMT process in the PS is coordinated by the combination of WNTs, BMPs, and fibroblast growth factors (FGFs), which allow the switch from E- to N-cadherin expression and the acquisition of a mesenchymal and migrating cell identity [20-22]. In particular, FGF4 and FGF8 play critical roles in embryonic posteriorization, as demonstrated by the fact that in their absence, mesoderm and endoderm are not produced as a consequence of failed EMT in the PS [21]. Analogously, lack of β-Catenin (β-CAT) or WNT3 expression results in failure to establish an AP axis during gastrulation [23, 24]. During EMT, the gene expression profile of the PS cells changes drastically, with downregulation of epiblast markers such as SOX2 and E-cadherin and upregulation of mesoderm specific transcriptional regulators such us Snail Family Transcriptional Repressor 1 (SNAI1), Mix Paired-Like Homeobox 1 (MIXL1), Eomesodermin (EOMES) and T-box transcription factor T (TBXT or T) [25-28]. The completion of EMT in the PS marks the first exit from pluripotency event, with the formation of the first germ layer,

namely the nascent mesoderm. The nascent mesoderm ingresses below the epiblast and migrates to extend towards the anterior side of the embryo, rapidly increasing embryo size and cell number. Many morphological and cellular specification events occur to ensure the formation of mesoderm-derived tissues in the trunk and thoracic region (see "*Trunk development"* chapter for a detailed description of these events).

Cells contributing to the endoderm lineage also partially arise from the PS from E6.5 [29]. In particular, these cells migrate anteriorly between the epiblast and the VE layers, ultimately intercalating with the VE (**Figure 1**) [30, 31]. Interestingly, endoderm cells seem to retain E-cadherin expression, despite originating from the mesenchymal PS, indicating a partial EMT followed by a rapid mesenchymal-to-epithelial transition (MET) in the region of the PS that lack SNAI1 expression [31]. Definitive endoderm, which is fully specified in the mid gastrulation stage around E7.5, is further characterized by the activation of transcription factors such as Sry-Box transcription factor 17 (SOX17), GATA6, GATA4, and Forkhead box protein A2 (FOXA2), that contribute to the gene regulatory circuit necessary for the induction and maintenance of this germ layer.

Concomitant with the gastrulation events happening in the posterior end of the embryo, the anterior portion of the epiblast starts exiting pluripotency, acquiring an ectodermal identity. In particular, the combined signaling from the AVE, the mesoderm, and the endoderm layers extending from the PS facilitates the specification of the anterior neuroectoderm [32]. These instructive signals induce and maintain the expression of neural-specific genes such as Orthodenticle Homeobox 2 (OTX2) and Engrailed homeobox 2 (EN2), which activity is necessary for the future head and brain development after gastrulation [33, 34]. From the late gastrulation stage on, the head fold starts to emerge from the anterior neuroectoderm and, together with the anterior mesoderm and endoderm, will originate the cranial and brain regions [32].

1.1.3 Trunk development

During post-implantation development, the trunk portion of the mouse embryo harbors proliferating and multipotent progenitor cells that will sustain the formation of fundamental tissues for the future body plan. Three main lineages are specified in the trunk region: the mesoderm, which will originate most of the trunk tissues, including the heart, the bones, the skeletal muscles, limbs, blood vessels, kidneys, and gonads; the endoderm, which will give rise to the gastrointestinal tract; and the neuroectoderm, which will mainly contribute to the future spinal cord (**Figure 2A**). The trunk develops from the continuous expansion of the anterior portion of the PS during and after gastrulation. This elongation process is granted by the presence of self-renewing and proliferating progenitors in the embryo's most caudal end, providing cells that contribute to multiple lineages [35, 36]. These pools of progenitors are located in the proximity of the node-streak border (NSB) and of the caudal lateral epiblast (CLE) and start being produced during mid gastrulation stage (**Figure 2B**) [36]. In particular, these cells seem to be regionalized in the posterior portion of the embryo, and such localization is instructive for lineage allocation during differentiation. In fact, progenitor cells known as neuromesodermal progenitors (NMPs) are located adjacent to the NSB and will give rise to paraxial mesoderm and neuroectoderm, while lateral and intermediate mesoderm (LPM and IM) progenitors reside in proximity of the CLE [37-39].

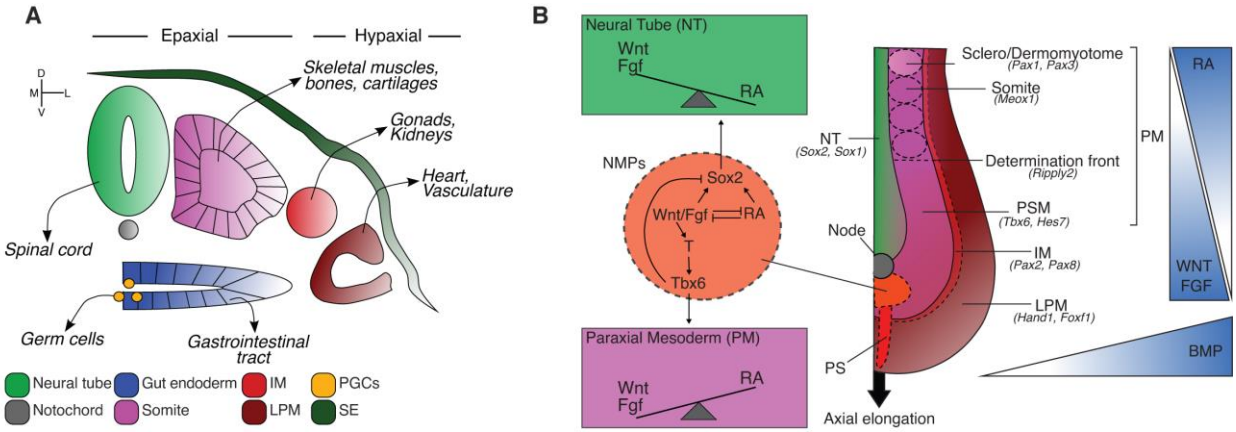

Figure 2. Trunk development

(A) Schematic representation of a transversal section of the trunk region of a developing embryo at mid-gestation and corresponding tissue organization. The neural tube and notochord occupy the medial region of the trunk, while the mesoderm tissues are organized along the mediolateral axis in epaxial (paraxial mesoderm) and hypaxial (IM and LPM) compartments. The gut endoderm, which hosts migrating PGCs, lays in the ventral portion of the embryo. The derivatives of each embryonic tissue are indicated in *italic*. **(B)** Spatial organization and gene regulatory network (GRN) orchestrating embryonic axial elongation and trunk tissue specification from NMPs. NMPs acquire their bi-potency in the posterior portion of a developing embryo during gastrulation at the node-PS border. The correct balance of two opposing signaling gradients (WNT/FGF and RA) guarantees further lineage specification along the AP axis by activating the corresponding GRNs for the NT and PM formation. Concomitantly, BMP signaling regulates mesodermal diversification along the mediolateral axis. IM, intermediate mesoderm; LPM, lateral plate mesoderm; PGCs, primordial germ cells; SE, surface ectoderm; M, medial; L, lateral; D, dorsal; V, ventral; NMPs, neuromesodermal progenitors; PS, primitive streak; PSM, pre-somitic mesoderm; NT, neural tube; PM, paraxial mesoderm. Figure realized using Adobe Illustrator.

1.1.3.1 Neuromesodermal progenitors

NMPs are retained in a proliferating state while continuously providing derivatives of both the neural tube and the paraxial mesoderm in a stem cell-like fashion (**Figure 2B**) [40]. Lineage labeling and tracing experiments have shown that NMPs are actual bi-potent progenitors, with clonal descendants identified in both the somitic and neural compartment [41]. NMPs acquire their competence during gastrulation due to opposing signaling gradients in the posterior portion of the developing embryo [42]. In particular, high WNT and FGF signaling, concomitant with low retinoic acid (RA), are necessary for the specification and maintenance of these progenitor cells' potential (**Figure 2B**) [43, 44]. The delicate balancing of these signaling gradients maintains NMPs in a bistable state, characterized by the co-expression of T and SOX2, mesodermal and neural-specific transcription factors, respectively [38, 45]. Both T and SOX2 are activated by dominant WNT signaling in the tailbud. In particular, WNT3A

signaling cascade is responsible for the direct induction of T [14]. Moreover, FGFs (specifically FGF4 and FGF8) are also necessary for T activation in mice [46]. In contrast, RA suppresses genetic programs related to WNT [39]. Fate decision is subsequently triggered by the unbalancing of such gradients, with FGF and WNT signaling pushing towards a mesodermal fate, while RA driving neural specification (**Figure 2B**) [47, 48]. In fact, *Wnt* downregulation biases NMPs towards the neural lineage. On the other hand, increased WNT activity results in upregulation of T in NMPs, which directly binds and activates T-Box Transcription Factor 6 (*Tbx6*) [38, 49]. TBX6 is a paraxial mesoderm-specific factor and reinforces the lineage commitment by suppressing SOX2 expression (**Figure 2B**) [49]. Despite recent advances in lineage tracing technologies and high-resolution microscopy, the mechanism by which NMPs contribute to both neural and mesodermal lineages and how the specific commitment is achieved remains elusive.

1.1.3.2 Somitogenesis

Once NMPs move out from the growth zone and start upregulating *Tbx6*, they acquire a mesodermal identity and start contributing to the nascent PSM (**Figure 2B**). The PSM will then give rise to the somitic mesoderm, building the future trunk skeleton and musculature blocks. Somites deposition in mice occur between E8.0 and E13.0 in a highly orchestrated process involving transcriptional regulation, signaling gradients, and mechanical forces (**Figure 3**). The anterior portion of the PSM sequentially subdivides into blocks of epithelialized tissue in a timed and periodic fashion (**Figure 3A**) [50]. This fascinating developmental mechanism has been studied for decades, and a primary theoretical model of somitogenesis was first proposed in 1976, the "*clock and wavefront model*" (**Figure 3B**) [51]. This model includes two main components for the rhythmic deposition of new somites during axial elongation: a temporal component ("*the clock*"), which consist of a series of molecular events that signal the anterior PSM when to segment a new somite, and a spatial component ("*the wavefront*"), the region of the PSM where the segmentation program starts [52]. From the molecular perspective, the segmentation clock is mainly controlled by the oscillation of Notch homolog 1 (NOTCH1) expression along the AP axis in the PSM (**Figure 3B**) [53]. NOTCH1 activity results in the oscillation of other genes in the PSM that are directly involved in feedback loops mechanism with NOTCH itself, including the Transcription factor Hes7 (HES7) and Lunatic fringe (LNFG) [53-55]. In fact, *Hes7* is activated at first in the posterior part of the PSM and mediates direct Notch activation. In turn, Notch activity suppresses *Hes7* expression and activates *Lfng*, which encodes a NOTCH receptors modulator and therefore represses Notch expression with a

feedback mechanism [55]. When the wave of high NOTCH reaches the most anterior portion of the PSM, concomitant with high TBX6 expression, Mesoderm Posterior BHLH Transcription Factor 2 (MESP2) is activated in a stripe of cells at the so-called "*determination front*" (**Figure 3B**) [50, 56]. The following somite segmentation is mediated by the action of Ripply Transcriptional Repressor 2 (RIPPLY2), a direct target of MESP2 in the determination front. RIPPLY2, in turn, suppresses *Tbx6* expression, favoring the consequent cellular epithelialization and segmentation of the new somite (**Figure 3B**) [50, 57]. Analysis of various mouse mutants has highlighted many somitogenesis regulators [52]. Embryos depleted of wither WNT3A, T, Fibroblast growth factor receptor 1 (FGFR1) or lacking both Transcription factor 1 (TCF1) and Lymphoid enhancer-binding factor-1 (LEF1) fail to produce somites and exhibit severe trunk morphological defects [58, 59]. Moreover, embryos lacking TBX6 not only fail to form somites, but generate ectopic neural tubes at the expense of the paraxial mesoderm compartments [60, 61].

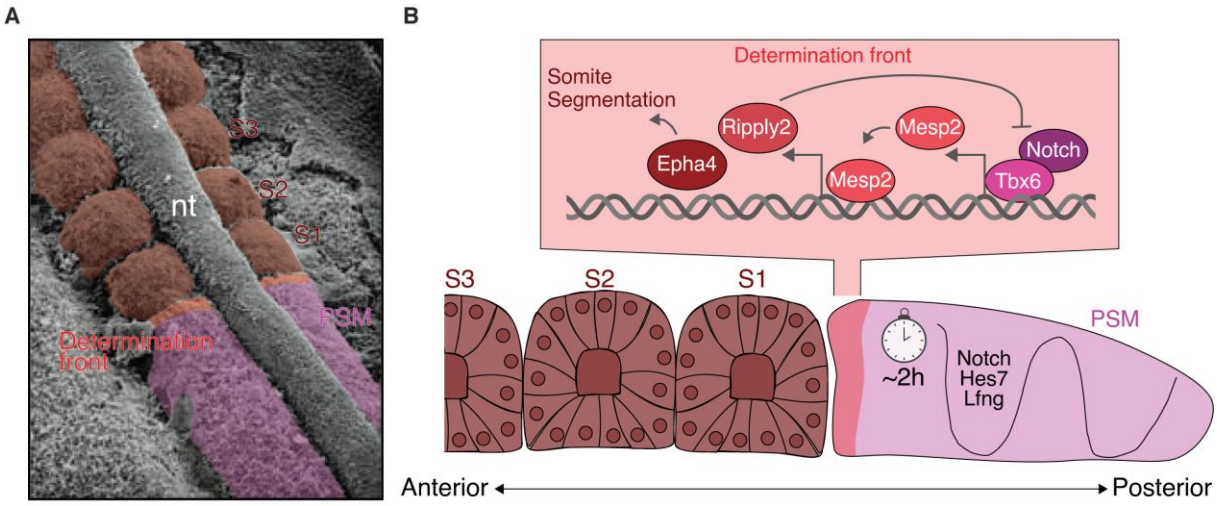

Figure 3. Somitogenesis

(A) Scanning electron micrograph of somite formation in a chick embryo. The PSM is highlighted in magenta, in red the determination front, and in brown the newly formed somites. The nt, laying in the middle of the two paraxial mesoderm rows, acts as a signaling center for mediolateral and dorsoventral tissue patterning. Somitic tissue organization is highly conserved in mammals. **(B)** Molecular mechanisms driving the rhythmic segmentation of the PSM in mouse embryos. In the PSM, the expression of NOTCH oscillates from the most posterior to the most anterior portion. This behavior results in the oscillating expression of other genes involved in somitogenesis, such as *Hes7* and *Lfng*. Roughly every two hours, following the so-called "*segmentation clock*," a high wave of NOTCH reaches the "*determination front*," a stripe of cells in the most anterior portion of the PSM, triggering the activation of the gene regulatory network responsible for somitic segmentation. High intracellular NOTCH in TBX6 expressing cells induces the expression of MESP2, which itself mediates the activation of *Ripply2*. RIPPLY2 inhibits *Tbx6* expression in a negative feedback loop, favoring cellular epithelialization and consequent somite segmentation. S, somite; nt, notochord; PSM, pre-somitic

mesoderm. Figure realized using Adobe Illustrator. Panel **(A)** scanning electron microscopy image is adapted from Kathryn W. Tosney.

Once somites have been formed from the most anterior part of the PSM, they undergo a series of patterning events necessary for cellular differentiation and specialization events. First, rostrocaudal compartmentalization is established, with the most posterior part of each somite expressing still high levels of NOTCH and resulting in the activation of UNC Homeobox (UNCX) expression. Concomitantly, the most anterior portion of each newly segmented somite upregulates T-box transcription factor 18 (TBX18) [50, 62, 63]. This somites' paired gene expression pattern is preserved and functionally relevant for the later dorsoventral patterning events [64]. With increased maturity, the dorsoventral somite patterning is also established. Cells belonging to the ventral portion of each somite, undergo EMT to form the mesenchymal sclerotome, which will later differentiate into the vertebral column and ribs. The forming sclerotome is characterized by the expression of transcription factors such as Paired box protein 1 and 9 (PAX1 and PAX9), and its transcriptional network is influenced by the combinatorial action of Sonic hedgehog (SHH) and Noggin (NOG) signaling gradients from the notochord and the neural tube [65-67]. The remaining dorsal part of the somite, also known as dermomyotome, will differentiate into the axial dermis (from the dermatome sub-compartment) and skeletal muscles (from the myotome sub-compartment) [68]. Initially, the dermomyotome is characterized by the expression of Paired box protein 3 (PAX3). While delaminating into the myotome, these cells lose *Pax3* expression and activate muscle-specific genes such as Myogenic differentiation 1 (*MyoD*) and Myogenic factor 5 (*Myf5*) [69, 70]. The somitic zone at the border between dermomyotome and sclerotome is known as syndotome, and will differentiate into the future tendons and cartilages [71].

1.1.3.3 Neural tube formation

The neural tube (NT) is the tissue precursor of the spinal cord and is an epithelialized tubular structure of distinct dual origins. The most anterior portion derives from the anterior neuroectoderm, while the most posterior part mainly originates from NMPs, as shown by different clonal tracing studies [37, 72, 73]. Once NMPs move anteriorly and acquire a neural-specific identity (see "*Neuromesodermal progenitors*" chapter for the molecular switch), they organize and form the neural plate (NP). The NP will subsequently fold in distinct steps during the primary neurulation event to acquire the NT epithelial morphology. This process is tightly

regulated at the cellular and molecular levels. More than 300 different genetic elements have been characterized to have a functional role in the NT closure [74]. Moreover, while interacting with the developing neural tube cells, the extracellular matrix has also been shown to play a critical role in ensuring neurulation (see "*Extracellular matrix in tissue morphogenesis*" chapter). Once the neural tube is closed, a series of molecular events, mainly guided by the SHH signaling coming from the ventrally located notochordal tissue, ensure the correct NT patterning along the dorsoventral (DV) axes [75]. SHH secreted from the notochord induces the activation of *Shh* itself in the most ventral part of the developing NT (floor plate) [76]. Therefore, a long-range morphogen gradient is established, directing the DV patterning and conferring positional information to the ventral progenitors [75, 77]. Concomitantly, WNT and BMP are secreted from the most dorsal portion of the developing tube (roof plate), while RA is secreted from the adjacent somites [75]. As an effect of this complex multi-signaling gradients interaction, a stratified gene expression pattern along the DV axis of the developing NT is established. The more ventral portion of the NT express, together with SHH and FOXA2, members of the Nk homeobox proteins (NKX), including NKX2.2, NKX2.9, and NKX6.1. On the other hand, the most dorsal part is characterized by the expression of transcription factors such as Paired box protein 6 and 7 (PAX6 and PAX7), but also members of the Developing brain homeobox protein family (DBX) [75]. These spatial gene expression modules, are delimited by discrete DV boundaries to ensure proper neuronal diversification. The proper "*gene expression code*" along the NT DV axes is, in fact, responsible for the specification of different neuronal classes, including the motoneurons and interneurons subtypes [75, 78].

1.1.3.4 Other tissues in the trunk

Other tissues and cell types originate in the developing trunk region during gastrulation and early organogenesis, including the gut tube, the LPM, the IM, and Primordial germ cells (PGCs) (**Figure 2**). All these tissues consolidate the integrity of the developing embryo by contributing to distinct organs.

The gut tube is specified from the definitive endoderm cellular layer at the mid gastrulation stage (see *"Gastrulation"* chapter for molecular details of the early events). By the end of gastrulation, the endoderm lineage is already partitioned along the AP axes as a result of different signaling gradients, including NODAL and WNT antagonist (anterior) and WNT, FGF, and BMP (posterior) [79, 80]. The combination of these signals results in a different gene regulatory network (GRN) for each segment along the AP axis. The caudal portion of the

developing gut tubes expresses high levels of CDX proteins and will develop into the hindgut. On the other hand, the rostral part of the tube expresses SOX2 and Hematopoietically-expressed homeobox protein (HHEX) and will give rise to the foregut [81, 82].

The LPM derivatives mainly contribute to the cardiovascular system and the heart. After gastrulation, the LPM consists of two bilateral layers of cells at the sides of the paraxial mesoderm [83]. LPM is specified together with the paraxial mesoderm from the ventral region of the PS as a result of high BMP signaling [84]. High levels of BMP are also necessary to maintain LPM lineage identity; in fact, after mesoderm induction from the PS, a BMP gradient is established along the mediolateral embryonic axis to ensure mesodermal diversification, with low levels of BMPs in the midline (paraxial mesoderm) and increasing levels in the lateral portions (LPM) [83]. As a result of these gradients, the LPM cells express a series of lineage markers, including Heart and neural crest derivatives expressed 1 and 2 (HAND1 and 2), Forkhead box F1 (FOXF1), and Paired Related Homeobox 1 (PRRX1).

The IM is the progenitor tissue of the kidneys and the gonads and is located between the paraxial and lateral plate mesoderms. Along the AP axis, the IM forms discrete domains that can generate distinct kidney tissues: pro-, meso- and metanephrons [85]. This patterning ensures the correct spatial localization of the developing kidneys. From the molecular perspective, the IM is characterized by the expression of transcription factors such as Paired box gene 2 and 8 (PAX2 and 8) and LIM homeobox gene (LIM1) [86, 87].

PGCs are first visible during gastrulation at E7.25 as a clump of around 40 cells in the posterior embryonic/extraembryonic border [88, 89]. They specify from the posterior epiblast from the combined action of WNT3 (high in the PS and nascent mesoderm) and BMP4 (high in the extraembryonic ectoderm (ExE)) [90]. WNT-induced transcriptional rewiring includes the activation of T and β-CAT, which directly bind and activate B lymphocyte-induced maturation protein-1 (BLIMP1) and PR/SET Domain 14 (PRDM14), PGCs lineage determinant genes [91]. Moreover, committed PGCs also express Developmental pluripotency-associated 3 (DPPA3), a factor required for PGCs identity maintenance [92]. From the most posterior portion of the tailbud, the PGCs enter the hindgut around E8.5 and start migrating anteriorly before colonizing the genital ridges at around E10.5 [90, 93].

1.2 The extracellular matrix

The extracellular matrix (ECM) is a critical component of various cellular and molecular processes and is necessary for proper tissue morphogenesis during development, given its ability to regulate proper cell movement, adhesion, and communication. The ECM is a highly heterogeneous assembly of different molecules that constitutes a mechanical and chemical environment surrounding the cell membranes (**Figure 4A**) [94]. Therefore, dissecting and understanding the composition and roles of its molecules have been challenging for decades. Technological advances in molecular biology and protein purification have made the characterization of the ECM structure and function more accessible, therefore providing a more detailed understanding of how different proteins contribute to the "*matrisome,*" the list of core ECM components of a specific matrix [95]. This analysis revealed one of the crucial properties of the ECM, namely its functional diversity [96]. In fact, ECM components are modular, and the different assemblies of these modules guarantee the realization of specialized functions during development and cell homeostasis. The main molecular components of the ECM can be divided into two main classes: fibrous proteins, such as collagen and elastin, and glycoproteins, including fibronectin, vitronectin, laminin, and proteoglycans (**Figure 4A**) [94, 97]. Their complex interaction in organized layers on the surface of cell membranes is defined as "*basement membrane*" or "*basal lamina*" and confers mechanical and instructive signals to the cells in close contact, especially during development and tissue morphogenesis [98].

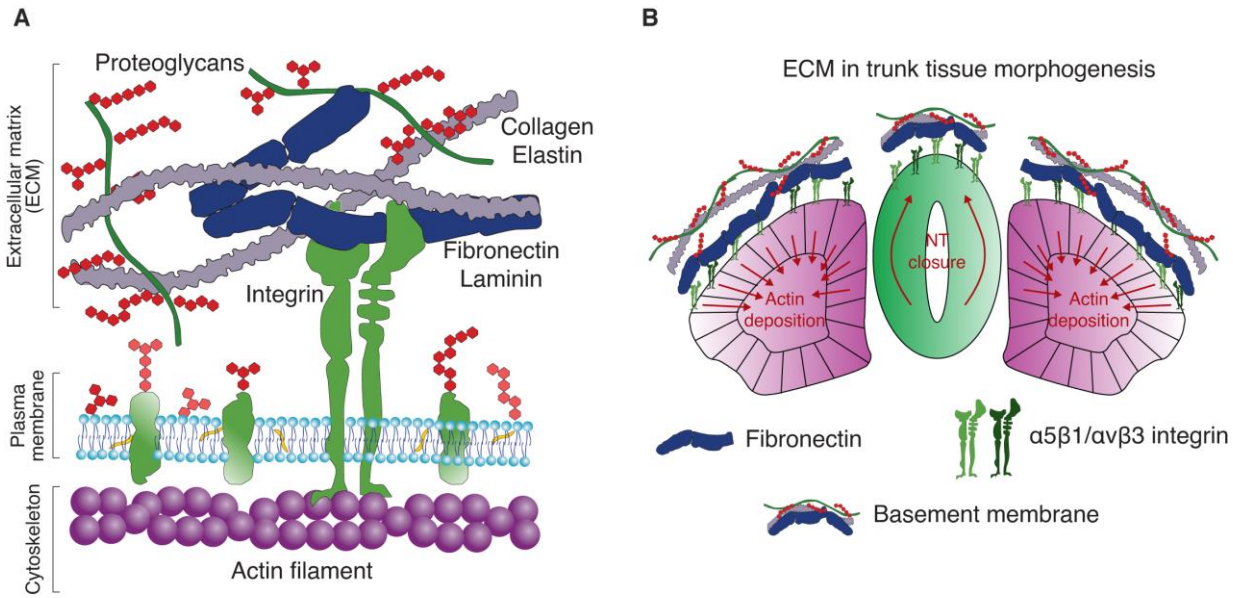

Figure 4. The extracellular matrix

(A) Schematic representation of the cell/matrix interface. The ECM is a heterogeneous assembly of various molecules, including fibrous and glycoproteins, that confers structural and chemical support to the cellular environment. Fibronectin and laminins organize in complexes with the proteoglycans to form the basement membrane, which contacts cellular receptors such as integrins. These interactions trigger various cellular responses through the plasma membrane, including re-organization of the cytoskeleton that results in cellular movement and migration. **(B)** Simplified depiction of ECM roles in trunk tissue morphogenesis. The basement membrane is crucial for both somitic segmentation and neural tube closure, as it induces apicobasal polarity via asymmetric actin deposition and cellular epithelialization. The interaction between fibronectin and α5β1/αvβ3 is known to be instrumental for proper morphogenesis of the neural and paraxial tissues. ECM, extracellular matrix; NT, neural tube. Figure realized using Adobe Illustrator.

1.2.1 Fibrous proteins

Collagens are the most abundant and predominant proteins in the ECM, with 28 different collagen types identified. They are organized in homo- or heterotrimers, with three α-chain peptides organized in a triple-helix, the structural hallmark to ensure collagen fibers functionality [99]. There are 40 different collagen α-chain peptides, and this variability confers a high degree of modularity and a wide range of functionalities to the collagen fibers [97]. The collagen triple helixes are first assembled in the endoplasmatic reticulum, then processed in the Golgi apparatus and finally exported in the extracellular environment, where another proteolytic processing step leads to the production of mature collagen molecules able to self-assemble into fibrils and ultimately organize in fibers [97, 100]. The assembly of collagen fibrils is fundamental for these molecules' function, given that the fibrillary structure provides binding

sites for other ECM and cellular proteins, guaranteeing mechanical robustness to the ECM [101].

If collagen fibers are relatively stiff, elastic fibers are extensible and deformable and therefore more suitable to preserve some tissues' structural integrity, including the skin, the blood vessels, and the lungs [94]. Elastic fibers are prevalently constituted by the protein elastin that, in contrast to collagens, is encoded by one gene [102]. Elastic fibers are also assembled step-wise but, in contrast to collagen fibers where the fibrils self-assemble, elastin processing is guided by helper proteins that mediate the positioning of the elastin monomers [103]. Collagen and elastic fibers represent the structural component of the ECM, but in order to exert their functions, they need to interact with binding proteins and receptors on the cell surface. Several classes of glycoproteins act as a "*glue*" and ensure functional cell-matrix interaction [97].

1.2.2 Glycoproteins

A single gene encodes for fibronectin and, after translation and processing in the cytoplasm, its product is secreted as a 220 kDa glycoprotein dimer in the extracellular environment, where it assembles into fibrillar structures [104]. Fibronectin structure and protein domains are crucial to its functions: it consists of three main domains of repeating units that contain binding sites for both other ECM molecules and cell surface receptors, such as integrins [95, 104]. In fact, it contains the tripeptide Arg-Gly-Asp (RGD) domain which is recognized by integrin receptors [105, 106]. Integrin-fibronectin interactions induce a series of cytoskeleton and structural rearrangements due to induced cellular contractility [106]. These mechanical forces induce further fibronectin conformational changes that allow clustering of more fibronectin molecules to ensure structural and functional continuity in the basement membrane of developing tissues [104, 107].

Vitronectin is a multifunctional glycoprotein component of the ECM. As for fibronectin, it also contains an RGD domain to interact with integrins, even though the two glycoproteins are recognized by different receptors [97, 108]. Moreover, it contains collagen-binding domains to guarantee proper cell-matrix bridging [97].

Proteoglycans are glycoproteins organized around a core protein that is covalently linked to glycosaminoglycans (GAGs). Different GAG chains associate with the core protein to produce different classes of proteoglycans, including heparin sulfate, chondroitin sulfate,

hyaluronan, and keratin sulfate [109, 110]. The saccharidic nature of the GAGs confers to proteoglycans their primary function. In fact, thanks to their long and negatively charged chains, they can bind water molecules to provide hydration and stress resistance to the ECM. For these reasons, they are highly abundant in the ECM of cartilages and the nervous system [111, 112].

Laminins represent a substantial basement membrane component and are involved in cell differentiation, proliferation, and migration. They consist of three peptide chains (α, β, and γ) coiled around each other to form a triple-helical domain [113]. Five α, four β, and three γ chains have been identified, and their combinations result in 16 different heterodimers with respective functions [114, 115]. Laminins can interact with cellular receptors, such as integrins and non-integrin receptors, and ECM structural proteins like collagens [94, 116, 117].

1.2.3 ECM in tissue morphogenesis and development

ECM components play pivotal roles in early and late development, given the mechanical and chemical support they provide to expanding tissues [118]. In particular, two main functions are exerted by the ECM during development: the first is an active function, where it signals the cells it interacts with to rearrange their cytoskeleton, change their shapes or induce cell polarity; the second is a passive and structural function, providing support for the developing tissues without directly instructing cellular behaviors [119]. It is often hard to distinguish between the two, and in many instances, both roles are played simultaneously. During gastrulation and early organogenesis, when tissues are not fully formed and organized yet, the active signaling function is crucial to ensure proper morphogenesis [36]. Integrins-ECM interaction has been shown to regulate several biological processes, including cell migration and acquisition of cell polarity, but also morphogenetic events such as vasculature development, somitogenesis, and neural tube formation (**Figure 4B**) [120-127].

During post-implantation development, somite formation is dependent on the correct basement membrane deposition at the ECM-cell interface [122]. In particular, α5β1 and αvβ3 integrins bind fibronectin to stimulate PSM cells to acquire cell polarity and undergo MET (**Figure 4B**) [128, 129]. In fact, mouse mutants for one or both of the α5 integrins or the fibronectin RGD domain display severe somitogenesis-associated defects, ultimately resulting in segmentation arrest [130-132].

Similarly, basement membrane assembly, particularly integrin-fibronectin interaction, is necessary for NT closure and morphogenesis (**Figure 4B**) [129, 133]. Despite undergoing NT

closure, mice lacking fibronectin exhibit a severe phenotype, characterized by a posteriorly kinked NT and anteriorly deformed headfold [128]. Interestingly, laminin KO mice also show neurulation defects, primarily in the cranial and occipital region [134, 135].

1.3 Mouse *in vitro* models of embryogenesis

Given the size and developmental modes of the mammalian embryo, which develops *in utero*, the complete picture of the molecular and cellular events shaping embryogenesis is missing. Since the derivation of the first mouse embryonic stem cell (mESCs) line, developmental biologists have been trying to model embryonic developmental transitions *in vitro*, intending to generate accessible, observable, and tractable systems to investigate embryogenesis (**Figure 5**) [136, 137]. The idea of building *in vitro* models of embryogenesis was based on the initial observations that teratocarcinoma cell lines retain the potential to aggregate and spontaneously differentiate *in vitro* in what were defined as embryoid bodies (EBs) in the ‘60s and ‘70s (**Figure 5**) [138, 139]. EBs generated from mESCs have been widely used in the ‘80s to study stem cell differentiation events, but the fidelity with which embryo-like developmental transitions were recapitulated was still unclear at that point [140, 141]. Only in the last fifteen years technological advances in synthetic biology, microscopy, and genomics have allowed the development and characterization of a multitude of mouse *in vitro* models of embryogenesis, recapitulating critical transitions of pre-, peri- and post-implantation development (**Figure 5**) [142-146]. These works build on the pioneering observation that EBs can establish an AP polarity and induce a localized PS-like domain under self-organizing signaling gradients, and that modulation of these gradients induces embryo-like molecular and cellular responses [147, 148]. Since then, the field has grown enormously, with the optimization of protocols that include extraembryonic cells, ECM components, and signaling molecules to mimic different developmental stages.

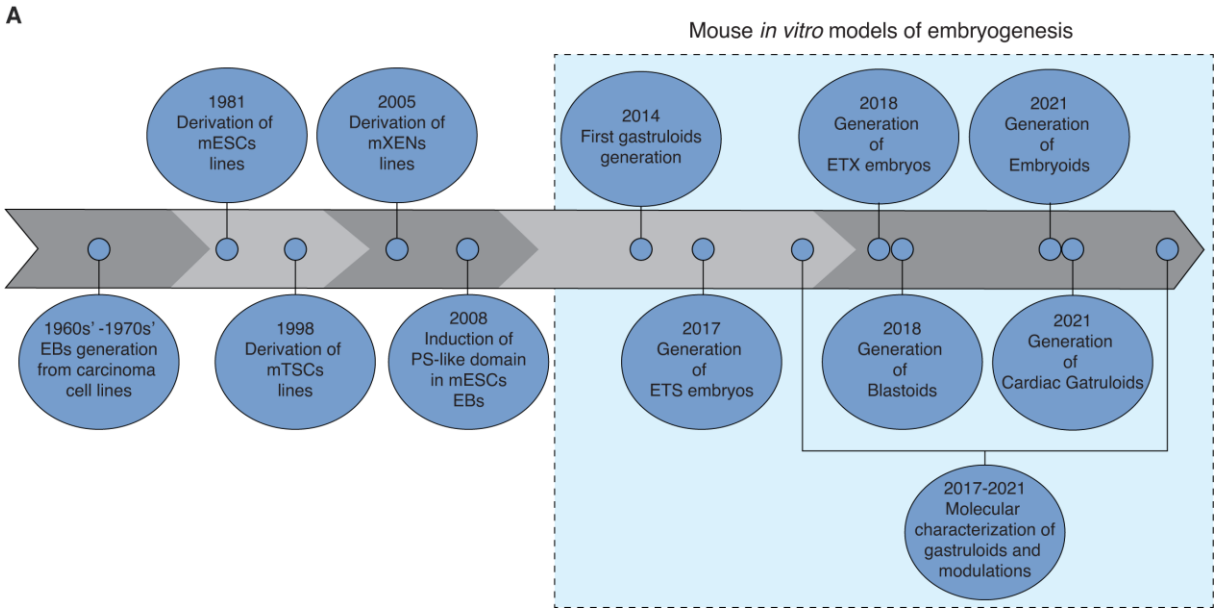

Figure 5. History of mouse *in vitro* models of embryogenesis

(A) Timeline of the main events that have led to the development and optimization of mouse *in vitro* models of embryogenesis. From the initial implementation of carcinoma cell lines in the '60s and '70s, through the first derivation of mESCs, mTSCs, and mXENs lines, to the recent optimization of relevant models of pre-, peri, and post-implantation development and their molecular characterization. EB, embryoid body; PS, primitive streak; mESCs, mouse embryonic stem cells; mTSCs, mouse trophoblast stem cells; mXENs, mouse extraembryonic endoderm cells; ETS, ESC- and TSC-derived embryos; ETX, ESC-,TSC- and XEN-derived embryos. Figure realized using Adobe Illustrator.

1.3.1 Pre- and peri-implantation models

Pre- and peri-implantation models of embryogenesis aim to recapitulate some of the main developmental transitions leading to blastocyst formation and the organization of the embryonic and extraembryonic tissues. Moreover, they aim to provide tractable systems to study signaling factors and cell-cell interactions in the early stages of development.

The "earliest" model in terms of developmental progression is blastoid, aggregates of mESCs and mouse trophoblast stem cells (mTSCs) that result in blastocyst-like structures under specific signaling conditions [149, 150]. mESCs are first aggregated in microwells in EB-like structures, over which TSCs are seeded. Once the two cell types start aggregating and interacting, in the presence of WNT and cyclic adenosine monophosphate (cAMP), blastoids spontaneously form a blastocoel with TSCs forming a cyst-like structure hosting a mESCs clump, reminiscent of the cellular organization of the developing blastocyst [149]. Blastoids have several cellular and molecular commonalities with the *in vivo* blastocysts, including size,

shape, the expression of pluripotency associate transcription factors in the ICM-like compartment, the expression of TE-like markers in the TSCs-derived cyst, and the potential to implant *in utero* (even though development does not progress beyond decidualization) [149]. The accessibility of this model, combined with signaling molecules modulation, have revealed critical morphogenetic gradients and their genetic interaction with transcription factors involved in the embryonic architecture of the early blastocyst, including the roles of BMP and NODAL in the regulation of the trophectoderm proliferation, epithelialization, and GRN establishment [149]. The complexity of the blastoid model was further increased in a following work that implemented the use of expanded potential mESCs (EPSCs) to induce the formation of the PrE layer in the blastocyst-like structures [151, 152].

Concomitant with the development of blastoids, ESC- and TSC-derived embryos (ETS embryos) have been derived from aggregates of mESCs and mTSCs combined in an extra-cellular 3D matrix surrogate (Matrigel) (Harrison et al 2017; Harrison et al 2018; Benton et al 2014). This model, built on a previous protocol able to induce epiblast-like rosettes in mESCs aggregates embedded in Matrigel, mimics the developmental transitions of the peri-implantation embryo, from E5.0 to E6.5 [153-155]. ETS embryos develop in a basal medium supplemented only with FGF4 to maintain TSCs in a proliferative state, stressing the importance of the ECM in the induction of self-organizing stem-cell properties. Matrigel is in ETS embryos mimicking the PrE layer surrounding the natural embryo, providing structural and chemical support. The striking spatial and cellular organization of ETS embryos recapitulates the one of the natural embryo, with morphogenetic steps leading to embryo cavitation, activation of embryo-like molecular programs, basement membrane deposition, and generation of morphogen gradients [154]. These processes result in ETS symmetry breaking, the induction of a PS-like mesodermal domain, and, under BMP stimulation, the formation of PGC-like cells [154]. Therefore, in less than 120h of *in vitro* development, ETS embryos can recapitulate the spatiotemporal organization of the developing peri-implantation embryo.

The derivation and maintenance of extraembryonic endoderm stem cells (XEN) led to the substitution of matrigel as an ECM surrogate in ETS embryo in order to mimic the presence of the PrE and VE during *in vitro* embryogenesis [156-158]. This modification results in the self-assembly of the three stem cells types (ES, TS, and XEN) in so-called ETX embryos [158]. ETX embryos recapitulate a peri-implantation embryo's cellular composition and organization, with the XEN cells organizing in an outer layer of cells surrounding the developing epiblast and TE-like compartments. Moreover, ETX embryos undergo symmetry breaking and EMT

events without the need for external signaling molecules, indicating the prominent role of the extraembryonic endoderm to instruct the neighbor epiblast [158]. Strikingly, ETX embryos can undergo gastrulation-like events, including the axial specification of nascent mesoderm and the subsequent specification and allocation of the definitive endoderm layer [158]. Transcriptomic analysis also revealed the high similarities between ETX and natural embryos.

Pre- and peri-implantation models of embryogenesis faithfully recapitulate the early stages of mouse embryogenesis but fail to progress in development beyond the early post-implantation stage. Therefore, post-implantation models of development have been optimized, mainly building on the gastruloid model.

1.3.2 Gastruloids

Gastruloids represent the most advanced *in vitro* model of embryogenesis in terms of developmental stage. In fact, during the 120-168 hours of development in a dish, they can recapitulate critical aspects of gastrulation and early organogenesis, including the formation of the three germ layers, the establishment of the three embryonic body axes, and the formation of spatially discrete domains of gene expression reminiscent of the developing mid-gestational embryo [159]. Gastruloids were originally generated almost a decade ago as small aggregates (150-300 cells) of mESCs that under specific signaling conditions self-organize and form elongating structures displaying post-occipital embryonic cellular complexity (**Figure 6**) [160-162]. The original protocol implies an initial 48h aggregation phase, where mESCs aggregates undergo lineage priming and enter an epiblast-like stage [160-162]. Subsequently, dominant WNT signaling is provided for 24h with the addition of CHIRON (CHIR99021) in the culture medium to induce the generation of a PS-like domain and the localized expression of T, necessary for the realization of gastrulation-like events [162, 163]. The signaling stimulation is then withdrawn around 72h after aggregation, and gastruloids self-organize and elongate for additional 48 to 96 hours, to reach a cellular and molecular complexity reminiscent of an E8.5-E9.5 developing embryo.

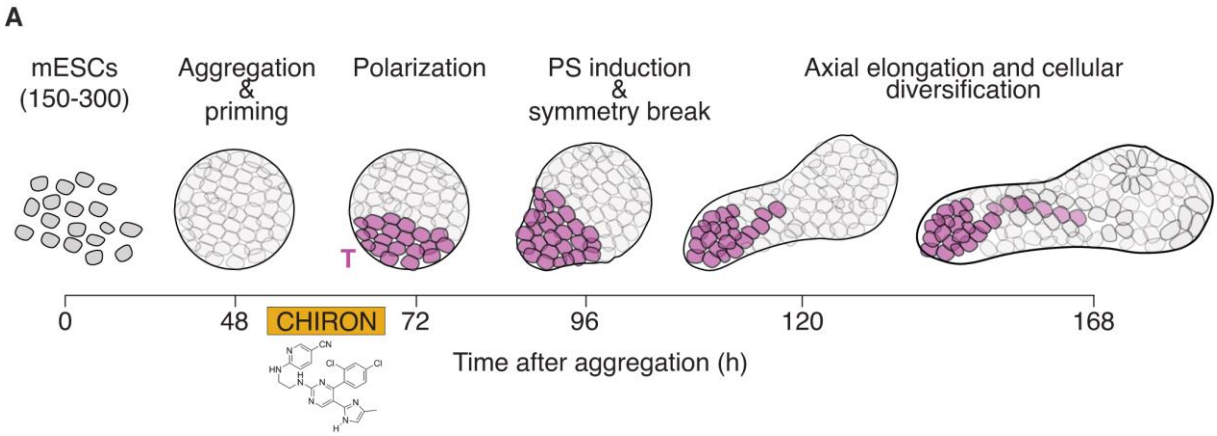

Figure 6. Gastruloids development

(A) Schematic representation of the cellular and morphological events during mouse gastruloids *in vitro* development. mESCs (150-300) are seeded in individual wells and left aggregating for 48h in the "*aggregation and priming*" phase. After aggregation, a 24-hour long pulse of CHIRON induces the polarization of the cell aggregates, with the localized expression of T in the perspective posterior domain (magenta). Around 96h after aggregation, gastruloids break their symmetry and induce a PS-like domain and elongate. From 96h to 168h, diverse cellular types emerge and organize along the three main body axes in a way reminiscent of the natural embryo. mESCs, mouse embryonic stem cells; PS, primitive streak; CHIRON, CHIR 99021. Figure realized using Adobe Illustrator. Schematic inspired from an idea of Vikas Trivedi.

The initial works developing and using gastruloids revealed their ability to undergo embryo-like movements and collective cellular behaviors such as symmetry breaking, axial elongation, AP polarization, and three germ layers specification [161-163]. Interestingly, these events occur in the absence of extraembryonic tissues and do not depend on their signaling stimulations (such as BMPs), indicating their dispensability to successfully realize gastrulation-like events [161]. Detailed molecular characterization with time-resolved RNA-sequencing revealed that, during their maturation, gastruloids mimic the transcriptional changes occurring in the post-occipital embryo, including the establishment of *Hox* collinear gene expression domains along the AP axis [163, 164]. On top of the AP axis, gastruloids have been shown to establish also the DV and mediolateral axes [163].

Building on these initial protocols and characterizations, in recent years, many laboratories have leveraged this technology to reveal the developmental plasticity of gastruloids to induce different cellular fates depending on the culture conditions and signaling stimulations, in an attempt to model more advanced embryonic structures *in vitro*. In the presence of WNT inhibition (XAV939), gastruloids elongate but develop anterior neural domain, as demonstrated

by the localized expression of OTX2 [165]. Interestingly, the development of anterior neural domains has been recently reported in gastruloids co-cultured with extraembryonic stem cells (TS or XEN), in line with the known role of extraembryonic tissues in the induction of brain development [166, 167]. A recent work has also detailed the developmental potential of gastruloids for gut-like structures generation, including the epithelialization events occurring while the tube is forming and the spatial and temporal activation of the GRN governing endoderm induction and maturation [168]. When exposed to well-known cardiogenic factors such as Vascular endothelial growth factor 165 (VEGF), basic Fibroblast growth factor (bFGF), and Ascorbic acid (AA), gastruloids upregulate a cardiac-specific gene expression program and develop an anterior beating heart-like domain and primordial vasculature system [169]. Interestingly, the generated heart-like domain contracts at an embryo-like pace, and this contraction is dependent on intracellular calcium exchanges, highlighting its physiological relevance [169]. Advances in single-cell RNA-sequencing technologies (scRNA-seq), combined with high-resolution imaging, have provided further evidence of the embryonic complexity of gastruloids [168-170]. In fact, mature gastruloids comprise the development of the three embryonic germ-layers and further show the presence of NMPs, PSM, somitic, spinal cord, endothelial, gut, and PGCs-like cell populations [170]. An active segmentation clock also controls the generation of somite-like cells in gastruloids in a temporally controlled and embryo-like fashion [170]. Interestingly, if embedded in 10% percentage Matrigel, gastruloids develop a string of somite-like structures along the AP axis [170].

A more recent work has also drastically modified the gastruloid protocol to achieve an even more complex embryo-like organization by simultaneously inducing both an anterior and a posterior domain [171]. This protocol is based on merging two distinct aggregates that have been differently exposed to signaling molecules to create a dual signaling center and, consequently, a controlled internal morphogenetic gradient [171]. This strategy results in structures (termed "*embryoids*") that develop both a posterior PS-like domain and an anterior brain-like structure. Interestingly they form a concrete axial mesoderm structure (notochord) which is absent in gastruloids [171].

2 AIM OF THE PROJECTS

In vitro models of embryogenesis represent potent tools to recapitulate some of the developmental transitions occurring *in utero* at the cellular and molecular level, with the advantage of allowing tractable, reproducible, and high-throughput investigation. For example, gastruloids undergo symmetry breaking, axial elongation, and patterning of their cellular domains along the three embryonic body axes. Moreover, these models react to external stimuli in a way that is reminiscent of the natural embryo response. Despite these recent advances, gastruloids deviate from the natural embryo in many aspects, including the lack of proper tissue morphogenesis. As such, in gastruloids, we can recover the presence of cells with a somitic transcriptional identity but without them organizing into somites. Similarly, neural and endodermal cells get specified, but they do not organize in epithelial tubular structures like the gut or neural tubes. The correct morphogenetic organization of tissues during development is pivotal for their consequent functionality and therefore represents a critical feature lacking in current post-implantation models limiting their implementation in multidisciplinary studies.

This project aimed to overcome these barriers by developing and characterizing a gastruloid-based *in vitro* model to recapitulate the morphogenetic events that embryos undergo *in utero*. To this end, in the first project, I queried the effect of an ECM surrogate (Matrigel) in the induction of tissue morphogenesis in gastruloids. Next, I focused on benchmarking the obtained models' similarities to the *in vivo* embryo by performing an in-depth molecular, cellular and functional characterization with the ultimate goal of providing the molecular basis of the unlocked morphogenetic potential. In the second project, I aimed to provide a detailed step-by-step protocol to reproduce the newly developed experimental strategy, allowing laboratories worldwide to implement our optimized model.

3 RESULTS

3.1 Paper 1: *Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites*

Authors:
Veenvliet JV*, **Bolondi A***, Kretzmer H, Haut L, Scholze-Wittler M, Schifferl D, Koch F, Guignard L, Sampath Kumar A, Pustet M, Heimann S, Buschow R, Wittler L, Timmermann B, Meissner A, Herrmann BG.
* Veenvliet and Bolondi contributed equally to this study.

Published in ***Science***, 11th Dec 2020, Vol 370, Issue 6522
DOI:10.1126/science.aba4937

Personal contribution:
Together with my shared first author (Veenvliet JV) I designed all the experiments and analysis, interpreted the data, prepared the visualizations and figures, and wrote the manuscript. In detail, I performed experiments that contributed to all main and supplemental figures during the initial project realization as well as during the peer review process. Together with Kretzmer H and Veenvliet JV, I designed the computational analysis and interpreted the results. First, second (revised) and third (final) version of the manuscript, including figures, were written by me together with Veenvliet JV, Meissner A and Herrmann BG. Detailed contributions of other co-authors are available in the publication document and at page 35 of this thesis under the section *Author contributions*.

Link to the original publication:
https://doi.org/10.1126/science.aba4937

Science
AAAS

Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites

Jesse V Veenvliet[1,†*], Adriano Bolondi[2,†], Helene Kretzmer[2,%], Leah Haut[1,2,%], Manuela Scholze-Wittler[1], Dennis Schifferl[1], Frederic Koch[1], Léo Guignard[3], Abhishek Sampath Kumar[2], Milena Pustet[1], Simon Heimann[1], Rene Buschow[4], Lars Wittler[1], Bernd Timmermann[5], Alexander Meissner[2,6,7,8,*], Bernhard G Herrmann[1,9,*]

[1] Dept. of Developmental Genetics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany
[2] Dept. of Genome Regulation, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany
[3] Max Delbrück Center for Molecular Medicine and Berlin Institute of Health
[4] Microscopy and Cryo-Electron Microscopy, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany
[5] Sequencing Core Facility, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany
[6] Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[7] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[8] Institute of Chemistry and Biochemistry, Freie Universität Berlin, 14195 Berlin, Germany
[9] Institute for Medical Genetics, Charité - University Medicine Berlin, Campus Benjamin Franklin, Hindenburgdamm 30, 12203 Berlin, Germany

† These authors contributed equally
[%] These authors contributed equally

[*] Corresponding authors, veenvlie@molgen.mpg.de, meissner@molgen.mpg.de, herrmann@molgen.mpg.de

Short Title: Trunk-like structures generated *in vitro*

One Sentence Summary: A platform for generating trunk-like structures with precursors of spinal cord, bone and muscle from stem cells in a dish

Post-implantation embryogenesis is a highly dynamic process comprising multiple lineage decisions and morphogenetic changes inaccessible to deep analysis *in vivo*. Here, we show that pluripotent mouse embryonic stem cells (mESCs) form aggregates that upon embedding in an extra-cellular matrix compound induce the formation of highly organized "Trunk-Like-Structures (TLS)" comprising the neural tube and somites. Comparative single-cell RNA-seq analysis confirms that this process is highly analogous to mouse development and follows the same step-wise gene-regulatory program. *Tbx6* knockout TLS develop additional neural tubes mirroring the embryonic mutant phenotype, and chemical modulation can induce excess somite formation. Trunk-like-structures thus reveal an unprecedented level of self-organization and provide a powerful platform for investigating post-implantation embryogenesis in a dish.

Vertebrate post-implantation development comprises a multitude of complex morphogenetic processes resulting from self-organization of stem cells and their descendants shaping the embryonic body plan *(1)*. Recently developed stem cell models represent powerful platforms for deconstructing the dynamics of these processes *in vitro (1,2)*. The most advanced models in terms of developmental stage accomplished to date are gastruloids, self-organizing mESC aggregates forming elongating structures comprising post-occipital embryo derivatives of all three germ layers, but lacking proper morphogenesis *(1,3)*. *In vivo*, the extracellular matrix (ECM) provides chemical signals and exerts mechanical constraints via the basement membrane, which has a critical role in tissue morphogenesis *(4,5)*. *In vitro*, matrigel can serve as an ECM surrogate, and embedding of gastruloids in 10% matrigel allowed the formation of a string of somite-like structures with anterior-posterior polarity *(6)*. Lower percentages of matrigel facilitated complex morphogenesis in organoids *(7)*.

Matrigel embedding drives trunk-like morphogenesis

To achieve more advanced embryo-like morphogenetic features in gastruloids we employed embedding in matrigel under various media conditions (**Fig. 1A**). To facilitate high-throughput characterization and quantification, we generated mESCs with *T::H2B-mCherry* (hereafter T^{mCH}) and *Sox2::H2B-Venus* (hereafter $Sox2^{VE}$) reporters, marking the mesodermal (ME) or neural (NE) lineage respectively (**Fig. S1A**). Embedding of 96h aggregates in 5% matrigel resulted in segmentation in the T^{mCH+} ME domain and formation of a $Sox2^{VE+}$ neural tube like structure (**Fig. 1A-C; Fig. S1B,C; Movie S1,2**). The vast majority of structures (hereafter referred to as TLS for **T**runk-**L**ike-**S**tructures) elongated and formed a T^{mCH+} pole at the posterior end, with segmentation occurring in about half the TLS (**Fig. 1D**). Whole-mount *in situ* hybridization for Tcf15 and Uncx confirmed somite identity and revealed embryo-like anterior-posterior polarity (**Fig. 1E**) *(8)*. In 61% of the TLS bilateral somites were observed (**Fig. S1D**). Additional WNT activation using CHIR99021 (hereafter TLS^{C}) or combined with BMP inhibition by LDN193189 (hereafter TLS^{CL}) improved the physical separation of neighboring segments without affecting T^{mCH+} pole formation or elongation (**Fig. 1A-D; Fig. S1E-H**), and resulted in an excess of segments at the anterior end, arranged like a “bunch of grapes” (**Fig. 1B,C,F**) *(9)*. Moreover, the ME domain expanded at the expense of the NE compartment, with apparent disorganization of the posterior end and neural tissue (**Fig. 1B; Fig. S1G,H**), a phenotype explainable by shifting the lineage choice of neuromesodermal progenitors (NMPs) - bipotent cells giving rise to both post-occipital NE and ME -, towards ME due to dominant WNT signaling, which has not been observed before *in vivo (10)*. In

addition, T^{mCH+}/$Sox2^{VE+}$ putative NMPs were reduced as confirmed by flow cytometry (**Fig. S1I**).

Phalloidin and N-cadherin staining demonstrated that the cells of the neural tube and somites of TLS, TLS^{C} and TLS^{CL} show proper apico-basal polarity, a characteristic of epithelial tissues, with F-actin and N-cadherin accumulating at the apical side **(Fig. 1G-H; Fig. S2A,B**). In contrast, in gastruloids F-actin accumulation appeared random and organized epithelial structures were not detected (**Fig. S2C**). Whole-mount immunofluorescence analysis of FOXA2 and SOX17, transcription factors characteristic for endoderm, revealed gut formation in a subset of TLS (**Fig. S3A-D**). Cells at the posterior base were SOX17-negative, but co-expressed FOXA2 with high levels of T^{mCH} (**Fig. S3B**). Thus, our data show that embedding in matrigel is both necessary and sufficient to drive complex, embryo-like tissue morphogenesis of the three embryonic germ layers.

We next performed a detailed morphometric analysis of trunk-like-structures and their substructures. The data demonstrate remarkable reproducibility of the three protocols with respect to size and shape of the whole-structure, somites and neural tube, whereas the gut-like-structure shows more variation (**Fig. S3E; Fig. S4A-D**). Time-resolved whole-structure morphometry showed similar morphogenetic changes over time (**Fig. S4B**). TLS in general were bigger than gastruloids (**Fig. S4A**). Comparisons of the different TLS protocols revealed that TLS^{C} and TLS^{CL} were slightly bigger than TLS and formed more somites, whereas the neural domain of the former was reduced in length and narrowing towards the anterior end (**Fig. S4A-C**). In all protocols, somites were similar in shape, but smaller than their embryonic counterparts (**Fig. S4D**).

To assess if the segmentation clock, an oscillator involved in somitogenesis is active in TLS, we performed live imaging *(11)*. In line with recent observations, we found that segmentation occurs in a rhythmic fashion at an embryo-like pace in all three TLS conditions (**Fig. S5A-C**; **Movie S3-5**) *(6)*. Remarkably, TLS^{C} and TLS^{CL} show consecutive formation of multiple somites (**Movie S6-9**).

Transcriptional characterization highlights selective responses to chemical modulation

To characterize the structures in detail at the molecular level we performed RNA-seq analysis (**Fig. S6A**), and found that TLS model the post-occipital embryo, similar to gastruloids (**Fig. 2A**) *(3)*. Compared to TLS, both TLS^{C} and TLS^{CL} showed significant upregulation of genes involved in (pre)somitic development (e.g. *Tbx6*, *Msgn1*, *Hes7 (8,10,12)*) at the expense of NE marker genes (e.g. *Sox1*, *Pax6*, *Irx3 (13)*), corroborating the flow cytometry and imaging results (**Fig. 2A; Fig. S6B-C**). The analysis of marker gene sets for NMPs, their direct descendants undergoing lineage choice (NMP ME & NMP NE), and for committed NE and ME cells substantiated this finding. Compared to TLS, TLS^{C} and TLS^{CL} displayed reduced expression of markers in all clusters including ME (**Fig. 2B**). In contrast, on average (P)SM specific markers were upregulated, while intermediate ME (IM) and lateral plate ME (LPM) markers were downregulated in TLS^{C}, and further reduced in TLS^{CL} (**Fig. 2B; Fig. S6D**) *(3,12)*, in line with the known role of WNT- versus BMP-signaling in PSM versus IM and LPM specification (**Fig. 2A; Fig. S6D**) *(8)*.

We next searched for gene expression differences possibly underlying improved physical separation of somites observed in TLS^{C} and TLS^{CL}. Among the most strongly upregulated genes compared to TLS was *Wnt6*, which acts as a somite epithelialization factor *in vivo* (**Fig. S6E**) *(14)*. In addition, multiple ephrins and their receptors, and other factors involved in somite epithelialization were upregulated (**Fig. 2C; Fig. S6E**) *(15,16)*. We observed expression changes of selected somite polarity markers and their inducers, in line with the role WNTs, SHH, BMPs and their antagonists play in somite compartmentalization *in vivo* (**Fig. S6F,G**) *(8,17)*. Overall our data show that exposure to CHIR or CHIR/LDN improved segment boundary formation, but affected somite cellular composition.

Tissue morphogenesis and remodeling genes are upregulated in TLS

Principal Component Analysis (PCA) indicates a high transcriptional similarity between gastruloids and TLS despite profound morphological differences (**Fig. 2D**). The latter are better highlighted by Gene-Set-Enrichment-Analysis (GSEA), which shows that matrigel embedding promotes tissue morphogenesis and remodeling (**Fig. 2E; Fig. S6H**). Zooming in on embryonic and tissue morphogenesis gene sets revealed markers of blood vessel development among upregulated genes suggesting the induction of capillary morphogenesis in TLS (**Fig. S6I**). GSEA also showed an enrichment of cell adhesion terms and overall a significant upregulation of corresponding marker genes in TLS (**Fig. 2E; Fig. S7A,B**). The most pronounced increase

was observed for integrins, transmembrane receptors mediating cell adhesion to the ECM important for e.g. neural tube formation, blood vessel development, and segmentation (**Fig. 2F; Fig. S7A,B**) *(18-21)*. Since binding of integrin to the glycoprotein fibronectin (FN1) and matrix assembly play an important role in somite and neural tube formation *(21,22)*, we asked if FN1, which is highly expressed in TLS, was retained by matrigel, as recently shown in 2D culture (**Fig. S7C**) *(23)*. Importantly, while FN1 protein clearly accumulated at the TLS-matrigel interface, that was not observed in gastruloids where FN1 likely diffuses into the medium (**Fig. 2G; Fig. S7D; Movie S10**). Taken together, our data suggest that the activation of morphogenetic programs by matrigel embedding is driven by extracellular matrix assembly involving integrins and fibronectin.

Single-cell RNA-seq demonstrates embryo-like dynamics of cell differentiation

Based on the above results, for an in-depth characterization we decided to focus on the TLS condition, as it produced the most *in-vivo*-like structures. After confirming reproducibility at the molecular level (**Fig. S8A-C**), we performed a time-resolved scRNA-seq analysis on a total of 20,294 post-processed cells sampled from TLS at 96, 108, and 120 hours (**Fig. S9A**). Clustering analysis identified 14 different cell states. The larger clusters corresponded to derivatives of the PSM and NE that flank putative NMPs, while smaller clusters comprised endoderm, endothelial cells and Primordial Germ Cell Like Cells (PGCLCs) (**Fig. S9B**). The main clusters organized into a continuum of states recapitulating spatio-temporal features of the developing post-occipital embryo (**Fig. 3A**). Across the three time points sampled, progenitor subtypes gradually decreased in favor of more mature neural and somitic cells as development progresses (**Fig. 3B; Fig. S9C**). As expected, putative NMPs co-expressing T and SOX2, or T^{mCH}, $Sox2^{VE}$ and CDX2 were located at the posterior end at 96h and 120h (**Fig. 3C**, **Fig. S10A-C**) *(10,24)*. TLS-NMPs thus display an *in-vivo*-like NMP signature and have the highest differentiation potential, as they give rise to differentiating cells of both neural and mesodermal lineages (**Fig. 3C**, **Fig. S10D-F**).

RNA Velocity analysis revealed neural and somitic trajectories rooted in the NMPs, further suggesting that TLS development recapitulates the developmental dynamics observed in the mid-gestational embryo (**Fig. 4A; Fig. S11A**) *(25)*. *In vivo*, NMPs and their descendants are arranged in an order of progressive maturity along the posterior to anterior axis *(8)*. Accordingly, ordering of TLS-derived cells along a pseudo-temporal trajectory showed that the somitic trajectory reflects the genetic cascade observed in the embryo (**Fig. 4B; Fig. S12A**).

For example, the trajectory from Fgf8^{+} NMPs and PSM via the determination front marked by Mesp2, to Meox1^{+} somites was faithfully recapitulated and the embryo-like spatial arrangement was confirmed by whole-mount *in situ* hybridization (**Fig. 4C**) *(8)*. Likewise, the genetic cascade from NMPs to neural progenitors reflected the *in vivo* differentiation path in space and time (**Fig. 4D**). Notably, subclustering of the neural cells demonstrated that TLS generate both dorsal and ventral neural subtypes, with dorsal subtypes being more prevalent (**Fig. S12B**) *(13)*. The analysis of *Hox* gene expression at consecutive time points showed *in-vivo*-like collinear activation, as described for gastruloids (**Fig. S12C**) *(3)*. To test if TLS somites establish cell states segregated along the dorsal-ventral (DV) and anterior-posterior (AP) somite axes *in vivo*, we reclustered all somitic cells. At 96h we detected two main groups, corresponding to the Uncx^{+} posterior and Tbx18^{+} anterior somite domains, in line with A-P polarity established during segmentation (**Fig. S13A,B**) *(6,8)*. At 120h, we found distinct clusters of Pax3^{+} (putative dorsal dermomyotomal) and Pax1^{+} (putative ventral sclerotomal) cells, and a small cluster of Lbx1^{+}/Met^{+} putative migratory limb muscle precursors (**Fig. S13C-F**) *(8,26)*. In addition, Scx^{+} syndetome cells were detected (**Fig. S13G**), and Uncx and Tbx18 expression were anti-correlated (**Fig. S13H**).

Primordial germ cells (PGC) specification in the embryo occurs between E6.0 and E6.5 via *T*-mediated activation of *Blimp1* and *Prdm14*, and at E7.5, nascent PGCs can be identified as a group of DPPA3^{+} cells in the posterior primitive streak, which later migrate along the hindgut to the gonads *(27,28)*. We assigned PGCLC identity using marker genes characteristic for PGCs and identified their location in TLS (**Fig. 4E; Fig. S14**). At 76h, roughly corresponding to stage E6.5, we detected T/Prdm14^{VE} co-expressing cell clusters (**Fig. 4E, Fig. S14A,B**). At 108h we found a group of Sox2$^{VE\text{-}high}$ cells that co-expressed DPPA3 (**Fig. 4E; Fig. S14C**). At 120h, Sox2$^{VE\text{-}high}$ cells were detected in contact with FOXA2^{+} cells, and DPPA3^{+} cells in contact with a T^{mCH+} gut-like epithelial structure (**Fig. 4E**; **Fig. S14D**). These data show that trunk-like structures contain cells displaying characteristics typical for PGCs.

TLS display a high complexity of cell states that match their in vivo counterparts

Single cell comparison of 120h gastruloids with 120h TLS identified different proportions of the major cell states (**Fig. S15A-C**). A more refined analysis revealed a higher complexity of cell states in TLS (**Fig. S16A-E, S17A-F, S18A-D**), and expression of later (more posterior) *Hox* genes suggests development into more advanced trunk stages (**Fig. S18E,F**). The comparison of TLS with TLS^{CL} showed that sclerotomal and more mature neural cells are virtually absent and somitic as well as endothelial cell identities are altered in the latter (**Fig. S15B,C, Fig. S16A,B, Fig. S17A-G, Fig. S18A,B**). Application of RNA Velocity confirmed that in TLS^{CL}, NMPs are highly biased towards the mesodermal lineage, while contribution to the neural lineage is diminished as compared to TLS-NMPs, which is further corroborated by upregulation of pPSM and downregulation of neural marker genes (**Fig. 15D,E**) *(8,13)*.

To investigate how close the cellular states identified in TLS resemble those in embryos, we mapped our single-cell transcriptomes to a scRNA-seq compendium of post-occipital embryonic cellular subtypes (**Fig. 5A**) *(29)*. The data revealed globally high accordance of TLS and embryonic cell states including characteristic marker genes, and pairwise comparison of mapped clusters identified only a small fraction of differentially expressed genes (**Fig. 5B,C; Fig. S19A-D**). Of note, PS- and early-NMP-like cells are exclusively present at 96h and replaced by late-NMP-like cells at 108h and 120h (**Fig. 5D**). Taken together, our scRNA-seq analyses demonstrate that TLS execute gene-regulatory programs in a spatiotemporal order resembling the embryo.

Knockout TLS display the embryonic mutant phenotype

Finally, to explore the utility of trunk-like structures further, we tested in a proof of concept experiment if gene ablation would reproduce the embryonic mutant phenotype. *In vivo*, loss of *Tbx6* results in trans-differentiation of prospective PSM and subsequent formation of ectopic neural tubes at the expense of PSM and somites (**Fig. 5E**) *(30,31)*. We deleted *Tbx6* from *Tbx6::H2B-Venus* ($Tbx6^{Ve}$) mESCs and generated TLS, which clearly failed to form somites even upon treatment with CHIR or CHIR/LDN (**Fig. 5E,F; Fig. S20A,B**). Quantitative PCR analysis on FACS-purified $Tbx6^{VE+}$ cells revealed upregulation of neural markers at the expense of (P)SM markers in $Tbx6^{-/-}$ cells, thus recapitulating the *in vivo* phenotype at the molecular level (**Fig. S20C**). Finally, whole-mount immunofluorescence analysis for SOX2 showed that $Tbx6^{-/-}$ TLS generated ectopic $Tbx6^{VE+}$ neural tubes, whereas gastruloids, TLS^{C}

and TLS^{CL} formed an excess of morphologically indistinct $SOX2^{+}$ tissue (**Fig. 5G; Fig. S20D-F**).

Discussion

Here we report the generation of embryonic stem cell derived trunk-like-structures (TLS), and demonstrate that TLS faithfully reproduce key features of post-occipital embryogenesis including axial elongation with coordinated neural tube, gut and somite formation as well as PGCLCs. Accordingly, genetic manipulation of TLS faithfully reproduced the morphogenetic and molecular changes observed *in vivo*. Thus, TLS will enable deeper analysis of the ontogeny of mutant phenotypes and provide additional tools for investigating morphogenetic mechanisms unavailable *in vivo*. We also envision that TLS^{C} and TLS^{CL} models may become important for testing current concepts of somitogenesis, for instance the hypothesis that somite size and shape are controlled by local cell-cell interactions *(9)*.

Mechanistically, our data highlight a crucial role for the ECM surrogate in unlocking the potential of *in vitro* derived mESC aggregates, though future efforts will have to address the exact functional contribution of individual components and biophysical properties (**Fig. S21**), possibly using modular synthetic 3D matrices *(7,32)*. Alternatively, the single-cell expression catalogue of TLS and gastruloids can provide initial guidance for further exploration of cell-cell and cell-matrix interactions and their control of embryonic architecture (**Fig. S7A,B; Fig. S22,23)**.

In conclusion, trunk-like structures provide a scalable, tractable, readily accessible platform for investigating lineage decisions and morphogenetic processes shaping the mid-gestational embryo at an unprecedented spatiotemporal resolution.

References and Notes:

1. M. N. Shahbazi, E. D. Siggia, M. Zernicka-Goetz, Self-organization of stem cells into embryos: A window on early mammalian development. *Science*. **364**, 948–951 (2019).
2. M. N. Shahbazi, M. Zernicka-Goetz, Deconstructing and reconstructing the mouse and human early embryo. *Nat. Cell Biol.* **20**, 878–887 (2018).
3. L. Beccari *et al.*, Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature*. **562**, 272–276 (2018).
4. I. Matsuo, R. Hiramatsu, Mechanical perspectives on the anterior-posterior axis polarization of mouse implanted embryos. *Mech. Dev.* **144**, 62–70 (2017).
5. H. K. Kleinman, D. Philp, M. P. Hoffman, Role of the extracellular matrix in morphogenesis. *Curr. Opin. Biotechnol.* **14**, 526–532 (2003).
6. S. C. van den Brink *et al.*, Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* (2020), doi:10.1038/s41586-020-2024-3.
7. J. A. Brassard, M. P. Lutolf, Engineering Stem Cell Self-organization to Build Better Organoids. *Cell Stem Cell*. **24**, 860–876 (2019).
8. J. Chal, O. Pourquié, Making muscle: skeletal myogenesis in vivo and in vitro. *Development*. **144**, 2104–2122 (2017).
9. A. S. Dias, I. de Almeida, J. M. Belmonte, J. A. Glazier, C. D. Stern, Somites without a clock. *Science*. **343**, 791–795 (2014).
10. F. Koch *et al.*, Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev. Cell*. **42**, 514–526.e7 (2017).
11. O. Pourquié, The segmentation clock: converting embryonic time into spatial pattern. *Science*. **301**, 328–330 (2003).
12. J. Chal *et al.*, Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nat. Biotechnol.* **33**, 962–969 (2015).
13. A. Sagner, J. Briscoe, Establishing neuronal diversity in the spinal cord: a time and a place. *Development*. **146** (2019), doi:10.1242/dev.182154.
14. J. von Maltzahn, N. C. Chang, C. F. Bentzinger, M. A. Rudnicki, Wnt signaling in myogenesis. *Trends Cell Biol.* **22**, 602–609 (2012).
15. M. Rowton *et al.*, Regulation of mesenchymal-to-epithelial transition by PARAXIS during somitogenesis. *Dev. Dyn.* **242**, 1332–1344 (2013).
16. L. Durbin *et al.*, Eph signaling is required for segmentation and differentiation of the somites. *Genes Dev.* **12**, 3096–3109 (1998).
17. C. Marcelle, M. R. Stark, M. Bronner-Fraser, Coordinate actions of BMPs, Wnts, Shh and noggin mediate patterning of the dorsal somite. *Development*. **124**, 3955–3963 (1997).
18. D. Jülich *et al.*, Cross-Scale Integrin Regulation Organizes ECM and Tissue Topology. *Dev. Cell*. **34**, 33–44 (2015).
19. J. H. Shawky, L. A. Davidson, Tissue mechanics and adhesion during embryo development. *Dev. Biol.* **401**, 152–164 (2015).
20. S. Koshida *et al.*, Integrinalpha5-dependent fibronectin accumulation for maintenance of somite boundaries in zebrafish embryos. *Dev. Cell*. **8**, 587–598 (2005).

21. E. Nikolopoulou, G. L. Galea, A. Rolo, N. D. E. Greene, A. J. Copp, Neural tube closure: cellular, molecular and biomechanical mechanisms. *Development*. **144**, 552–566 (2017).
22. G. G. Martins *et al.*, Dynamic 3D cell rearrangements guided by a fibronectin matrix underlie somitogenesis. *PLoS One*. **4**, e7429 (2009).
23. J. Lu *et al.*, Basement membrane regulates fibronectin organization using sliding focal adhesions driven by a contractile winch. *Dev. Cell*. **52**, 631–646.e4 (2020).
24. S. Amin *et al.*, Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche. *Cell Rep.* **17**, 3165–3177 (2016).
25. G. La Manno *et al.*, RNA velocity of single cells. *Nature*. **560**, 494–498 (2018).
26. M. Buckingham *et al.*, The formation of skeletal muscle: from somite to limb. *J. Anat.* **202**, 59–68 (2003).
27. S. Aramaki *et al.*, A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev. Cell*. **27**, 516–529 (2013).
28. M. Saitou, M. Yamaji, Primordial germ cells in mice. *Cold Spring Harb. Perspect. Biol.* **4** (2012), doi:10.1101/cshperspect.a008375.
29. S. Grosswendt *et al.*, Epigenetic regulator function through mouse gastrulation. *Nature*. **in press**, (2020).
30. D. L. Chapman, V. E. Papaioannou, Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature*. **391**, 695–697 (1998).
31. D. Concepcion *et al.*, Cell lineage of timed cohorts of Tbx6-expressing cells in wild-type and Tbx6 mutant embryos. *Biol. Open*. **6**, 1065–1073 (2017).
32. A. Ranga *et al.*, Neural tube morphogenesis in synthetic 3D microenvironments. *Proc. Natl. Acad. Sci. USA*. **113**, E6831–E6839 (2016).

Acknowledgments: We are grateful for the support and feedback received from members of the Herrmann & Meissner laboratories, in particular Stefanie Grosswendt, Zachary Smith and Atsuhiro Taguchi. We thank Dijana Micic & Judith Fiedler for animal care, Norbert Mages for assistance with (sc)RNA-Seq, Claudi Giesecke-Thiel & Uta Marchfelder for assistance with FACS, Thorsten Mielke & Beatrix Fauler for help with microscopy, Sandra Währisch, Fabian Tobor, Ipek Gassaloglu and Polly Burton for technical assistance, and Nikolaus Rajewsky (MDC/BIMSB) for providing access to the NanoString. **Funding:** J.V.V. was partly funded by an Alexander von Humboldt Fellowship. This work was supported by NIH grant HG006193 (A.M.) and the Max Planck Society. **Author contributions:** B.G.H. initiated the study; J.V.V. and B.G.H. conceived the project; J.V.V., A.M. and B.G.H. supervised the project. J.V.V., A.B. and L.H. designed, performed and quantified most experiments. H.K. performed bulk & sc-RNA-Seq computational analysis with help of J.V.V. and A.B. M.S-W., D.S., F.K., M.P. and A.S.K generated and derived the mESC reporter lines. L.W. performed tetraploid complementation. S.H. performed pilot experiments to optimize culture media. R.B. helped with image acquisition and performed image analysis with L.G., A.B. and J.V.V. B.T. supervised next-generation sequencing. J.V.V. drafted the first version of the manuscript. The final manuscript was written by J.V.V, A.B., A.M. and B.G.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability**: All data is available in the main text or the supplementary materials. Bulk and single-cell RNA-Seq data have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE141175. All computational code used in this study is available upon request.

List of Supplementary Materials:

Materials and Methods

Figures S1-S24

Movie S1-S10

External Data Tables S1-S10

References (*33-71*)

MDAR Reproducibility Checklist

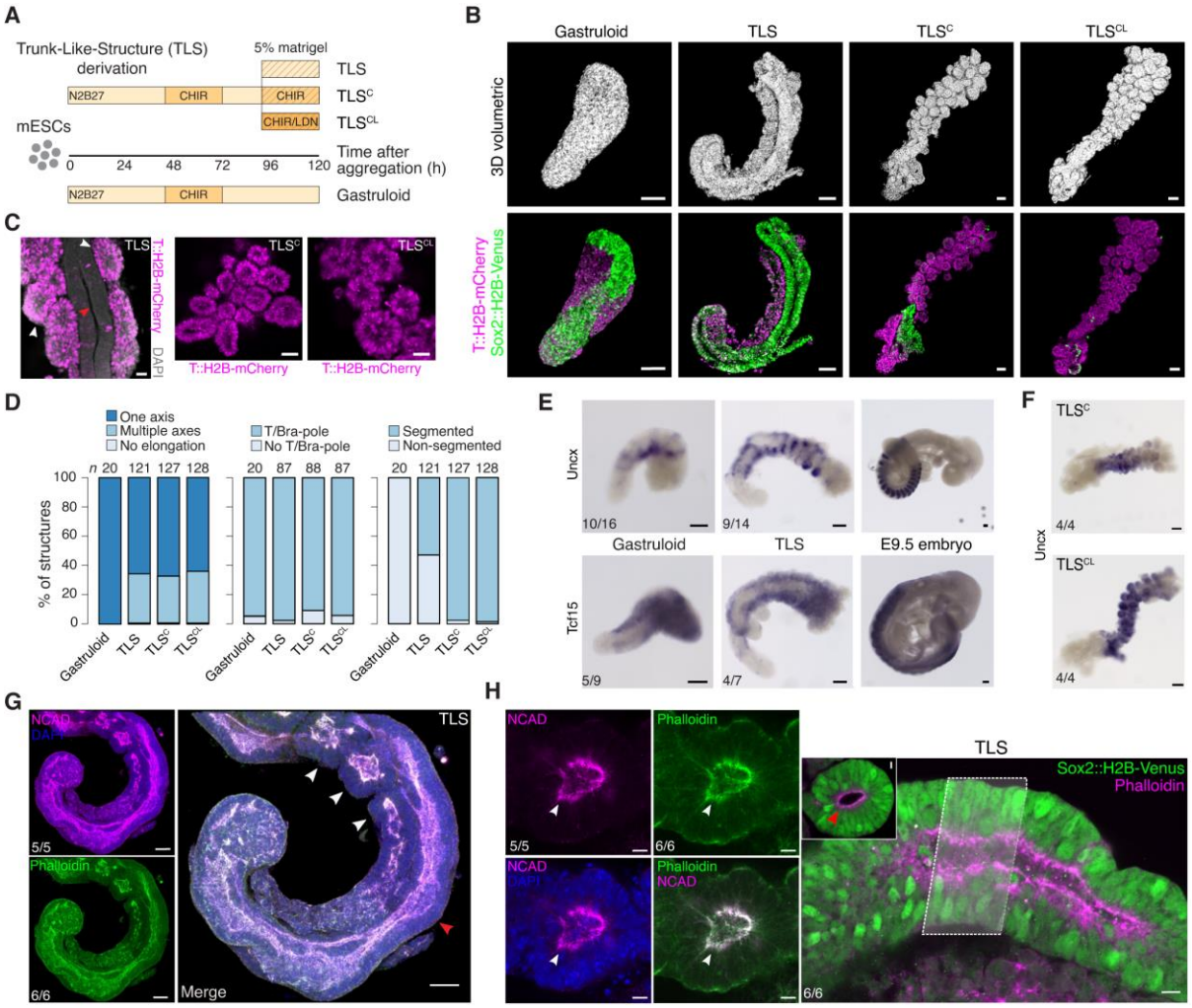

Fig. 1. Generation of Trunk-Like Structures with somites and a neural tube.
(**A**) Schematic overview: 200-250 mESCs were aggregated in ultra-low-attachment plates; *Wnt* agonist CHIR99021 (CHIR) was added between 48 and 72 hours (*3*). At 96h aggregates were i) cultured as gastruloids *(3)* ii) embedded in 5% matrigel (TLS), iii) treated with WNT signaling activator CHIR99021 (CHIR) alone (TLS^{C}) or iv) in combination with the BMP signaling inhibitor LDN193189 (TLS^{CL}). The two compounds have been reported to induce a (pre-)somitic mesoderm fate in 2D and 3D differentiation protocols (*8*). (**B**) 3D volumetric renderings (upper panel) and confocal sections (bottom panel) of gastruloids, TLS, TLS^{C} and TLS^{CL}. Scale bars 100µm. Each image is representative of at least ten biological replicates with similar morphology and expression patterns. (**C**) Segments in TLS are T^{mCH+} and positioned adjacent to the neural tube. In TLS^{C} and TLS^{CL} the segments are arranged in "bunches of grapes". Scale bars 25µm. Red arrowhead indicates neural tube, white arrowhead somites. (**D**) Quantification of morphogenetic features in gastruloids, TLS, TLS^{C} and TLS^{CL} (see Supplemental Information for scoring criteria). (**E**) Segments express somitic markers Uncx and Tcf15 as shown by whole-mount in situ hybridization. Note the characteristic stripy expression pattern of Uncx in TLS due to its restriction to the posterior somite half, whereas Tcf15 is expressed throughout the segments (as in the embryo). Scale bars 100µm. (**F**) In TLS^{C} and TLS^{CL} Uncx is detected throughout the segments, indicating loss of anterior-posterior polarity. Scale bars 100µm. (**G,H**) Confocal sections showing that cells of somites and neural tube display apical-basal polarity with NCAD and F-actin (Phalloidin) accumulating at the apical surface. Scale bars in panel **G** 50µm. Scale bars in panel **H** 10µm. White arrowheads indicate somites, red arrowheads indicate neural tubes.

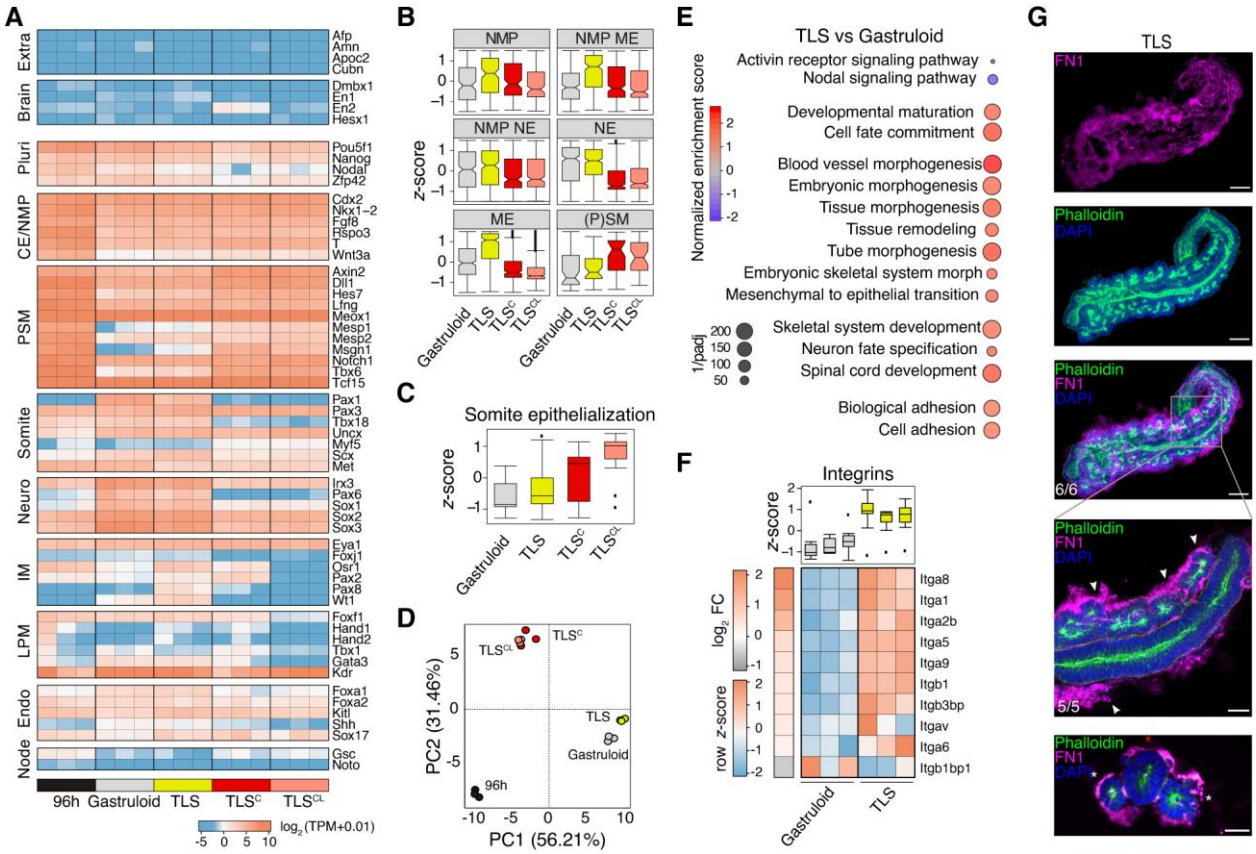

Fig. 2. Transcriptional characterization of trunk-like-structures

(**A**) Heatmap of log_2(TPM+0.01) expression (TPM, Transcripts Per Million) of selected genes associated with development of indicated embryonic structures in 96h aggregates, 120h gastruloids, 120h TLS, 120h TLS^C and 120h TLS^{CL}, as measured by RNA-seq. Replicates were derived from pools of independent biological samples (see **Fig. S6A** for exact experimental set-up). CE, caudal end, NMP, neuromesodermal progenitors, PSM, presomitic mesoderm, LPM, lateral plate mesoderm, IM, intermediate mesoderm. (**B**) Box plots showing distribution of marker genes for indicated cell types (average *z*-score). Boxes indicate interquartile range (IQR). Whiskers extends to 1.5xIQR from the hinge. Dots indicate outliers. Notches are centered on the median. List of genes used for each category and statistical analysis in **Data S1.** (**C**) Boxplot representing average *z*-score per column (pool of 3 replicates) for somite epithelialization factors (see **Fig. S6E** for individual genes) (**D**) PCA analysis of samples from (**A**) with color coding of individual samples (dots) as in (**A**). PC1 and PC2 represents the two components with highest percentage of explained variance. (**E**) Selected significant terms of Gene Set Enrichment Analysis (GSEA) enriched in TLS as compared to gastruloids of 120h. Full list of significant (FDR<0.05) terms is provided in **Data S2**. (**F**) Heatmap of scaled expression (row *z*-score) of integrins with significantly different expression (padj(FDR)<0.05) in gastruloids vs 120h TLS. Boxplot represents *z*-score per column (sample), with boxes indicating interquartile range, whiskers extending to 1.5xIQR from the hinge, dots showing outliers and central line representing median. Every column represents one of three biological replicates. Statistical analysis in **Data S1.** (**G**) 3D maximum intensity projection (top three images) and confocal section showing FN1 accumulation around TLS somites and neural tube (zoomed in image, white arrowheads). Phalloidin staining shows apical-basal polarity. Scale bar 100µm, 50µm for magnification. Bottom image: Lightsheet transversal section showing FN1 accumulation around the somites (white asterisk) and neural tube (red asterisk) in TLS. Scale bar 50µm.

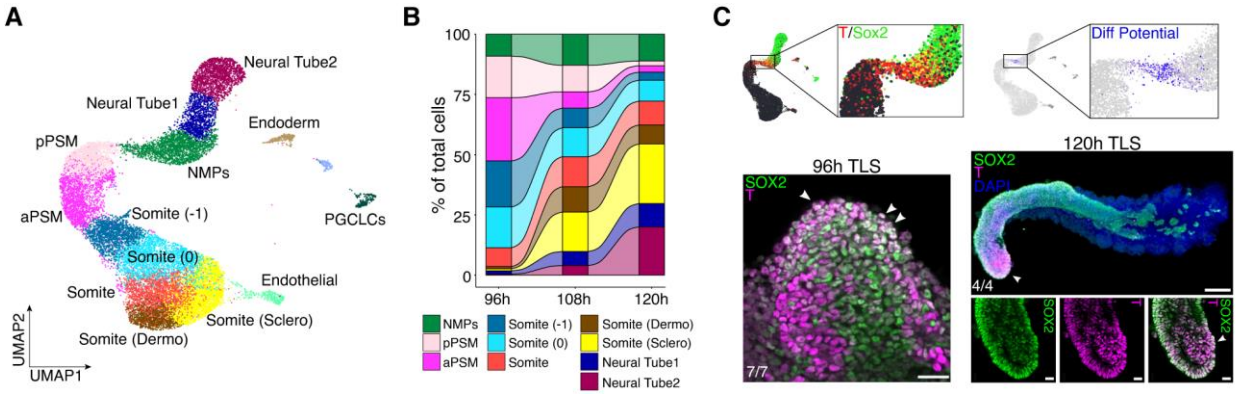

Fig. 3. Time-resolved single-cell RNA-Sequencing of TLS.

In total, 20,294 cells were sampled from 96, 108, and 120 hour TLS (see **Fig. S9A** for experimental set-up). (**A**) UMAP (Uniform Manifold Approximation and Projection) coloured by the fourteen clusters identified. (**B**) Alluvial plot of percentage of neuromesodermal progenitors (NMPs), posterior presomitic mesoderm (pPSM), anterior PSM, somitic and neural tube cells over time. (**C**) NMPs co-express Sox2 and T (top left panel, blending expression with blend.threshold=0.25), and are characterized by the highest differentiation potential (top right panel, see Supplementary Information for differentiation potential calculation). NMPs co-express T and SOX2 at 96h and 120h and reside at the posterior end of TLS (confocal sections, bottom left panel and magnifications, 3D maximum intensity projection, whole structure); white arrowheads, NMPs; red arrowheads, somites). Scale bars 50µm for 96h, 100µm for 3D maximum intensity projection, 20µm for magnifications.

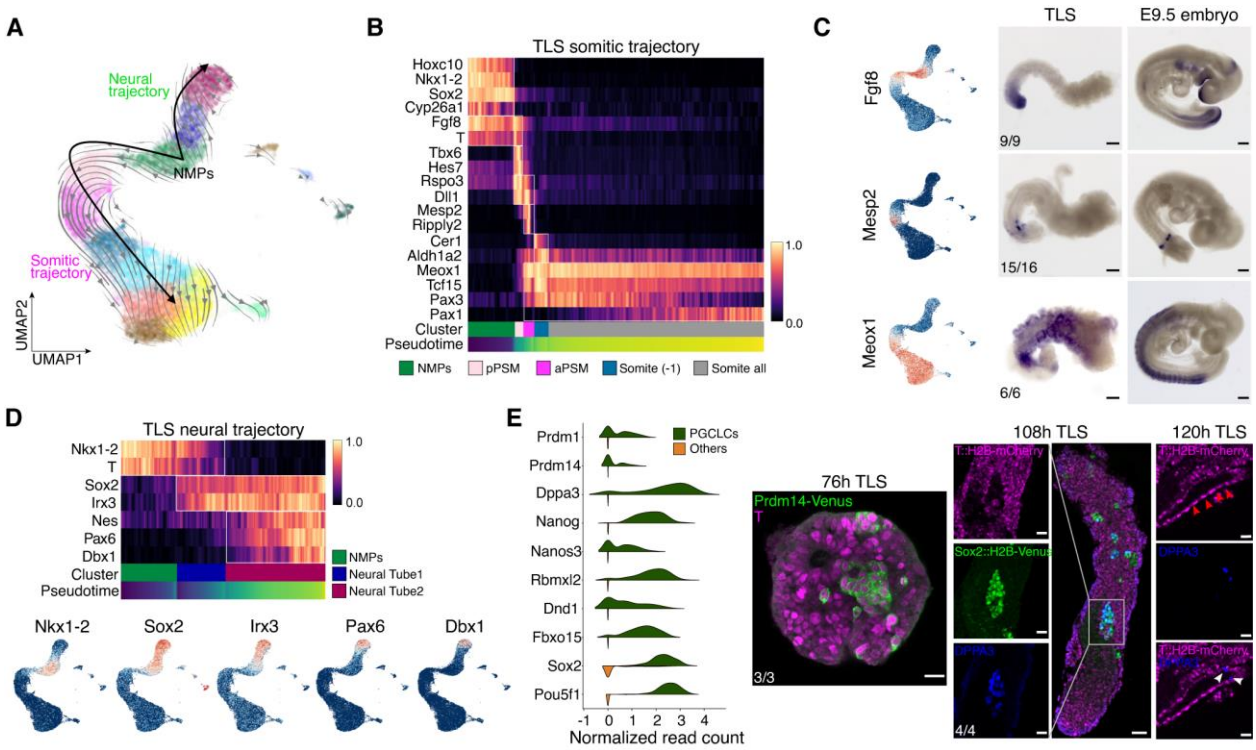

Fig. 4. TLS differentiation trajectory

(**A**) UMAP colored by identified clusters with trajectories inferred from RNA Velocity. Grey arrow flows represent calculated velocity trajectories. (**B**) Heatmap with scaled expression of genes involved in somitogenesis in 9,004 cells from 120h TLS rooted in NMPs and ordered by pseudotime. (**C**) UMAP coloured by expression of indicated genes (left panels), and whole-mount *in situ* hybridization for the same genes in TLS and E9.5 embryos (right panels). Numbers indicate the fraction of TLS with embryo-like expression. Scale bars TLS 100µm, embryo 200µm. (**D**) Heatmap with scaled expression of genes involved in neural development in 9,004 cells from 120h TLS rooted in NMPs and ordered by pseudotime (upper panel) and UMAP coloured by expression of indicated genes (bottom panel). (**E**) Split violin plots showing expression of marker genes for primordial germ cell like cells (PGCLC, left panel), and confocal sections of TLS showing PGCLCs specification dynamics: T/PRDM14^{VE} double positive at 76h, SOX2$^{VE\text{-}high}$/DPPA3^{+} PGCLCs at 108h, and DPPA3^{+} cells in close contact with the T^{mCH+} gut-like-structure at 120h. At 120h, Sox2$^{VE\text{-}high}$ cells in contact with FOXA2^{+} gut-like domain or DPPA3^{+} cells in contact with T^{mCH+} gut-like domain were observed in 7/9 TLS. Scale bars 20µm for 76h, 50µm for 108h left panel, 25µm for 120h. Red arrowheads indicate gut-like-structure, white arrowheads indicate DPPA3^{+} PGCLCs.

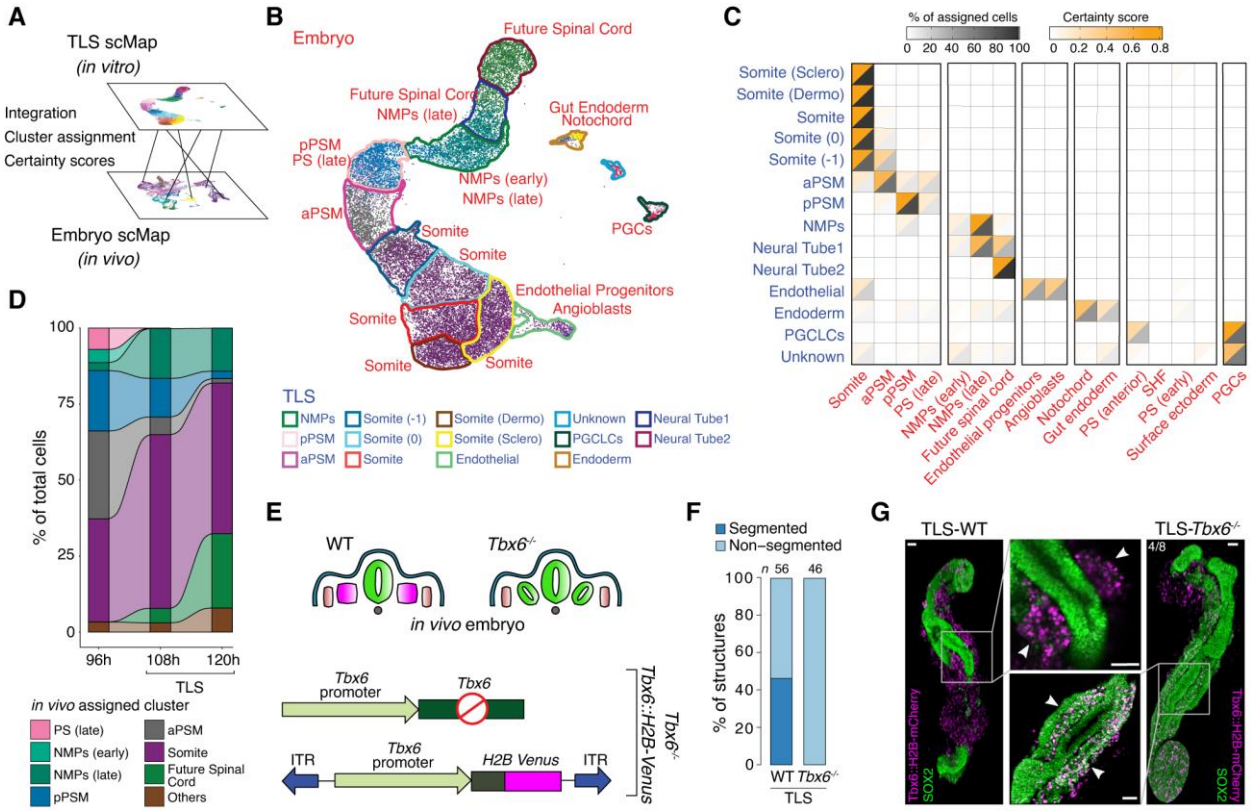

Fig. 5. TLS cell states are embryo-like and *Tbx6*$^{-/-}$ TLS recapitulate the embryonic knockout phenotype.

(**A**) Schematic of our comparative transcriptome analysis of TLS with post-occipital E7.5 and E8.5 embryos at the single-cell level. (**B**) TLS UMAP coloured by assigned embryonic cell type. TLS clusters are projected as corresponding coloured contours. Blue font, TLS clusters; Red font, embryo clusters. (**C**) Split heatmap with percentage of assigned cells (dark grey) and certainty score (orange) for TLS cells from the indicated cluster upon unbiased mapping to the *in vivo* counterpart. Font colour code as in (**B**). (**D**) Alluvial plot of percentage of cells assigned to indicated *in vivo* cluster in 96, 108 and 120h TLS. (**E**) Simplified schematics of *Tbx6*$^{-/-}$ *in vivo* phenotype and knockout/reporter constructs. (**F**) Quantification of segmentation phenotype in TLS-*Tbx6*$^{-/-}$. Data represent 3 different experiments performed with two independent mESC lines of each genotype. (**G**) Formation of ectopic neural tubes in TLS-*Tbx6*$^{-/-}$. Ectopic neural tubes are identified as SOX2^{+}/Tbx6^{VE+} tubular structures flanking the main SOX2^{+}-only neural tube. Green, SOX2; Magenta, Tbx6^{VE}. White arrowheads indicate Tbx6^{VE+} somites in WT, and Tbx6^{VE+}/SOX2^{+} ectopic neural tubes in *Tbx6*$^{-/-}$. Scale bars 50μm.

Science
AAAS

Supplementary Materials for

Mouse embryonic stem cells self-organize into trunk-like-structures with neural tube and somites

Jesse V Veenvliet†*, Adriano Bolondi†, Helene Kretzmer, Leah Haut, Manuela Scholze-Wittler, Dennis Schifferl, Frederic Koch, Léo Guignard, Abhishek Sampath Kumar, Milena Pustet, Simon Heimann, Rene Buschow, Lars Wittler, Bernd Timmermann, Alexander Meissner*, Bernhard G Herrmann*

† These authors contributed equally
*Corresponding authors. Email: veenvlie@molgen.mpg.de, meissner@molgen.mpg.de, herrmann@molgen.mpg.de

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Material and Methods

Animal work

All animal work was approved by the local authorities (LAGeSo Berlin, license numbers G0247/13, G0243/18 and G0243/18-SGr1). *T::H2B-mCherry/Sox2::H2B-Venus* embryos were generated via tetraploid complementation (*45*). For embryo isolation, mice were sacrificed by cervical dislocation and uteri were dissected in PBS. For the localization and validation of *Prdm14-Venus* Primordial Germ Cells *in vivo* all medium compositions are listed in **Data S3**. Zygotes were obtained by *in vitro* fertilization of B6D2F1 oocytes and Prdm14-Venus reporter sperm (*46*) (B6.Cg-Tg(Prdm14-Venus)1Sait/SaitRbrc:BRC) (No. RBRC05384), as previously described (*47*). Blastocysts were screened for expression of *Venus* reporter in the ICM, and were re-transferred in a clutch of 20-30 into the uterine horns of 2.5dpc CD1 pseudopregnant fosters. E7.5 embryos were dissected in cold M2 media and washed in PBS with 0.4% BSA.

Mouse embryonic stem cell culture

Except for the Prdm14-Venuse line (see above), all mouse embryonic stem cells (mESCs) used in this study were male and from an F1G4 background (*48*). mESCs were routinely maintained on 6cm plates (Corning 430166) gelatinized with 0.1% gelatin (1:20 dilution of 2% gelatin (*Sigma G1393*) in tissue-culture grade H_2O) and coated with mitotically inactive primary mouse embryo fibroblasts (3-4x10^4cells/cm^2) with standard mESC medium containing 15% FCS and 1000 U/ml leukemia inhibitory factor (*LIF, Chemicon ESG1107*) at 37°C and 5% or 7.5% CO_2. mESCs were split every second day with a dilution suitable to the proliferation velocity (between 1:5 and 1:9). ES+LIF medium was refreshed daily. For splitting, media was aspirated and cells were washed once with PBS and trypsinized (Tryspin-EDTA (0.05%) (Gibco 25300054)) for 5-10 minutes at 37°C. Trypsin was neutralized by 3ml ES+LIF and cells centrifuged for 5 minutes at 1000 rpm, after which the pellet was resuspended in ES+LIF. For freezing of mESCs, cell pellets were resuspended in ES medium with 20% FCS, and mixed in a 1:1 ratio with ES freezing medium. Cells were frozen down o/n in the –80 °C and transferred to liquid nitrogen the next day.

Derivation of Prdm14-Venus mouse embryonic stem cell line from mouse blastocysts

Mouse embryonic stem cells (mESCs) were derived from Prdm14-Venus reporter positive blastocysts (B6 x B6D2F1). Blastocysts were treated with Acidic Tyrode's solution for 2mins to remove the zona pellucida and washed in three serial drops of M2 media and three drops of mESC medium. Single blastocysts were seeded into each well of a 24-well dish coated with MEFs containing mESC medium supplemented with PD0325901. After blastocysts attached onto the feeder layer (~3 days), media was changed every day. Outgrowths were dissociated into single cells with 0.25% Trypsin, and reseeded in the same well. Cells were monitored for visible dome-shaped colonies. The mESC colonies were dissociated with TrypLE and passaged on feeders – P0 (Passage zero). Subsequently, Prdm14-Venus mESCs were maintained on feeders with mESCs medium (see above for mESC culture).

Generation of transgenic mouse embryonic cell lines

The mouse embryonic cell lines used for the experiments were all derived from F1G4 cells except for the Prdm14-Venus reporter line (see above). To generate the fluorescent reporter constructs, we genome-engineered the mouse BACs *RP24-530D23* (*T*), *RP23-249O15* (*Sox2*) and *RP23-421P23* (*Tbx6*), containing ~180 - 230kb of the C57BL/6 mouse genome surrounding the respective loci, via Red/ET recombineering (*49*). In short, the starting codon (ATG) of the genes was replaced with a reporter gene containing H2B-mCherry (for *T*) or H2B-Venus (for *Sox2*/*Tbx6*), followed by the rabbit *b-globin* polyadenlylation signal and an *FRT*-flanked selection cassette (hygromycin or neomycin), driven by the *Pgk* promotor. For random integration, 5µg of the modified BACs was linearized with PI-SceI (NEB) and electroporated into $3x10^6$ mESCs. Selection was performed using mESC medium containing 250µg/ml neomycin and 150µg/ml hygromycin for the T::H2B-mCherry/Sox2:H2B::Venus reporter line. For the Tbx6::H2B-Venus reporter line 350µg/ml hygromycin was used for selection. After selection, single clones were picked, expanded and checked for BAC integration by PCR. Two independent *Tbx6::H2B-Venus* mESC lines were generated. In both lines, the endogenous *Tbx6* locus was targeted by CRISPR/Cas9 (**Data S4**) to create null mutants. To this end, a double nicking approach with four guide RNAs (*50*) was used. The specific target sequences contained 5'-$GN_{19}NGG$-3', with *N* being any arbitrary nucleotide, *G* being guanine and *NGG* the Protospacer adjacent motif and GN_{19} the guide RNA. If there was no guanine at the 5'-end of the template sequence, an extra guanine was added 5' to the other 20 bases. Single stranded

oligonucleotides (**Data S4**) with an added 5'-*CACC-3'* at 5'-end and 5'-*AAAC*-3' at 5'-end of the complementary strand were annealed in 10X T4 ligation buffer *(Promega M1801)* by continuous cooling from 95°C to 25°C. Annealed oligos were cloned into the *BbsI* site of px335A_hCas9_D10A_G2P (gift from dr. Boris Greber) (**Data S4**) containing expression cassettes for hCas9 nickase, guide RNA and puromycin resistance.

One day prior to transfection, $3x10^5$ mESCs were seeded on fibroblast coated 6-well plates *(Costar 3516)* with 3ml ES+LIF per well. The next day, the medium was refreshed. Separate mixes of 110µl Opti-MEM reduced serum medium *(Gibco 31985062)* plus 25µl Lipofectamine2000 *(Thermo Fisher 2125239)* and 125µl Opti-MEM plus 8µg per vector were prepared. 125µl of each mix was then combined and incubated for 15 minutes at room temperature before being transferred to 1.25ml ES+LIF without Penicillin/streptomycin. After 5h of incubation with the transfection mix, mESCs were split and plated on puromycin resistant feeders. 24h after transfection, transient selection was started with ES+LIF containing 2µg/ml puromycin *(Gibco 10130127)* at day 1 and day 2 and ES+LIF containing 1µg/ml puromycin at day 3. After approximately one week, single clones were picked and expanded on 96-well plates *(Costar 3596)*. Clones were genotyped by PCR using primers spanning Exons 1-4. Deletion breakpoints were analyzed by Sanger sequencing of purified PCR fragments.

Generation of gastruloids

All medium compositions are listed in **Data S3**. Gastruloids were generated as described previously (*51*), with some minor modifications. First, mESCs were feeder freed. To this end, mESCs were trypsinized on the feeder plate as described above and washed with ES+LIF. Cells were pelleted by centrifugation at 1000rpm for 5 minutes at room temperature and resuspended in 2ml ES+LIF. On three gelatinized (0.1% gelatin) wells of a 6-well plate, cells were sequentially plated for 25 minutes, 20 minutes and 15 minutes during which cells were kept in the incubator at 37°C and 5% or 7.5% CO_2. With each transfer, cells were triturated to maintain a single cell suspension. Feeder-freed mESCs were then washed once in 5ml pre-heated (37°C) PBS containing $MgCl_2$ and $CaCl_2$ (*Sigma D8662*) and once in pre-incubated (37°C, 5% or 7.5% CO_2) 5ml NDiff227 (*Takara Y40002*). mESCs were then pelleted by centrifugation for 5 minutes at 1000rpm and resuspended in 500µl of NDiff227. 10µl of the cell suspension was mixed with 10µl of Trypanblue (*Bio-Rad 1450021*) for automated cell counting with *Luna Automated Cell Counter*. 200-250 live cells were then plated in a volume of 30 to 40µl

NDiff227 into each well of a 96-well round bottom, low attachment plate (*Cellstar 96 well suspension culture plate (655185)* or *Costar 7007 ultra-low attachment 96 well plate (7007)*). Cells were then allowed to aggregate for 48 h. After these 48h cells were pulsed with 3µM CHIR99021 (CHIR, *Merck Millipore*) in 150µl NDiff227. Between 72 and 120h, medium was refreshed every 24h by removing 150µl of the old media and adding the same volume of new, pre-incubated (37°C and 5% or 7.5% CO_2) NDiff227. For gastruloids treated with CHIR and CHIR+LDN, 5µM CHIR with or without 600nM LDN193189 (LDN, *Reprocell*) was added from 96h to 120h. For controls, an equal volume of diluent (DMSO) was added.

Generation of trunk-like-structures

All medium compositions are listed in **Data S3**. The gastruloid protocol described above was followed until 96h. Gastruloids were then embedded in 5% Growth-Factor-Reduced Matrigel (MG) (*Corning 356231*). To this end, fresh NDiff227 medium was pre-incubated for at least 20 minutes at 37°C and 5% or 7.5% CO_2. Pre-incubated medium was then put on ice for 5 minutes, after which MG was added to achieve a final concentration of 5% in the culture wells. Medium was then put at room temperature for 5 minutes, during which 150µl of old medium was removed from the aggregates. New medium with MG (150µl) was then added, and the cultures were returned to the incubator and further cultured at 37°C and 5% or 7.5% CO_2. TLS cultures were allowed to settle for at least 30 minutes before proceeding to further experimentation (e.g. live imaging). For TLS treated with CHIR or CHIR+LDN, 5µM CHIR alone or together with 600nM LDN was added prior to MG and left from 96h to 120h. For controls, an equal volume of diluent (DMSO) was added.

Spiking in of different ECM components

To test the effect of different ECM components, the gastruloids protocol was followed until 96h after aggregation. Then, 150ul of media was removed using a multi-channel pipette. Prior to the addition of Collagen IV and Laminin the pre-incubated (37°C and 7.5% CO_2) NDiff227 medium was pre-cooled on ice for 5 minutes, after which the components were added. When Collagen IV and/or Laminin were added in combination with low-melting-point (LMP) agarose half of the NDiff227 medium volume was pre-cooled, whereas the other was kept at 37°C to allow for the addition of LMP agarose. Immediately after adding the components and thorough mixing, 150ul of medium was added to each well of the 96-well plate and plates were put back

into the incubator. Laminin (*Sigma 114965-81-9* or *Corning 354232*) was used at a final concentration of 0.06mg/ml or 0.12mg/ml. Collagen IV (*Santa Cruz sc-29010* or *Cultrex 3458-096-02*) was used at a final concentration of 0.12 mg/ml. In the experiment where Laminin (*Corning 354239*) and Collagen IV were combined, Laminin was used at at concentration of 0.03mg/ml and Collagen IV at a concentration of 0.048 mg/ml. For the experiments with LMP agarose (*Promega V2111*), a 1.5% stock was prepared in tissue-culture grade 1x PBS and subsequently a 5% solution of this stock in NDiff227 was used. For Laminin, Collagen IV, Collagen IV + LMP agarose, LMP agarose, n = 2 independent experiments. For Collagen IV + Laminin, Collagen IV + Laminin + LMP agarose, n = 1 independent experiment. In all experiments trunk-like-structures (5% matrigel) and gastruloids (no ECM surrogate) were taken along as positive and negative controls respectively.

Whole-mount Immunofluorescence

Collected embryos were washed twice in PBS and then fixed in 4% PFA for 30 minutes under rotation at 4°C, washed 3x with PBS, and stored in PBS until immunofluorescent staining was performed. Prdm14::Venus reporter embryos were fixed with 4% PFA overnight at 4□C. Gastruloids or trunk-like-structures were picked using a p200 pipette with the tip cut-off at the 50µl mark. Gastruloids were washed twice with PBS and then fixed in 4% PFA for 1 h in glass vials (*Wheaton 224882*) at 4°C on a roller. For trunk-like-structures, individual structures were picked using a p200 pipette with the tip cut-off at the 50µl mark, and transferred to either 96-well plates (*Costar 3596*) or Ibidi 8-well glass-bottom plates (*Ibidi 80827*). Trunk-like-structures were washed twice with PBS + $MgCl_2$ and $CaCl_2$ + 0.5% BSA (*Sigma A8412*), once with PBS, and then fixed in 4% PFA for 1h at 4°C on a rocking platform. Subsequently, gastruloids or trunk-like-structures were washed twice in PBS for 5 min, permeabilized by 3 x 20 minutes incubation in 0.5% Triton-X/PBS (PBST) and blocked in 5% fetal calf serum/PBST (blocking solution) overnight at 4°C. For antibody staining, embryos/gastruloids/trunk-like-structures were transferred to Ibidi 8-well glass bottom plates. Primary antibody incubation was performed in blocking solution for 48-72h at 4°C, after which embryos/gastruloids/trunk-like-structures were washed three times with blocking solution and three times with PBST. After the last washing step, embryos/gastruloids/trunk-like-structures were incubated in blocking solution o/n at 4°C. The next day, secondary antibodies diluted in blocking solution were added, and structures were incubated for 24h at 4°C. Afterwards, embryos/gastruloids/trunk-like-

structures were washed three times with blocking solution and three times with PBST. The last PBST washing step after secondary antibody incubation included DAPI (0.02%, *Roche Diagnostics 10236276001*). DAPI was incubated o/n and washed off once with PBST. All primary and secondary antibodies are listed in **Data S3**.

For the fibronectin stainings, samples were picked from individual wells, and an equal volume of 4% PFA was added. Samples were then incubated for 15 minutes at 4°C, followed by 45 minutes at RT. Subsequently, samples were washed twice with PBS, and fixed for another 60 minutes at 4°C. Samples were then permeabilized with PBST (2x 20 minutes RT), and blocked o/n in blocking solution. The next day, antibodies were added for 24-48h (**Data S3**). Subsequently, samples were washed 2 times with blocking solution and 1 time with PBST, incubated in blocking solution for 1h at RT, and secondary antibodies and phalloidin were added. After o/n incubation at 4°C samples were washed twice in blocking and twice in PBS. The last included DAPI (0.02%, *Roche Diagnostics 10236276001*). DAPI was incubated o/n and washed off with 1xPBS, after which samples were processed for tissue clearing.

Tissue clearing

Prior to imaging, embryos, gastruloids and trunk-like-structures were cleared with RIMS (Refractive Index Matching Solution). To this end, samples were washed twice with PBS for 10 minutes, post-fixed in 4% PFA for 20 minutes and washed three times with 0.1M phosphate buffer (PB, 0.025M NaH_2PO_4 and 0.075M Na_2HPO_4, pH 7.4). Clearing was performed by incubation in RIMS (133% w/v Histodenz (Sigma-Aldrich D2158) in 0.02M PB) on a rocking platform at 4°C for at least one to several days.

Imaging

Embryos, gastruloids and trunk-like-structures stained with antibodies or carrying fluorescent reporters were imaged with the Zeiss AxioZoom v16 (wide-field), Zeiss Celldiscoverer7 (wide-field), Zeiss LSM710 (laser-scanning microscope with Airyscan), Zeiss LSM880 (laser-scanning microscope with Airyscan) or Zeiss Lightsheet Z1 with appropriate filters for mCherry, Venus, DAPI, Alexa Fluor 488, Alexa Fluor 568, Alexa Fluor 647, and combinations thereof. Whole-mount in situ hybridization images were acquired with Zeiss AxioZoom v16.

Post-acquisition image processing was performed using Zen, Arrivis and/or Fiji. For live imaging of TLS in 96-well plates we used the Zeiss Celldiscoverer 7 with incubator chamber temperature set at 37°C and CO2 content at 5%. Acquisition intervals ranged from 15-45 minutes.

Scoring criteria WIFC

For WIFC structures were picked that displayed segmentation. For each protein, total sample size and reproducibility of the staining patterns are indicated in the figures. In detail:

Fig. 1B: TLS: T^{mCH+} segments (for definition see "Morphometric analysis of TLS") flanking a $Sox2^{VE+}$ tubular structure (unilateral or bilateral). TLS^{C} and TLS^{CL}: somitic segments arranged like a bunch-of-grapes and compromised neural domain. Multiple examples of each condition are provided in **Fig S1**.

Fig. 1G: Accumulation of NCAD and F-actin at apical side of the somites and neural tube.

Fig. 2G: Accumulation of FN1 at the structure-matrix interface and accumulation of F-Actin at the apical side of somites and neural tube.

Fig. 3C: T/SOX2 co-expressing cells at the posterior end (96h and 120h TLS).

Fig. 4E: Cluster of T^{+}/PRDM14-$Venus^{+}$ cells (76h TLS; more examples in **Fig. S14B**). Cluster of $DPPA3^{+}$/Sox2::H2B-$Venus^{high}$ cells (108h TLS, second example in **Fig. S14C**). $DPPA3^{+}$ cells in proximity of T^{+} tubular structure or Sox2::H2B-$Venus^{high}$ cells in proximity of the $FOXA2^{+}$ domain (TLS 120h, **Fig. S14D**).

Fig. 5G: $SOX2^{+}$/$Tbx6^{VE+}$ tubular structures flanking the main $SOX2^{+}$-only neural tube.

Fig. S2: Accumulation of NCAD or F-actin at apical side of the somites and neural domain (TLS^{C} and TLS^{CL}) or the absence thereof (gastruloids).

Fig S3A: $FOXA2^{+}$ gut-like tubular structure stretching (part of) anterior-posterior axis.

Fig. S3B: $T^{CH\text{-}high}$/$FOXA2^{+}$ cells at the posterior base of the gut-like tubular structure (upper panel). $T^{CH\text{-}high}$/$SOX17^{-}$ cells at the posterior base of the gut-like tubular structure (lower panel).

Fig. S3C,D: $FOXA2^{+}$ gut-like tubular structure stretching (part of) anterior-posterior axis with $T^{CH\text{-}high}$/$FOXA2^{+}$ cells at its posterior base.

Fig. S7D: Accumulation of FN1 around the structure (TLS^{CL}) or lack thereof (gastruloids).

Fig. S10A: T/SOX2 co-expressing cells at the posterior end.

Fig. S10B: $CDX2^{+}/T^{CH+}/SOX2^{VE+}$ triple-positive cells at the posterior end (CDX2 fluorescence intensity over anterior-posterior axis quantified in **Fig. S10B**).

Fig. S14D: $Prdm14^{VE+}/DPPA3^{+}$ cluster (right panel).

Fig. S20D,E: $Tbx6^{VE+}$ somitic segments arranged like a bunch-of-grapes and compromised neural domain (WT). Expansion of neural domain with intermingled $Tbx6^{VE+/}SOX2^{+}$ cells (KO).

Fig. S20F: Expansion of neural domain with intermingled $Tbx6^{VE+/}SOX2^{+}$ cells.

Morphometric analysis of TLS

Scoring of segmentation, T-poles, axes of elongation and somite bilaterality in TLS and gastruloids

For quantification of the number of structures with a T^{+}-pole, we employed the *T::H2B-mCherry* reporter line. If a structure displayed multiple axes of elongation, it was sufficient for one end to have a $T::H2B\text{-}mCherry^{+}$ pole to be scored as having a "T-pole". We defined gastruloids/trunk-like-structures as "segmented" when at least four neighboring segments had developed along the antero-posterior axis at 120h. Segments were defined as T^{mCH+} substructures that displayed i) indentations and ii) opposite curvatures at segment borders. Finally, if a structure displayed two or more axes of elongation, it was scored as "multiple axes", if showing just one as "one axes", if none "no elongation". For all conditions, at least three independent experiments, each one including the specific treatment condition, were analyzed. Structures that grew out of focus in a way that they could not be rated for one of the categories were excluded for the respective category.

To analyze if somites formed uni- or bilaterally, we recorded $SOX2^{VE}$, T^{mCH} and brightfield with the Axiozoom V16 (Zeiss). Structures were scored as "bilateral" if at least 3 neighboring segments (see above for criteria) were detected bilaterally from the neural tube.

We note that because of H2B-reporter stability mCherry (as well as Venus) reporter signal is detected all along the anterior-posterior axis. However, there is a clear enrichment of T^{mCH} at the posterior end, which we used to score for the "T/pole" (see e.g. **Fig S1B,C**).

Whole structure 2D morphometry

Gastruloids, TLS, TLS^{C} and TLS^{CL} structure contours were identified from Axiozoom transmitted light images by variance of a based thresholding algorithm on a Zeiss ZEN Blue workstation (Version 3.1). In detail, transmitted light image channels were variance based segmented with the ZEN image analysis module and the images were filtered to exclude background objects (single cells, debris, imaging artifacts). Each individually generated mask was quality checked before exporting metric data sets. Area, perimeter, aspect ratio (perimeter/area), compactness, roundness and circularity were calculated using the identified masks. Significance was tested by performing one-way ANOVA followed by Tukey's multiple comparison test and raw data with associated statistics are provided in **Data S5**.

Time-resolved whole-structure 2D morphometry

TLS, TLS^{C} and TLS^{CL} were identified from Zeiss CD7 CellDiscoverer wide-field brightfield images as described above for whole structure morphometry. Area, circularity, roundness and compactness were measured over time between 97h to 119h after aggregation at intervals of 15-45min.

2D Neural Tube morphometry

TLS, TLS^{C} and TLS^{CL} fluorescent images acquired with Zeiss Axiozoom V16 were segmented in order to identify the neural tube and to compute their length, width, perimeter, area, area neural/area whole structure, aspect ratio, solidity and circularity.

The extraction of the neural tube morphological measurements was performed according to the following steps:

1. Masking away of the region expressing T^{mCH}

2. Segmenting the region expressing Sox2^{VE} but not T^{mCH}: segmenting the neural tube
3. Manually correcting the segmentations when necessary
4. Skeletonizing the neural tube region
5. Extracting the measurements

Notations:

Let an image I be a function $I: \Omega \subset \mathbb{R}^2 \rightarrow \mathcal{I} \subset \mathbb{N}$, where Ω is the spatial domain of definition of the image and $\mathcal{I}$ being the dynamic range of the image, in our case $\mathcal{I} = [0, 2^{16} - 1]$.

Let $Im(I) = (I(x,y)|(x,y) \in \Omega)$ be the image of Ω by the function I. It is therefore the sequence of reached intensities by the image I.

Let I_{Sox2} be the image of intensity for the Sox2^{VE} channel.

Let I_T be the image of intensity for the T^{mCH} channel.

Note that since I_{Sox2} and I_T are two channels of the same sample, at the same time:

$$\Omega_{Sox2} = \Omega_T = \Omega$$

Masking away the region expressing T^{mCH}:

Cells expressing Sox2^{VE} are either neural tube cells, neural tube precursors or neuromesodermal progenitors. Cells expressing T^{mCH} are either somite cells, somite precursors or neuromesodermal progenitors. To extract the neural tube but not the precursor cells, we have to detect cells expressing T^{mCH} and to remove them from the set of cells expressing Sox2^{VE}.

To extract the cells expressing T^{mCH}, we first applied a Gaussian filter with a sigma of 5 pixels to I_T generating the smoothed image I_T^G. We then separated the background from the foreground in I_T^G by thresholding the image with a threshold th_T^b, obtained by applying the Otsu method (*52*). Using this threshold, we extracted the foreground intensity values and built $\mathcal{F}_T = \left(i \in Im(I_T^G) \backslash i > th_T^b\right)$, the sequence of foreground intensity values. Then, in order to separate intensity values resulting from actual T^{mCH} expression and intensity values resulting from autofluorescence, we made the hypothesis that the intensities of the two groups of cells were observing different distribution. We therefore used the Otsu method a second time to best split

$\mathcal{F}_T$ in two distributions resulting in a second, and final, threshold value th_T. Let call $Otsu^2$ this operation of two consecutive Ostu thresholding of an image I resulting in a threshold value th, in our case:

$$th_T = Otsu^2(I_T^S)$$

We can now build the mask M_T image of T^{mCH} positive cells:

$$M_T: (x,y) \in \Omega \rightarrow \begin{cases} 0 if I_T(x,y) < th_s \\ 1 otherwise \end{cases} \quad \text{Eq 1}$$

We then altered the $Sox2^{VE}$ image using the mask of T^{mCH} positive regions creating a new image $I_{Sox2}^{M_T}$:

$$I_{Sox2}^{M_T}: (x,y) \in \Omega \rightarrow \begin{cases} I_{Sox2}(x,y) if M_T(x,y) = 0 \\ \tilde{I}_{Sox2} otherwise \end{cases},$$

where $\tilde{I}_{Sox2}$ is the average intensity of I_{Sox2}.

Segmenting the region expressing $Sox2^{VE}$, reconstructing the neural tube:

We then smoothed $I_{Sox2}^{M_T}$ with a Gaussian kernel of sigma 5 giving the smoothed image I_{Sox2}^{G} and built the mask image M_{Sox2} using the threshold $th_{Sox2} = Otsu^2(I_{Sox2}^{G})$ as described in (Eq 1). We then applied a morphological operations (*53*) to smooth and remove morphological irregularities. We first applied a closing with a radius of 10 pixels followed by an opening with a radius of 5 pixels generating a new, filtered, mask containing multiple connected components. We kept the connected component that had the largest volume as the neural tube segmentation and discarded the other connected components generating our final mask M^{NT} and its corresponding set of pixels: $P^{NT} = \{p \in \Omega | M^{NT}(p) = 1\}$.

Manual inspection and curation:

Ultimately, we manually inspected the segmented neural tubes and corrected them when necessary using Fiji (*54*) creating our final mask M_{cur}^{NT}. We used this mask to perform the measurements. In order to be able to measure the neural tube width as a function of the anterior-posterior position, we manually marked the anterior and posterior ends of the neural tube based on the position of the neuromesodermal progenitor domain.

Skeletonizing the neural tube region:

To compute the length and the width of the neural tube we skeletonized (*55*) the segmented region generating a mask image S where $S(x,y) = 1$ if and only if the pixel in position (x,y) belongs to the skeleton of M_{cur}^{NT}. Because of some irregularities of the segmentation, the reconstructed skeleton could have several branches. In order to remove the auxiliary branches, product of the irregularities, we first built a graph $G^S = (V,E)$ of the skeleton where each vertex $p \in V$ is a pixel of the skeleton and two vertices are connected ($(p_i, p_j) \in E$) if and only if there were 4-connected in the image:

$$G^S = (V,E) V = \{(x,y) | S(x,y) = 1\} E = \{(p_i, p_j) \in V^2 | \|p_i - p_j\| \leq 1\},$$

where $\|\cdot\|$ is the L2 norm.

Ultimately, we identified the neural tube skeleton as the set of vertices from G^S that belong to the longest path $LP \subset V$ of G^S building the mask M_{Skel}^{NT} of the skeleton:

$$M_{Skel}^{NT}(x,y) = \begin{cases} 1 if (x,y) \in LP(G^S) \\ 0 otherwise \end{cases}$$

and we built the subgraph $Skel = (V_{Skel}, E_{Skel})$of G^S containing only the longest path:

$$V_{Skel} = LP E_{Skel} = E \cap LP^2.$$

Note that the longest path was computed by enumerating all the extremities of G^S (vertices with exactly one neighbour) and computing all the pairwise distances between the extremities. The couple of extremities raising the longest distance is the longest path in G^S.

Extracting the measurements:

Once the segmentation of the neural tube and the construction of its skeleton was done we extracted the set of necessary measurements. In order to extract measurements in physical units (in our case µm) we had to use the pixel size. For all our images the pixel size was isotropic, let denote it as vs.

Length:

We computed the length (L) of the neural tube as the length of the longest path of G^S normalized by the voxel size:

$$L = vs \cdot \sum_{(p_i,p_j)\in E^{LP}} \|p_i - p_j\|$$

Width:

Since the width changes at different position within the neural tube, we computed the width (W) as the median width along the skeleton of the neural tube.

We computed the width (w_i) of the neural tube, for a given position $p_i \in \Omega$, as the distance between that position and the closest pixel that does not belong to the segmented neural tube:

$$w_i = vs \cdot \min_{p_j \notin P^{NT}} \|p_i - p_j\|.$$

Then W is computed as previously mentioned:

$$W = \dot{w}$$

where:

$$w = \{w_i | p_i \in V_{Skel}\}$$

and $\dot{\cdot}$ is the median.

Perimeter:

We computed the perimeter (P) as defined in (*56*) from the mask image M^{NT} and normalized it by the voxel size:

$$P = vs \cdot perimeter(M^{NT})$$

where $perimeter$ is the perimeter function described in (*56*) and implemented in (*57*).

Area:

We computed the area (A) as number of pixels belonging to the segmentation mask times the voxel size:

$$A = vs^2 \cdot |\{v \in \Omega | M^{NT}(V) = 1\}|$$

Aspect ratio:

We computed the aspect ratio (AR) as the ratio of the structure as the length over the width:

$$AR = \frac{L}{W}$$

Solidity:

We computed the solidity (S) as described in (*58*):

$$S = \frac{A}{A_{CH}(M^{NT})}$$

where A_{CH} is the area of the convex hull of a given binary mask, we used the implementation provided in (*57*).

Circularity:

We computed the circularity (C) as described in (*58*):

$$C = 4\pi \frac{A}{P^2}$$

we used the implementation provided in (*57*).

One-way ANOVA followed by Tukey's multiple comparison test was used to calculate p-values and raw data with associated statistics are provided in **Data S5**.

Whole structure, Neural Tube and Gut Tube 3D morphometry

TLS, TLS^{C} and TLS^{CL} fluorescent images acquired with confocal microscopy were segmented in order to identify whole structure, neural tube and gut to compute their length, width, area, volume, aspect ratio, solidity and sphericity.

The following detailed analysis for the gut measurements have been applied also to the whole structure and to the neural tube.

The extraction of the gut morphological measurements was performed according to the following steps:

1. Segmenting the region expressing either SOX17 or FOXA2: segmenting the gut
2. Skeletonizing the gut region
3. Extracting the measurements

<u>Segmenting the gut tube:</u>

In order to segment the gut we combined a set operations on the 3D images to build the mask best representing the gut. First we resampled the image from its original anisotropic sampling into an isotropic sampling. The new voxel size was chosen so the image had 500 voxels along the x dimension. Once resampled, we applied a Gaussian filter with a sigma value of 3 voxels. We then manually chose a threshold value such that the whole gut was covered creating a binary mask. Next, we applied a morphological closing with a radius of 10 voxels onto the previously created binary mask. If necessary, we followed the closing operation with a morphological opening with a radius of 5 voxels. We discarded the connected components raised by this thresholding that did not belong to the gut, filled the potential holes and if necessary, applied a

last morphological closing with a radius of 20 to ensure having only one single connected component and having a surface homeomorph to a sphere.

Skeletonizing the gut tube:

As we did for the neural tube, we created the morphological skeleton of the previously mentioned mask in order to measure gut length and computed the distance chamfer transformation to compute the gut width.

Extracting the measurements:

To extract the measurements we proceeded similarly as for the neural tube with a few adjustments to fit the 3D images:

Length:

The length was computed as described for the neural tube.

Width:

The width was computed as described for the neural tube using the 3D chamfer transformation.

Area:

The area was computed as defined in (*56*) using the implementation from (*58*), from the mask image, and normalized it by the voxel size

Volume:

We computed the volume (V) as number of voxels belonging to the segmentation mask times the cube of the voxel size:

$$V = vs^3 \cdot |\{v \in \Omega | M^{Gut}(v) = 1\}|$$

Aspect Ratio:

The aspect ratio was computed as described previously.

Solidity:

We computed the sphericity (S) as described in (*58*):

$$S(M) = \frac{V(M)}{V_{CH}(M)}$$

Where M is a given binary mask, $V(M)$ is the volume of the mask M and $V_{CH}(M)$is the volume of the convex hull of the mask M. We used the implementation provided in (*57*).

Sphericity:

We computed the sphericity (Sph) as described in (*58*):

$$Sph(M) = \pi^{\frac{1}{3}} \frac{(6.V(M))^{2/3}}{A(M)}$$

Where M is a given binary mask, $V(M)$ is the volume of the mask M, and $A(M)$ is the area of the mask M. We used the implementation provided in (*57*).

All gut tube measurements can be found in **Data S5**.

Somite morphometry

Somite sizes were measured in Fiji 1.8.0 (*54*) using the DAPI channel from confocal high resolution Z-stacks to identify epithelialized somitic structures. For E8.5 mouse embryos, TLS, TLS^{C} and TLS^{CL} somites were manually segmented and masks for each somite were stored and used to calculate area, perimeter, length, width, circularity, roundness, solidity and number of somites/TLS. Individual masks were drawn inspecting all Z-stacks for each identified somites and selecting the Z position containing the biggest surface. For embryo somites, the 8 most posterior ones were measured, corresponding to the median of the number of somites measured flanking one side of the neural tube in TLS. In detail:

Area:

Area was calculated as the area of the mask in square pixels and then recalibrated in square micrometers.

Perimeter:

Perimeter was calculated as the length of the outside boundary of the selected mask.

Length:

Length was calculated drawing the maximum distance between two points of each mask along the anterior-posterior axis.

Width:

Width was calculated drawing the maximum distance between two points of each mask perpendicular to the anterior-posterior axis.

Circularity:

Circularity for each mask was calculated as

$$C = 4\pi \frac{A}{P^2}$$

where A is area and P is perimeter of the mask.

Roundness:

Roundness for each mask was calculated as

$$R = 4 \frac{A}{\pi(max\_axis)^2}$$

where max_axis is the longest distance between two points of the mask.

Solidity:

Solidity for each mask was calculated as

$$S = \frac{A}{A_{CH}}$$

where A_{CH} is mask convex area.

One-way ANOVA followed by Tukey's multiple comparison test was used to calculate p-values. Raw data with associated statistics are provided in **Data S5**.

CDX2 expression along the anterior-posterior axis in TLS

CDX2 expression along the anterior-posterior axis was performed in Fiji 1.8.0 (*54*). In detail, stacks of composite images of DAPI and CDX2 stainings imaged by confocal microsopy were loaded into Fiji, brightness/contrast were automatically adjusted and background was subtracted for all stacks. A 3D Projection was then generated using the default settings, a line (width 50 pixels) was manually drawn from the posterior to the anterior end of the structure along the midline, and fluorescent intensity measured using the "Multichannel Plot Profile" function of the BAR plug-in.

Segmentation clock measurements

To measure the interval between the formation of subsequent segments (or bursts of segmentation in case of TLS^{C} and TLS^{CL}) we analyzed live-imaging recordings of T^{mCH} and brightfield channels in trunk-like-structures developing from 97h onwards. Interval time of the recordings was between 15 and 45 minutes. Raw data associated with **Fig. S5B,C** can be found in **Data S5**.

Flow cytometry and FACS

Individual trunk-like-structures were washed with cold PBS+0.5%BSA 2 times, after which 50µl of trypsin was added. Trunk-like-structures were then incubated for 10 minutes at room temperature, after which samples were dissociated by trituration for 50 times using a p200 pipette to achieve a homogenous suspension. The reaction was then stopped with the addition of 100µl of 5% BSA in PBS. Before cell counting and/or sorting on a FACS Aria II (*Becton Dickinson*) or a BD FACSCelesta Flow Cytometer (*Becton Dickinson*, counting only) the cell suspension was filtered through a 35µm mesh. Equal numbers of samples for each condition were harvested in three independent experiments. For FACS, samples were dissociated as described above, and the $Tbx6^{VE+}$ and $Tbx6^{VE-}$ fraction was sorted in 1.5ml low-binding tubes with 350µl RLT (*Qiagen*) + 1% v/v B-mercaptoethanol (*Sigma*). Flow cytometry data were later analyzed using FlowJoV10.

RNA isolation, reverse transcription and quantitative PCR

RNA isolation was performed using the RNeasy Micro Kit (*Sigma 74004*) according to the manufacturer's instructions with the following modifications: genomic DNA was digested on column with the addition of an extra 1µl of RNase-free DNase I (*Roche*) to ensure complete digestion, and after the 80% ethanol column wash and column centrifugation the remaining ethanol was removed with a p10 pipette tip and columns were left to air-dry for 5 minutes. Quantitative reverse transcriptase PCR (qRT-PCR) was then performed using a two-step protocol. First, RNA was reverse transcribed using the Quantitect Reverse Transcription Kit (*Qiagen*), according to the manufacturer's instructions. qPCR was carried out on a StepOnePlus Real-Time PCR System (*Life Technologies*) using GoTaq qPCR Master Mix (*Promega*) with validated gene-specific primers (**Data S4**). Fold change was calculated from ΔCt using *Tbp* as housekeeping gene (**Data S6**).

Whole-mount in situ hybridization

Probe synthesis

For synthesis of DIG labeled RNA probes, plasmids containing the cDNA of interest and promoters for sense and anti-sense strand synthesis were used from the MAMEP database (http://mamep.molgen.mpg.de). In order to obtain a sufficient amount of material, some plasmids were first retransformed into *E. coli* (DH5α) and afterwards isolated with a mini-prep kit (*Qiagen*) according to the manufacturer's instructions as described below. Probes were synthesized by *in vitro* transcription using a PCR product of the desired cDNA or the linearized plasmid.

Retransformation in DH5α and mini-prep

DH5α cells were thawed on ice and 1µl of the plasmid was added to 100µl of competent cells and incubated on ice for 30 minutes. Subsequently, cells were heat shocked at 42°C for 45 seconds in a water bath and then immediately cooled down on ice. 500µl of LB medium was added and the mix was incubated for 1h at 37°C (heating block) under shaking. Then, 25µl and 250µl of the mixture were plated on separate Agar-plates containing Ampicillin (Amp) for selection of transformed bacteria (all vectors carried a gene for Ampicillin resistance). The plates were incubated overnight at 37°C. The next day, single colonies were picked in 5ml LB+Amp and incubated overnight at 37°C under shaking. Plasmids were isolated with the

QIAprep Spin Miniprep Kit (*Qiagen*) following the manufacturer's instructions. DNA concentration was measured with a Nanophotometer (*Implen*).

Restriction Digest

After plasmid digestion (250ng of the plasmid, 2µl 10x buffer, 0.2µl 100x BSA and 2µl of each restriction enzyme in a 20µl total volume), expected band sized for the vectors and inserts was confirmed on a 1% Agarose/TAE-gel, stained with SybrSafe (1:20.000).

Polymerase-Chain-Reaction (PCR)

For PCR, 5µl plasmid (1 ng/µl) was used in a 50µl total reaction volume containing 5µl 10x PCR buffer (*Invitrogen*), 1.5µl $MgCl_2$ (*Invitrogen*), 10µl dNTP Mix (10 mM for each nucleotide), 0.25µl forward Primer (100 µM), 0.25µl reverse Primer (100 µM) and 0.2µl Taq DNA Polymerase (5u/µl) (*Invitrogen*). For the plasmid of Uncx a different PCR strategy was used, because of its CG richness. Here, the 50µl reaction contained 5µl Plasmid (1ng/µl), 1µl dNTPs, 0.25µl U5, 0.25µl L2, 0.2µl Taq DNA Polymerase (5u/µl) (*Qiagen*) and 10µl 5x Q-solution (*Qiagen*). Primers used for the respective vectors are listed in **Data S4**. All PCR products were checked on a 1% Agarose/TAE-gel.

Linearization of plasmids for in vitro transcription

Plasmids were linearized with the appropriate restriction enzyme (1.5µg DNA, 2.5µl 10x buffer, filled up with DEPC-H_2O to 25µl). The reaction was incubated for 1h at 37°C. Subsequently, 8µl ammonium acetate (10M) and 80µl ice-cold 100% ethanol were added, followed by centrifugation of the sample for 30 minutes at 4°C and 13200 rpm. The supernatant was then removed and 150µl of 70% ice-cold ethanol was added. The sample was then centrifuged again for 10 minutes, the supernatant was removed, and the pellet was air dried and dissolved in 9.5µl DEPC-H_2O.

In vitro transcription

For *in vitro* transcription, 9µl of PCR product or 1.5µg of linearized plasmid was incubated with 3µl 10x Transcription buffer, 3µl ACG nucleotides (4mM each nucleotide), 0.75µl digUTP-UTP Mix (4mM), 1.5µl DTT, 1µl RNase inhibitor and 60U of the respective RNA polymerase, in a total reaction volume (filled up with DEPC-H_2O) of 30µl. The reaction was incubated for 2h at 37°C (T7, T3) or 40°C (Sp6). Subsequently, 3µl of RNase-free DNaseI (10u/µl) was added and the reaction was incubated for 15 minutes at 37°C. The RNA probe

was then purified with ProbeQuant G50 Sephadex columns (*Pharmacia*). Adding 20µl DEPC-H_2O to the probe, spinning it through the resin of the column and adding again 30µl DEPC-H_2O increased its volume to 100µl. Aliquots of 30µl were immediately put on dry ice and stored at -80°C. The RNA probes were checked on a 1% Agarose/TAE-gel.

Fixation and methanol series

Collected embryos, gastruloids and trunk-like-structures were washed twice in DEPC-PBS and then fixed in 4% PFA overnight at 4°C. The next day, samples were transferred into 100% methanol via a methanol series, including two washing steps in DEPC-PBS and a transfer from 25% to 50% to 75% to 100% methanol/DEPC-PBS (10 minutes each). Upon transfer to 100% methanol, samples were washed twice in 100% methanol and stored in 100% methanol at -20°C. All steps were performed for at least 10 minutes at 4°C.

Pre-hybridization & Hybridization

Unless stated differently, all steps were performed for 10 minutes at 4°C under rocking. For the composition of the solutions used for in situ (Pre-hybridization, Hybridization, Post-hybridization washes and antibody incubation, Post-antibody washes, Staining) we refer to **Data S3**. First, samples were pre-hybridized by transferring to DEPC-PBST via a reverse methanol series (100%, 75%, 50%, 25% for 10 minutes each). Subsequently, samples were incubated in 6% H_2O_2/DEPC-PBST at 4°C, trunk-like-structures/gastruloid for 10 minutes and embryos for 20 minutes, followed by three washes with DEPC-PBST. ProteinaseK/DEPC-PBST (10µg/ml) treatment was performed at 4°C, (7 minutes for trunk-like-structures/gastruloids, 10 minutes for embryos). ProteinaseK activity was quenched with Glycine/DEPC-PBST (2mg/ml) and two washes with DEPC-PBST. Subsequently, samples were post-fixed in 0.2% Glutamine/ 4% PFA/DEPC-PBS for 30 minutes at room temperature and then washed twice with DEPC-PBST. Samples were then incubated with 68°C pre-warmed Hyb for 15 minutes at room temperature after which Hyb was refreshed and structures incubated for 2h at 68°C. If not immediately used for RNA probe hybridization, the samples were pre-cooled for 15 minutes at room temperature and stored at -20°C. Prior to hybridization with the RNA probe, the samples and Hyb solution were pre-warmed at 68°C and incubated in fresh Hyb for 15 minutes at 68°C. Meanwhile, the RNA probe was diluted in Hyb (200ng/ml) and pre-heated for 13 minutes at 80°C in a heating block. The samples were then incubated with the RNA probe at 68°C overnight. The Hyb solution of the 15 minutes incubation step was

stored at -20°C for the first washing step on the next day. All steps were performed under rocking.

Post-hybridization washes, antibody incubation and post-antibody washes

The next day, samples were washed once with Hyb (stored from day before) for 30 minutes at 68°C, twice with Solution 1 for 30 minutes at 68°C, twice with Solution 3T for 30 minutes at 68°C, twice with Solution 3T for 1h at 68°C and three times with TBST for 15 minutes at room temperature. During the 1h washes with Solution 3T, antibody solution was prepared. TBST with one grain of embryo powder was pre-heated for 30 minutes at 70°C in a water bath and cooled down on ice. Subsequently, 1% v/v lamb serum and 0.2% v/v Anti-DIG antibody (*Roche*) were added and incubated for 1h at 4 °C in the dark while rocking. The mix was centrifuged for 10 minutes at 4000rpm at 4°C and 1% lamb serum/TBST was added to the supernatant (final antibody concentration 1:2000). After finishing the washing steps, the samples were blocked with 10% lamb serum/TBST for 2.5h at room temperature and incubated with the antibody solution overnight at 4°C. All steps were performed under rocking. The next day, samples were washed twice with TBST for 15 minutes, twice for 30 minutes and six times for 1h, all at room temperature. The final washing step was performed overnight in TBST at 4°C.

Staining

Samples were washed four times in freshly prepared NTMT for 15 minutes at room temperature under rocking. In the meantime, the staining solution BM Purple (*Roche*) was pre-warmed at room temperature and centrifuged for 1 minute at 13200 rpm. The supernatant was then used for staining the samples and incubated until a clear and specific signal appeared. The first 15 minutes of staining were performed under rocking, afterwards without. For stopping the staining reaction, the samples were washed once with NTT and twice with PBST, for at least 10 minutes each at room temperature under rocking. The samples were stored in 4% PFA/PBS at 4°C. Embryos, trunk-like-structures and gastruloids were imaged with the *AxioZoom.V16* (Zeiss).

Scoring criteria WISH

For WISH structures were picked randomly. For each gene, total sample sizes and reproducibility of the displayed staining patterns are indicated in the figures. In detail:

Uncx: staining bilateral of the anterior-posterior midline (gastruloids), stripy pattern in segments (TLS), staining in the segments (TLS^{C}, TLS^{CL}).

Meox1: staining bilateral of the anterior-posterior midline (gastruloids), staining in the segments (TLS, TLS^{C}, TLS^{CL}).

Tcf15: staining bilateral of the anterior-posterior midline (gastruloids), staining in the segments (TLS, TLS^{C}, TLS^{CL}).

Mesp2*: in-vivo*-like expression stripe at the posterior end (TLS)

Fgf8: staining at posterior end (TLS).

Single-cell transcriptome profiling of TLSs and gastruloids

96h, 108h TLS, 120h TLS, gastruloids and TLS^{CL} were generated as described above. For 96h, 6 structures were selected based on the presence of a T^{mCH+} pole and absence of multiple axes formation; for 108h TLS, 5 structures were selected based on the presence of a T^{mCH+} pole, one axis of elongation, and initiation of neural tube formation, but no already occurring segmentation in the T^{mCH+} domain; for 120h TLS, 3 structures were selected based on the presence of a T^{mCH+} pole, one axis of elongation, clear formation of a neural tube $Sox2^{VE+}$ domain and segmentation in the T^{mCH+} domain; for gastruloids, 5 structures were selected based on the presence of a T^{mCH+} pole, one axis of elongation and a $Sox2^{VE+}$ domain; for TLS^{CL}, 3 structures were selected based on the presence of of a T^{mCH+} pole, elongation along one axis, segmentation in bunches of grapes at the anterior T^{mCH+} domain. TLSs and gastruloids were picked with a p200 with the pipette tip cut-off at the 50µl mark, and serially washed through pipette transferring (cut 200µl tip) in wells filled with 200µl of 1xPBS/0.4%BSA (5 transfers) to get rid of the Matrigel. TLSs of the same time point were then pooled together and dissociated in 200µl TrypLE Express (Gibco) for 15 minutes (96h), 20 minutes (108h TLS) or 25 minutes (120h TLS, gastruloids and TLS^{CL}) at 37ºC, with pipetting every 5 minutes intervals. The cell suspension was filtered using Scienceware Flowmi Cell Strainers, 40µm. Cells were washed

twice with 1ml 1xPBS/0.4%BSA with centrifugation steps performed for 5 minutes at 1200rpm in low DNAbind Eppendorf tubes. The cell concentration was determined using a hemocytometer and cells were subjected to single-cell RNA sequencing (10x Genomics, Chromium™ Single Cell 3' v3; one reaction per timepoint/sample) aiming for a target cell recovery of up to 10,000 sequenced cells per sequencing library (timepoint). Single-cell libraries were generated according to the manual, with one modification: fewer PCR cycles (*n*=8) were ran than recommended during cDNA amplification or library generation/sample indexing to increase library complexity. Libraries were sequenced with a minimum of 230 million paired end reads according to parameters described in the manual.

Bulk RNA-seq of gastruloids and trunk-like-structures

96h aggregates, gastruloids, TLS, TLS^{C} and TLS^{CL} were generated as described above. We sequenced 3 biological replicates per condition (96h aggregates, gastruloids, TLS, TLS^{C} and TLS^{CL}). For 96h, 10 structures per replicate were selected (see previous section for selection criteria) and pooled; for gastruloids, 6 structures per replicate were selected (based on the presence of a T^{mCH+} pole and one axis of elongation) and pooled; for TLS, TLS^{C} and TLS^{CL}, 6 structures per replicate were selected (see previous section for selection criteria; criteria for TLS^{C} were the same as for TLS^{CL}) and pooled. All samples were washed twice with 1xPBS/0.4%BSA. Subsequently, 350µl of RLT Plus buffer containing 1% β-mercaptoethanol (Thermo) was added to dissociate the structures and lyse the cells. After dissociation by trituration and vortexing, samples were frozen at -80°C. The next day, RNA was extracted using RNeasy Plus Micro Kit (*Qiagen*) and RNA concentration and quality was measured using the Agilent RNA 6000 Pico kit on an Agilent 2100 Bioanalyzer. All samples analyzed had a RINe value higher than 8.0, and were subsequently used for library preparation. mRNA libraries were prepared using KAPA Stranded RNA-Seq Kit (*KapaBiosystem*) according to the manufacturer's instructions. 500ng of total RNA was used for each sample to enter the library preparation protocol. For adapter ligation dual indexes were used (NEXTFLEX® Unique Dual Index Barcodes NOVA-514150) at a working concentration of 71nM (5µl of 1uM stock in each 70µl ligation reaction). Quality and concentration of the obtained libraries was measured using Agilent High Sensitivity D5000 ScreenTape on an Agilent 4150 TapeStation. All libraries were sequenced using 75bp-paired end sequencing (150 cycles kit; FC-410-1002) on a HiSeq4000 platform at a minimum of 23.7 million fragments per sample.

Expression profiling of individual TLS using NanoString

mESCs, mouse embryos and TLS were generated, selected and dissociated as described above. For mESCs, a total of 20,000 cells were harvested and frozen in 350µl RLT Plus buffer containing 1% β-mercaptoethanol (Thermo). Three embryos for each developmental stage (E8.5 and E9.5) were decapitated in order to obtain the post-occipital portion, pooled, dissociated, and frozen in 350µl RLT Plus buffer containing 1% β-mercaptoethanol. For 120h TLS, 9 individual structures were selected (based on the selection criteria outlined before), dissociated independently and frozen in 350µl RLT Plus buffer containing 1% β-mercaptoethanol. RNA was isolated from the 12 samples in parallel using RNeasy Plus Micro Kit (*Qiagen*) and RNA concentration was measured using Qubit™ RNA HS Assay Kit.

To profile the expression of 41 genes and 4 housekeeping genes (*Polr1b*, *Hprt*, *Abcf1*, *Gusb*), a range of 36-161 ng total RNA/sample were used in a NanoString nCounter Element assay to profile the 12 individual samples. Probe hybridization was set up according to manufacturer's instructions and performed for 24h (MAN-10040-05). Reactions were ran on the NanoString nCounter SPRINT Instrument. False negative probes detected up to 5 counts, which informed the magnitude of potential false negative signal. Thus, 6 counts were conservatively removed from all measurements. Positive control probes normalization step was applied (geometric mean) and finally the combination of the four housekeeping genes was used to obtain the final normalized counts table. To assess variability in the expression of individual marker genes, the $\log_2$(MAX(all_samples)/MIN(all_samples)) was used as a proxy with a high value representing high variability and vice versa. See **Data S7** for probes design and sequences and for normalized gene counts.

Computational analysis

If not stated otherwise: All statistics and plots are generated using R version 3.6.0 "Planting of a Tree" and Seurat version 3.0 (*59*).

Single-cell transcriptome profiling of TLSs

Preprocessing

The Cell Ranger pipeline version 3 (10x Genomics Inc.) was used for each scRNA-seq data set to de-multiplex the raw base call files, generate the fastq files, perform the alignment against

the mouse reference genome mm10, filter the alignment and count barcodes and UMIs. Outputs from multiple sequencing runs were also combined using Cell Ranger functions.

Quality control

The initial quality control was performed using scanpy (*60*). Cells with less than 10,000 or more than 40,000 counts, a mitochondrial-fraction above 0.1 and less than 3,000 genes were flagged as insufficient and therefore removed from the analysis (see below).

Cluster determination

Single cell data created for the three developmental time points (96h, 108h and 120h) were loaded to Seurat (*59*), with a minimum requirement of 200 features and 3 cells and filtered for previously flagged barcodes. Subsequently, the expression data were independently normalized, variable features were detected and log-normalized and scaled to 10,000 (default settings). Next, for downstream integration of the three time points sequenced, a PCA was run for each time point prior to integration anchor set detection (reduction = "rpca", dims = 1:30). Finally, these integration anchors were used to integrate the data points using the previously calculated anchor sets. A list of cell cycle markers loaded with Seurat was used to cell cycle score all cells and subsequently run the default workflow for scaling with vars.to.regress set to cell cycle scores for S and G2M phase (**Fig. S24A**). For downstream analysis and visualization of the integrated dataset, a PCA was run to then calculate a joint UMAP (dims = 1:30, n.neighbors = 10). Standard workflow steps were applied for cluster generation (FindNeighbors, dims = 1:20 and FindClusters, resolution = 0.5), resulting in a total of 15 clusters. Finally, two small clusters (Seurat_10 and Seurat_11) were removed due to the presence of stressed cells (high mitochondrial RNA counts and low total RNA counts) (Seurat_10; **Fig. S24B**) and doublets (shown by almost double amount of total RNAs counts and UMIs) (Seurat_11; **Fig. S24C**). All remaining clusters show a similar distribution of average UMIs and genes detected per cell (**Fig. S24D**). See **Data S8** for marker genes list for each identified cluster.

Subclustering of somitic and neural cells in TLS

For subclustering of somitic cells, data were first split by sampled timepoints and somitic clusters (“Somite (0)”, “Somite”, “Somite (Dermo)”, “Somite (Sclero)”) were extracted. Subclustering was then performed in Scanpy (resolution=0.3). For subclustering of the neural

tube cluster, cells assigned to “NeuralTube2” from TLS were extracted. Subclustering was then performed in Scanpy (resolution=0.65).

Comparison with gastruloids and TLS^{CL}

The Cell Ranger pipeline version 3 (10x Genomics Inc.) was used for the gastruloids and TLS^{CL} scRNA-seq data sets to de-multiplex the raw base call files, generate the fastq files, perform the alignment against the mouse reference genome mm10, filter the alignment and count barcodes and UMIs.

The initial quality control was performed in R using the Seurat package. Cells with a minimum feature count of 3,000, an RNA count of 10,000 to 80,000 and less than 5% mitochondrial genes, were kept for analysis. Subsequently, the expression data were independently normalized, variable features were detected and log-normalized and scaled to 10,000 (default settings).

Each cell of the gastruloids and TLS^{CL} was then clustered with respect to TLS cell states. After single cell transformation (SCTransform, default settings), transfer anchors were found with the 120h TLS cells as reference (FindTransferAnchors, dims=1:30) and cell state labels were transferred (TransferData, dims=1:30).

Next, for downstream integration of the three conditions (120h TLS, gastruloids and TLS^{CL}), a PCA was run for each time point prior to integration anchor set detection (reduction = “rpca”, dims = 1:30). Finally, these integration anchors were used to integrate the conditions using the previously calculated anchor sets. A list of cell cycle markers loaded with Seurat was used to cell cycle score all cells and subsequently run the default workflow for scaling with vars.to.regress set to cell cycle scores for S and G2M phase. For downstream analysis and visualization of the integrated dataset, a PCA was run to then calculate a joint UMAP (dims = 1:10).

Subclustering of somitic, neural and endothelial cells in the TLS, gastruloid and TLS^{CL} integrated dataset

For subclustering of somitic cells, somitic clusters (“Somite (0)”, “Somite”, “Somite (Dermo)”, “Somite (Sclero)”) were extracted from the integrated data-set. Subclustering was then performed in Seurat 3.1 (resolution=0.8). For subclustering of endothelial cells, cells assigned to “Endothelial” were extracted. Subclustering was then performed in Seurat 3.1 (resolution=0.1). For subclustering of the neural cells, cells assigned to “NeuralTube2” were extracted. Subclustering was then performed in Seurat 3.1 (resolution=0.4).

Comparison with mouse embryo

A previously established reference data set of mouse post-implantation development (E6.5 to E8.5) *(29)* was utilized for gene expression comparison to *in vivo* data as well as cell states and proportions comparisons. The *in vivo* data were filtered to include only the relevant embryo time points (E.7.5, E8.0 and E8.5) and all extra-embryonic and occipital cell-states (that are not generated in gastruloids *(3)* and TLS (this study)) were removed. See **Data S8** for marker genes lists for each identified cluster. Both data sets (the mouse embryo (*in vivo*) and the TLS (*in vitro*)) were normalized in parallel (SCTransform, default settings). To adjust for different resolutions in cell state detection between the embryo and TLS datasets, the *in vivo* "somites" and "PSM" cell states were subclustered (resolution=0.15) to identify a cell cluster with an anterior PSM (aPSM) signature (**Fig. S24E,F**). These cells were relabeled accordingly and subsequently treated as an individual cell state. The subclustering step also generated two clusters from the *in vivo* PSM original cell state, but due to marker genes similarity this subclustering was not taken into account in the further downstream analysis (**Fig. S24E,F**). Prediction scores were used as measurement for certainty in cell type matching ("certainty scores"). Scores shown are means across all cells even if not assigned to the respective cluster. Finally, the *in vivo* mouse and *in vitro* TLS cell types were matched using an integrated reference, by first finding of transfer anchors, followed by data transfer using the anchors and finally the predicted cell states were used as *in vivo*/*in vitro* matched cell state counterparts. For differential expression and conserved marker calculation both data sets were merged and genes located on the sex chromosomes were excluded to avoid biases resulting from comparison of only male TLS with *in vivo* cells from male and female embryos.

Analysis of conserved marker genes

The conserved marker genes were calculated on the integration of the TLS and embryo data as described above using the FindConservedMarkers function with default setting. Cut-offs used were set to a minimum percentage difference of 25% (for percent of cells expressing the gene of interest (GOI)), and an absolute minimum $\log_2$FC of 1. See **Data S9** for lists of top 25 conserved marker genes.

Analysis of differentially expressed genes between TLS and embryo

TLS and embryo data were integrated as described above. For the analysis, the *in vivo* cell states were used as a reference, and compared to the TLS cells attributed to that cluster. To this end, the percent of cells expressing a gene as well as the average expression for that gene in the *in*

vivo cell states as well as the *in vitro* cells attributed to that cluster was computed using the FindMarkers function with default settings. Differential expression between TLS and embryo was tested for either all the genes expressed in the cluster of interest, or the 20 top marker genes of each *in vivo* cell states. Cut-offs used were a minimum percentage difference of 25% (for percent of cells expressing the gene of interest (GOI)), and an absolute minimum log_2FC of 1. See **Data S9** for list of genome wide differentially expressed marker genes.

RNA velocity and pseudo-time analysis

RNA velocity was calculated using the velocyto tool *(25, 61*) and visualized using scanpy (*60*). Trajectory inference and pseudo-time analysis were done using the scanpy package PAGA (*62*). The previously calculated UMAP and cell cluster assignment was used for velocity projection and trajectories for pseudo-time. For pseudo-time analysis, somitic clusters ("Somite (0)", "Somite", "Somite (Dermo)", "Somite (Sclero)") were merged, the putative NMP cluster was set as root and 108h and 120h cells were analyzed separately.

Differentiation trajectories in TLS

We used Harmony (*63*), an algorithm to bridge time points, and combined it with Palantir (*63, 64*), an algorithm that infers cell fate potential, as previously published by the Peer lab (*63*); In line with their analysis, we used these algorithms to construct a temporal map of the developing TLS to identify when fate decisions occur.

After loading gene-barcode matrices, we performed normalization, variable gene selection and log transformation using the Harmony default settings. Next, we used Harmony to calculate the augmented affinity matrix and force directed layout for the 96h, 108h and 120h TLS samples. To infer cell fate potential and branch probabilities we used a 96h NMP cell as start cell (BC: GACCCAGAGTTGGCGA).

Using Palantir (default settings), we predicted two differentiation branches- one with an endpoint in the neural cluster and one with an end point in the somitic clusters and a hot spot of differentiation potential in the NMP cluster.

Subsequently, we used the MAGIC imputation algorithm (*65*) to visualize the expression of T and Sox2 along the two branches and visualized onto the layout as well as a gene trend line graph using the inferred pseudo-timing. Finally, for the NMP cells, the normalized expression of Sox2 and T (as determined by Seurat) were contrasted with their differentiation potential and the expression of Tbx6 and Sox1.

Bulk RNA-seq of gastruloids and trunk-like-structures

Processing and expression levels

RNAseq triplicates of 96h aggregates, 120h gastruloids, TLS, TLS^{C} and TLS^{CL} were pre-processed using cutadapt (*66*) to remove adapter and trim low quality bases. Reads were subsequently aligned against the reference genome mm10 using STAR (*67*) (parameter: outSAMtype BAM SortedByCoordinate --outSAMattributes Standard --outSAMstrandField intronMotif --outSAMunmapped Within --quantMode GeneCounts). Finally, Stringtie (*68*) was used for transcript assembly, e.g. calculation of strand-specific TPMs. The three biological replicates highly correlate between each other (**Fig. S24G**).

Differential gene expression

Differential expression analysis was done independently per group comparison using the R package DESeq2 (*69*) using the raw expression counts from STAR's reads per gene output. *z*-scores were calculated according to the formula ((VALUE(sample)-AVERAGE(all_samples))/STDEV(all_samples). TPM values for all genes can be found in **Data S10**.

Gene Set Enrichment Analysis (GSEA)

GSEA was performed using the R package "fgsea" (*70*) (minSize=5, maxSize=500, number of permutations = 10,000) taking all genes statistically differentially expressed between TLS and gastruloids (FDR < 0.05, no log_2FC cut-off).

Plotting

Plots were generated with GraphPad Prism 7, R 3.6.0 and R 3.6.1, Python, and PlotTwist *(71)*.

Data availability

Bulk and single-cell RNA-Seq data have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE141175.

Code availability

All computational code used in this study is available upon request.

Statistical analysis:

One-way ANOVA followed by Tukey's multiple comparison test was used to calculate p-values. Raw data with associated statistics are provided in **Data S1** and **Data S5.**

Science
AAAS

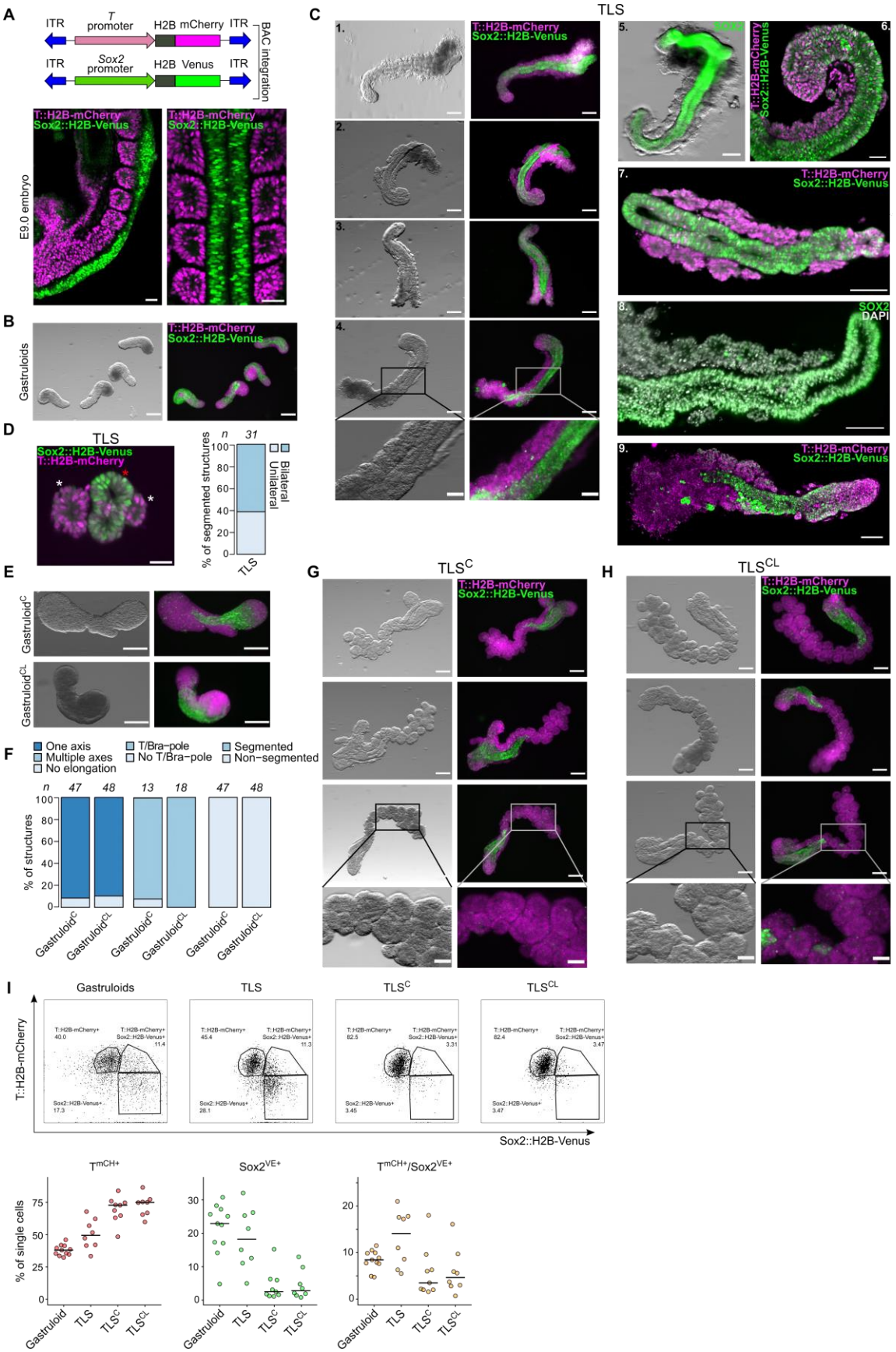

Fig. S1. Induction of embryo-like morphology in trunk-like-structures

(A) Schematic of *T::H2B-mCherry*; *Sox2::H2B-Venus* double reporter mESCs (upper panel) and *in vivo* validation by tetraploid complementation (bottom panel). Optical sections of light sheet imaging of an E9.0 mouse embryo caudal end (bottom-left panel) and of somites and neural tube (bottom-right panel). Scale bar 50µm. (**B**) Wide-field fluorescent and bright-field images of 120h gastruloids. Samples are representative for n=3 independent experiments. Scale bar 200µm. (**C**) Lightsheet, confocal and wide-field microsopy of 120h TLS. **1-5**. Wide-field; **6,9.** optical sections confocal imaging; **7,8.** Optical sections of light sheet imaging. TLS were obtained in n=6 different experiments. (**D**) Lightsheet imaging showing in transversal section somites and neural tube relative position (somites are bilateral to the neural tube) and quantification of the frequencies of uni vs bilateral TLS. White asterisks indicate somites, red asterisk indicates the neural tube. Scale bar 50µm. (**E**) Wide-field fluorescent and bright-field images of 120h gastruloids treated with CHIR or CHIR+LDN from 96 to 120h showing no signs of segmentation in the absence of 5% matrigel. Samples are representative for n=2 independent experiments. (**F**) Quantifications of gastruloids from conditions in (**E**). Treatment of gastruloids with CHIR and LDN does not significantly alter the axis formation and induction of a T^{mCH+} pole, and is not able to induce segmentation. Samples were obtained from n=2 independent experiments. (**G**) Wide-field fluorescent and bright-field images of 120h TLS^{C}. (**H**) Wide-field fluorescent and bright-field images of 120h TLS^{CL}. In **B,C,E** and **G,H**, scale bar is 200µm, except for magnifications, where scale bar is 50µm. Samples are representative for n=6 independent experiments. (**I**) Flow cytometry raw plots to demonstrate gating settings and corresponding quantification of percentage of T^{mCh+} (mesodermal), $Sox2^{VE+}$ (neural), and $T^{mCh+/}Sox2^{VE+}$ (neuromesodermal) cells in individual gastruloids and TLSs. Dots represent individual TLS/gastruloids, line indicates median. Gastruloids were measured with BD Celesta, TLSs with BD FACSAria II. Samples were obtained in n=3 independent experiments. Statistical analysis in **Data S1**.

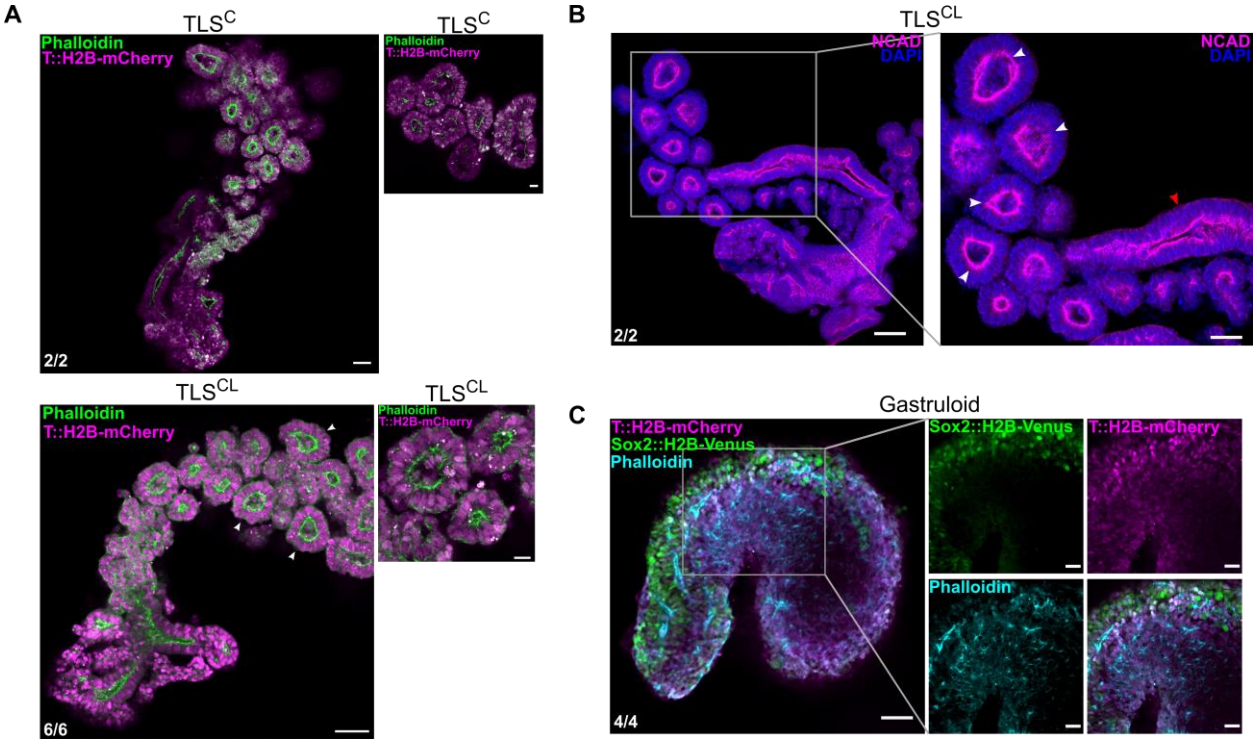

Fig. S2. Apical-basal polarity in trunk-like-structures and gastruloids

(**A**) Confocal sections showing F-actin localization in TLS^{C} and TLS^{CL} as visualized by Phalloidin staining. Note the clear apical-basal polarity of cells in the somites. Scale bar 50µm, 20µm for the magnification. (**B**) Confocal sections showing NCAD staining in TLS^{CL}. Both in the somites (white arrowheads) and neural domain (red arrowhead) cells are polarized with NCAD accumulated at the apical surface. Scale bar 100µm, 50µm for the magnification. (**C**) Confocal sections showing F-actin localization in gastruloids as visualiyed by Phalloidin staining. For each staining the proportion of structures exhibiting the reported pattern is shown. Scale bar 50µm, 20µm for the magnifications.

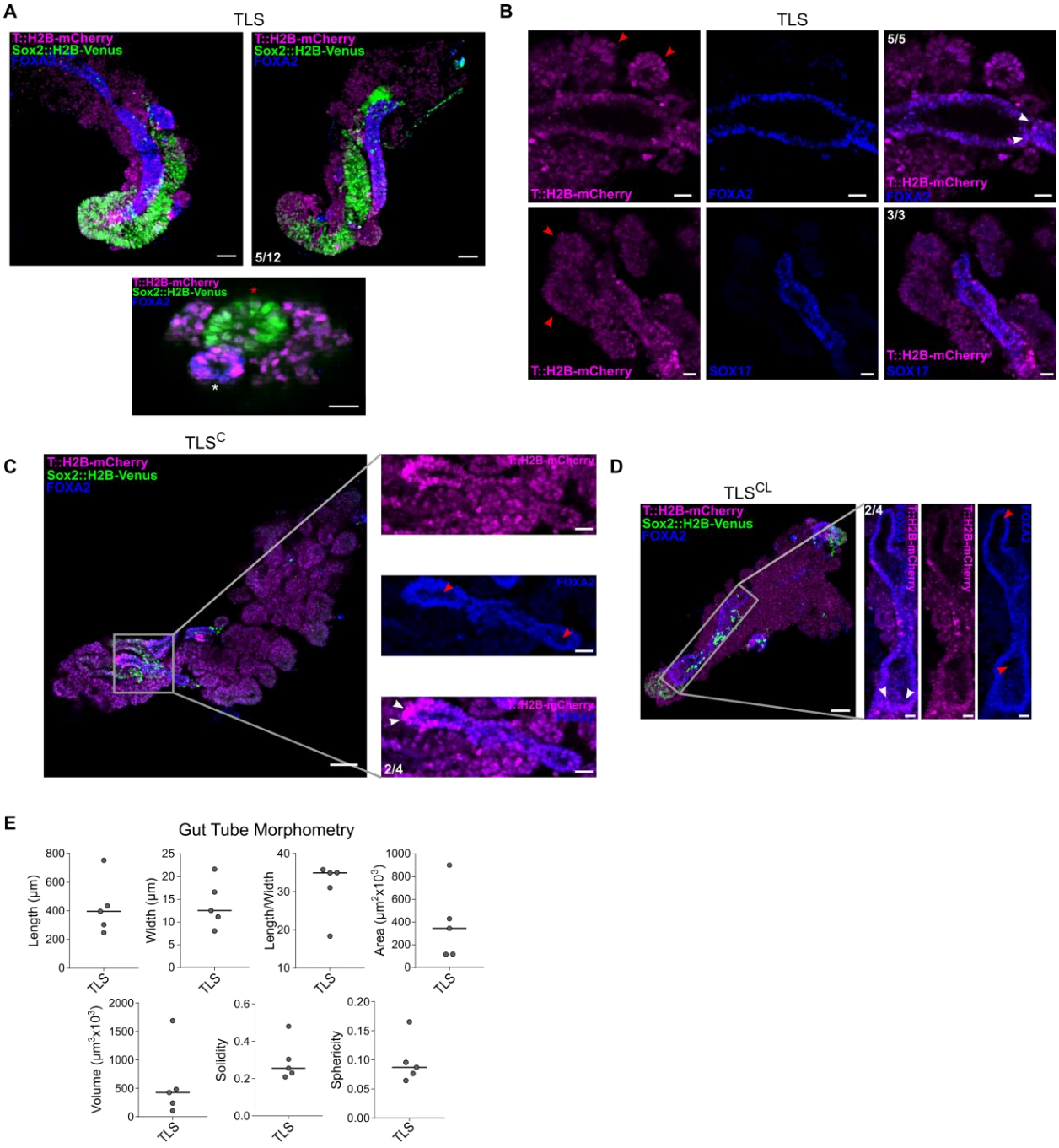

Fig. S3. Formation of gut-like-structures in trunk-like-structures

(A) 3D maximum intensity projection of TLS (upper panels) stained for FOXA2 (blue) showing the gut-like-structure. Scale bar 50µm. Transversal view of a 3D maximum intensity projection of TLS (lower panel) showing the relative position of neural tube and the gut-like-structure. Red asterisk indicates the neural tube, white asterisks indicates the gut-like-structure. Scale bar 20µm. (**B**) Confocal sections showing the FOXA2^{+} (upper panel) and SOX17^{+} (bottom panel) gut-like structure in TLS. Note FOXA2^{+}/T$^{mCH\text{-}high}$ cells at the base of the gut (white arrowheads). Red arrowheads indicate somites. Scale bars 25µm. (**C**) Confocal section showing gut-like structure in TLS^{C}. (**D**) 3D maximum intensity projection showing gut-like structure in TLS^{CL}. In the magnifications (right panels) in (**C,D**) white arrowheads indicate FOXA2^{+}/T$^{mCH\text{-}high}$ cells at the base of the gut. Red arrowheads indicate gut tubular

cavity. In **C-D** scale bar 50µm for 3D maximum intensity projections, 20µm for magnifications. For all stainings in **A-D** the proportion of structures exhibiting the reported pattern is shown. (**E**) Distributions of gut tube length, width, length/width, area, volume, solidity and sphericity in TLS (N=5). Line represents the median.

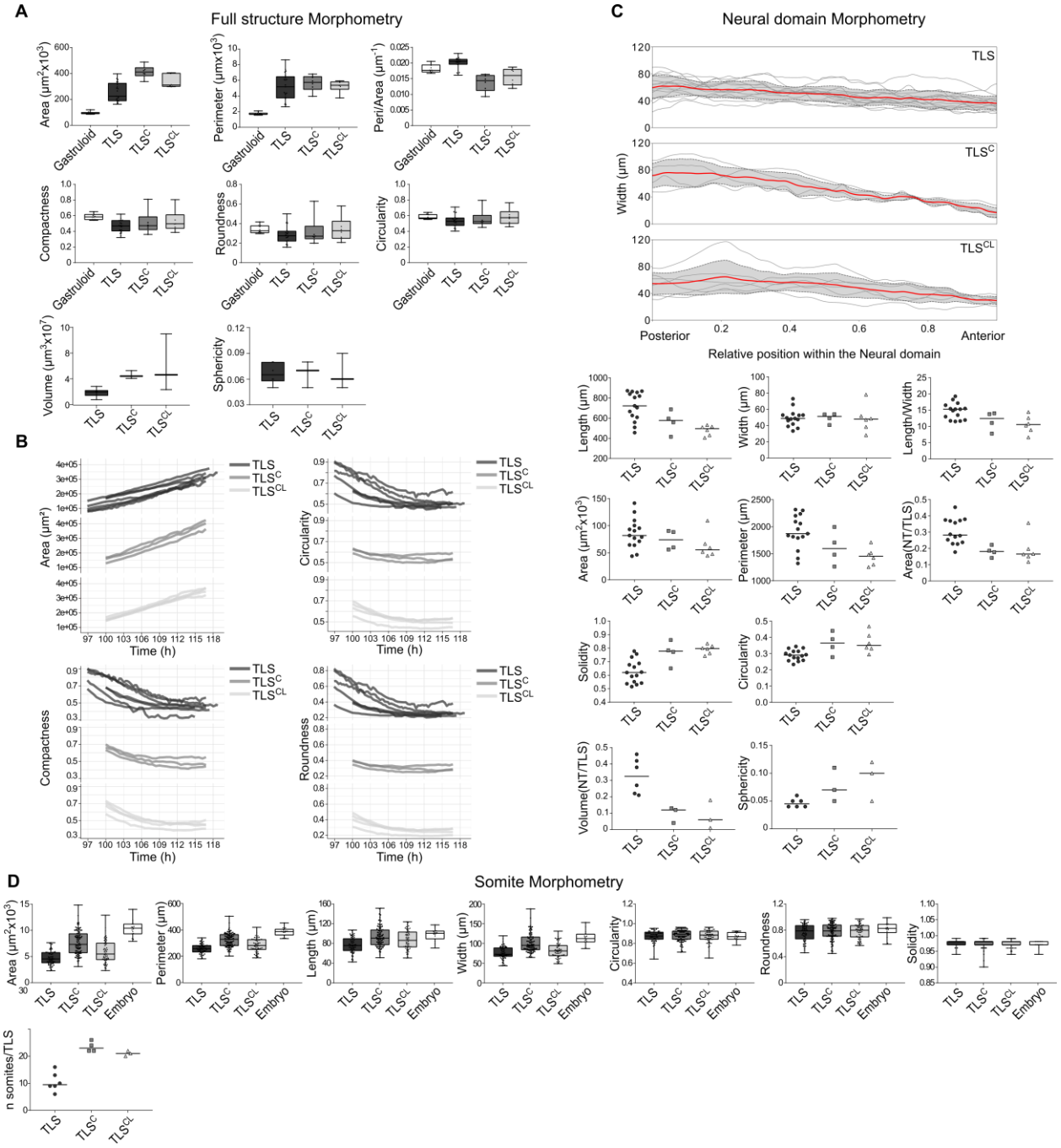

Fig. S4. Morphometric measurements in gastruloids, TLS, TLS^{C} and TLS^{CL}

(A) Boxplots showing area, perimeter, perimeter/area, compactness, roundness and circularity distributions for gastruloids TLS, TLS^{C} and TLS^{CL} whole structures and volume, sphericity for TLS, TLS^{C} and TLS^{CL} whole structures. Boxes indicate interquartile range. End of whiskers represent minimum and maximum. Symbols indicate individual structures. Central line represent the median. Gastruloids (N=10), TLS (N=21), TLS^{C} (N=6), TLS^{CL} (N=6). (**B**) Time-resolved whole-structure morphometry of TLS (N=7), TLS^{C} (N=3), TLS^{CL} (N=4). (**C**) Width of the neural domain along the anterior-posterior axis in individual TLS, TLS^{C} and TLS^{CL} (upper panel) and distributions of length, width, length/width, area, area neural/area whole structure, perimeter, solidity, circularity volume neural/volume whole structure and sphericity of the neural domain in TLS, TLS^{C} and TLS^{CL}. Symbols

indicate individual structures. Line represents the median. TLS (N=15), TLS^{C} (N=5), TLS^{CL} (N=6). (**D**) Boxplots showing area, perimeter, length, width, solidity, roundness and circularity distributions for TLS, TLS^{C},TLS^{CL} and E8.5 mouse embryo somites and distribution of number of somites per trunk-like-structure. Boxes indicate interquartile range. End of whiskers represent minimum and maximum. Symbols indicate individual somites. Central line represents the median. TLS (N=6; n=63), TLS^{C} (N=4; n=95), TLS^{CL} (N=3; n=63), embryo (N=2; n=16). N = number of gastruloids/TLSs/embryo; n = number of somites. Statistical analysis for all comparisons in **Data S5**.

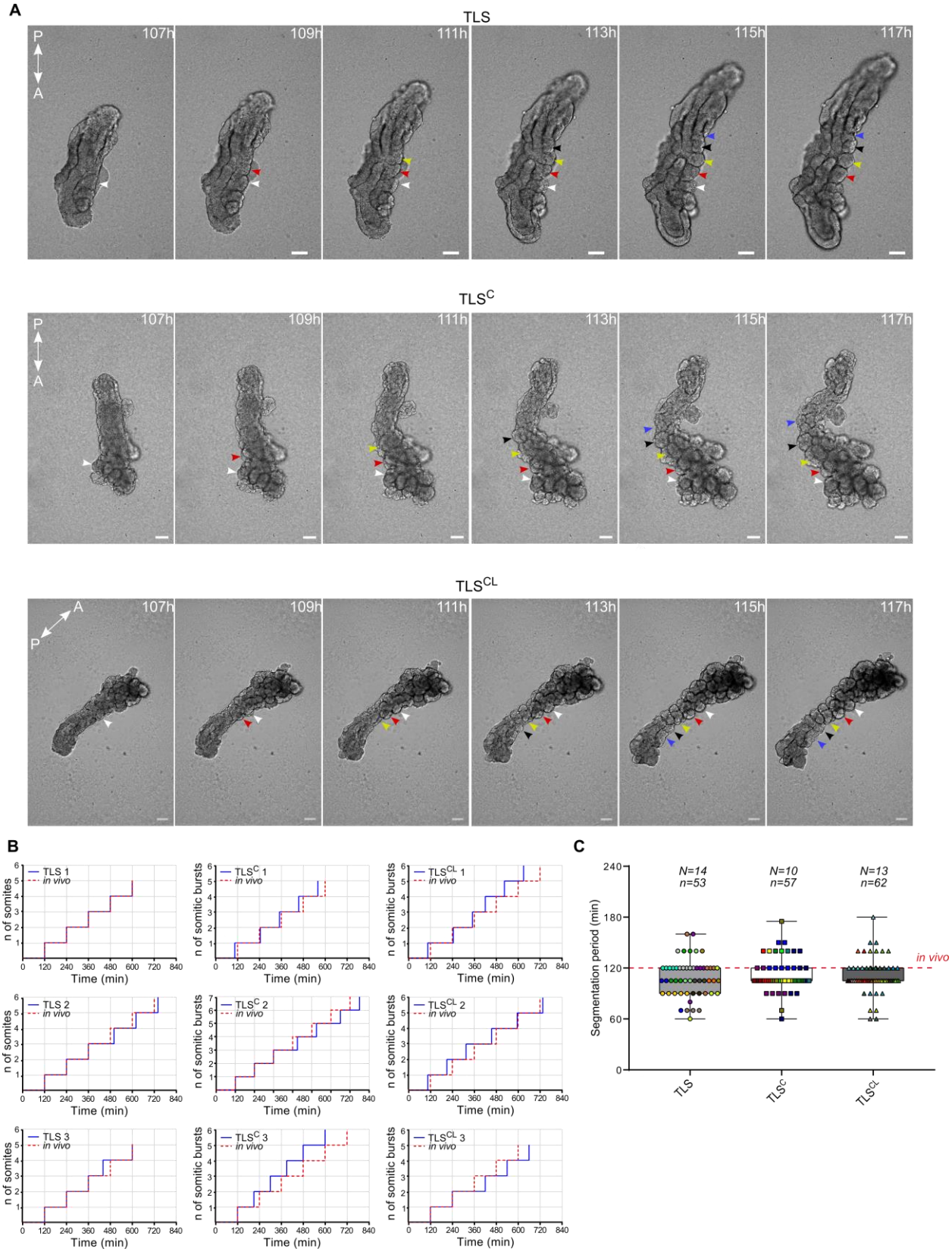

Fig. S5. Sequential segmentation in TLS, TLS^{C} and TLS^{CL} somites

(**A**) Stills from bright-field live imaging showing sequential generation of somites along the anterior-posterior axis in TLS, TLS^{C} and TLS^{CL}. Arrowheads indicate individual somites (in TLS) or somitic bursts (in TLS^{C} and TLS^{CL}) throughout time, with individual colours tracking individual somites or

somitic bursts. Scale bar 100µm. A, anterior; P, posterior. (**B**) Generation of new somites (in TLS) or somitic bursts (in TLS^{C} and TLS^{CL}) over time in individual structures. Red line indicates the *in vivo* 120 minutes interval. (**C**) Boxplot showing the segmentation period for TLS, TLS^{C} and TLS^{CL} as compared to *in vivo*. Boxes indicate interquartile range. End of whiskers represent minimum and maximum. Symbols indicate individual somites, that are color coded by the structure of origin. Central line represent the median. N = number of TLSs; n = number of somites. Structures were obtained in at least n=3 independent experiments.

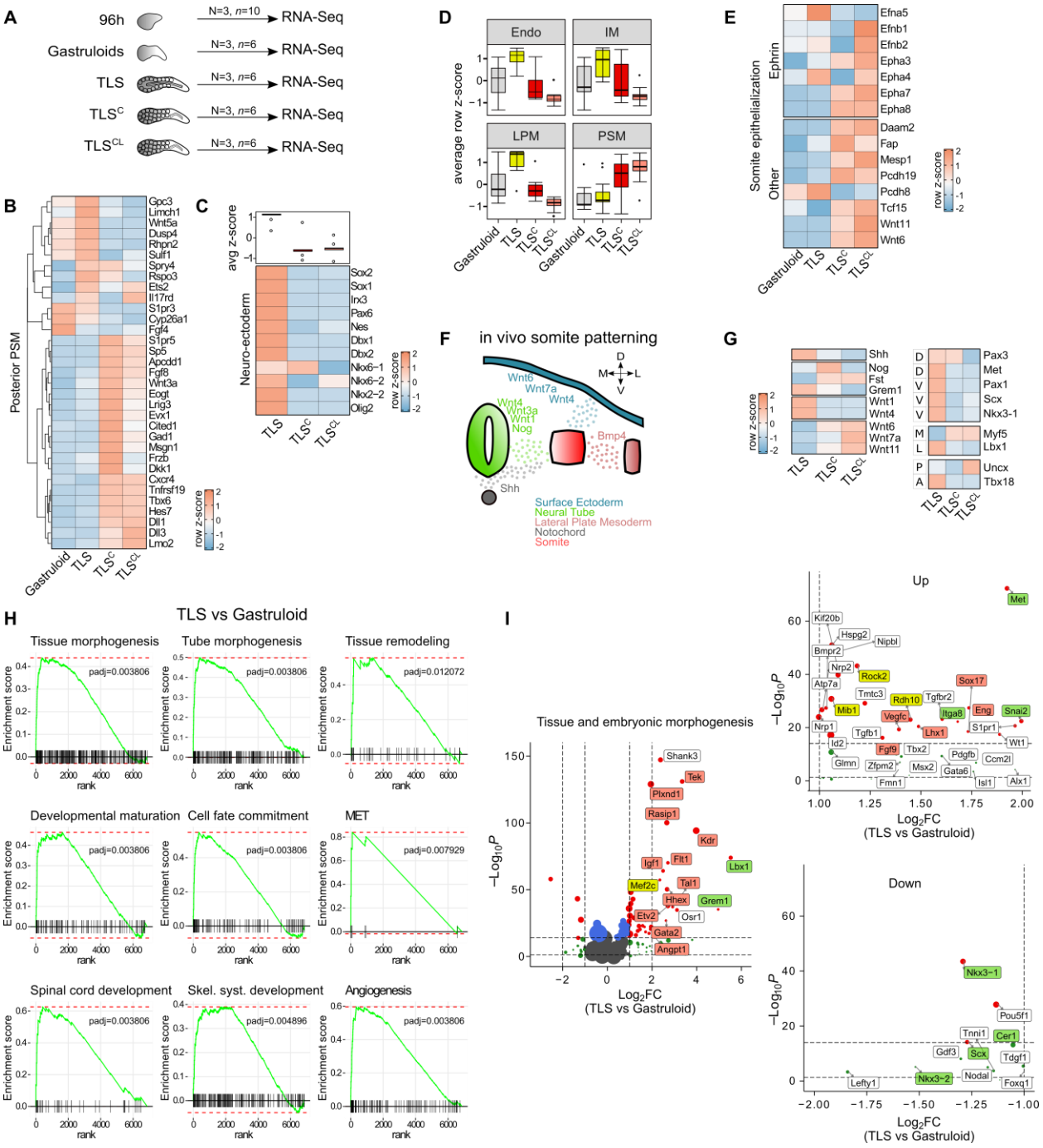

Fig. S6. Bulk RNA-Sequencing of gastruloids and trunk-like-structures

(**A**) Schematic of experimental set-up. N=number of replicates, *n*=number of pooled structures per replicate. (**B**) Heatmap with scaled expression (row *z*-score) of known marker genes of posterior presomitic mesoderm (pPSM) *(12)*. Scores are an average of the 3 independent replicates per condition. (**C**) Heatmap with scaled expression (row *z*-score) of neural tube known marker genes. Scores are an average of the 3 independent replicates per condition. Boxplot shows distribution of *z*-scores (per column) for different conditions. Dots indicate outliers. (**D**) Boxplots showing distribution of *z*-scores for known marker genes of indicated tissue layers. Endo, Endoderm; IM, Intermediate Mesoderm; LPM, Lateral Plate Mesoderm; PSM, Pre-Somitic Mesoderm. Boxes indicate interquartile range. Whiskers

extends to 1.5xIQR from the hinge. Dots indicate outliers. Central line represent the median. (**E**) Heatmap of scaled expression (row *z*-score) of somite epithelialization factors in gastruloids, TLS, TLS^{C} and TLS^{CL}. Scores are an average of the 3 independent replicates per condition. (**F**) Schematic overview of signaling factors involved in somite compartmentalization *in vivo*. M, medial; L, lateral; D, dorsal; V, ventral. (**G**) Heatmap with scaled expression (row *z*-score) of (left panel) signaling factors depicted in (**F**) and (right panel) marker genes for different somite compartments *in vivo*. A, anterior; P, posterior; M, medial; L, lateral; D, dorsal; V, ventral. Scores are an average of the 3 independent replicates per condition. (**H**) Gene Set Enrichment Analysis enrichment plots of selected significant pathways. (**I**) Volcano plot of genes involved in tissue and embryonic morphogenesis (first panel). Dot size scales with $\log_2$ of absolute expression. Red dots, absolute $\log_2FC>1$ and padj$<10e^{-15}$. Green dots, absolute $\log_2FC>1$. Upper dotted line, padj (FDR)$=10e^{-15}$; bottom dotted line, padj (FDR)=0.05. Up- and downregulated genes involved in embryo & tissue morphogenesis (second two panels). Dot size scales with $\log_2$ of absolute expression. Red dots, $\log_2FC>2$ and padj$<10e^{-15}$. Green dots, $\log_2FC > 2$. Upper dotted line, padj$=10e^{-15}$; bottom dotted line, padj(FDR)=0.05. Green label, involved in somitogenesis; orange label, involved in blood vessel development; yellow label, involved in both (*33–41*). List of genes used for each category in (**C**), (**D**) and (**E**) and statistical analysis in **Data S1.**

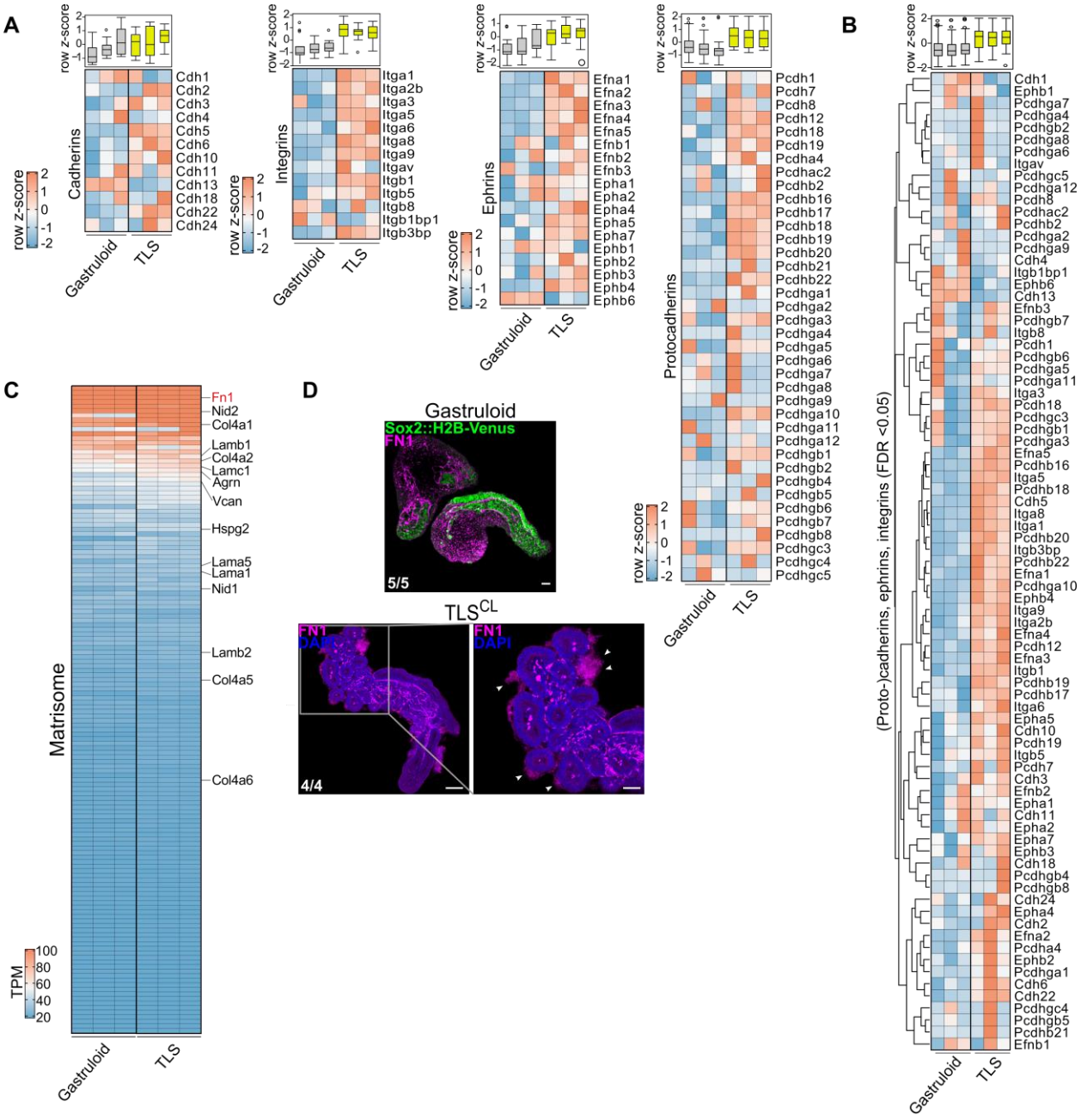

Fig. S7. Expression of cell adhesion molecules in gastruloids and TLS

(**A**) Heatmap with scaled expression (row *z*-score) of cadherins, protocadherins, and ephrins in gastruloids and TLS. Boxplots show distribution of *z*-scores (per column) for different samples. Boxes indicate interquartile range. Whiskers extends to 1.5xIQR from the hinge. Dots indicate outliers. Central line represent the median. Every column in each heatmap represents one of the three biological replicates. (**B**) Heatmap of scaled expression of genes in (**A**) with significantly differential expression (FDR<0.05). (**C**) Heatmap showing TPM values for the matrisome related genes in gastruloids and TLS. Highlighted are basement membrane components. Note that Fibronectin (Fn1) is the highest expressed basement membrane component in both gastruloids and TLS. Every column represents one of the three biological replicates. (**D**) Confocal section showing FN1 staining in gastruloids (upper panel) and TLS^{CL} (lower panels). White arrowheads indicate FN1 accumulated around the somite structures in TLS^{CL}. The

proportion of structures exhibiting the reported pattern is shown in the left bottom corner. Upper panel scale bar 50μm. Lower panel 100μm, 50μm for magnification. Statistical analysis in **Data S1.**

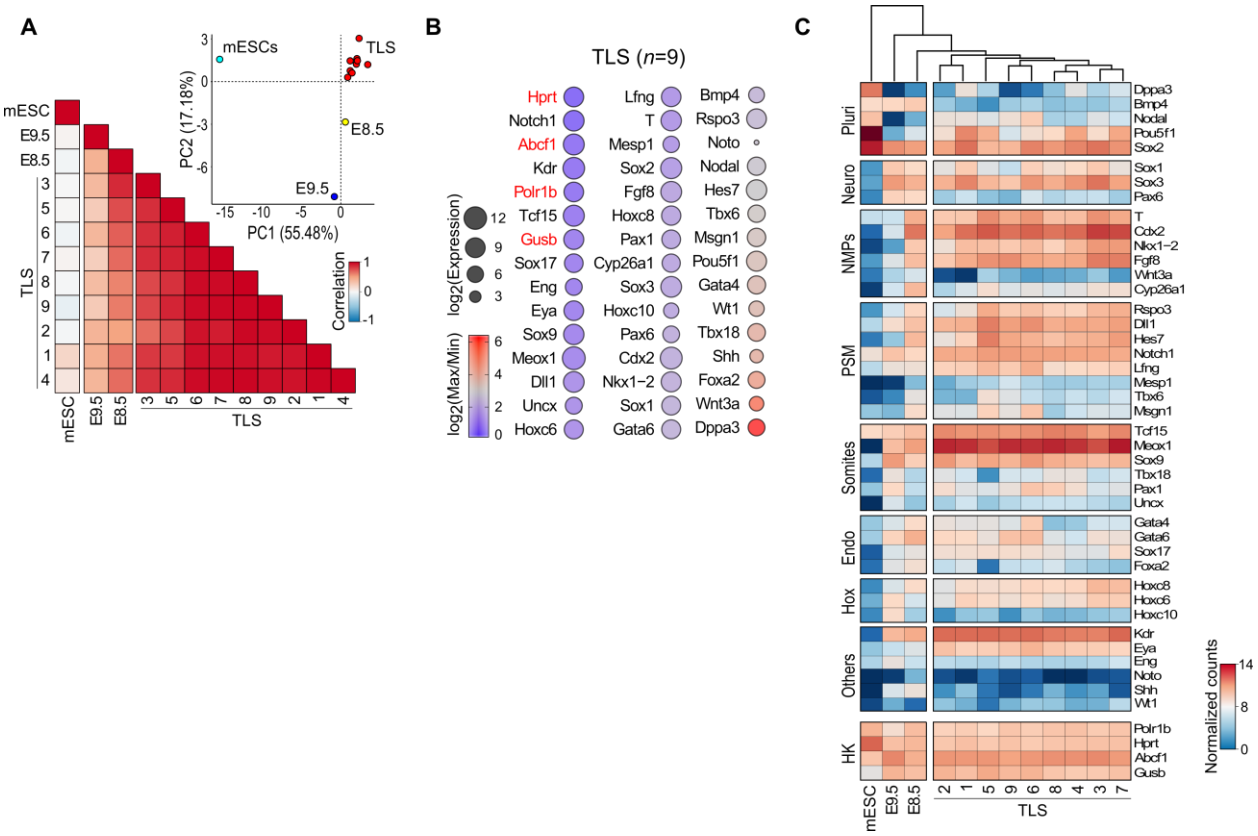

Fig. S8. NanoString analysis of individual TLS

(**A**) Heatmap correlation plot and PCA analysis of mESCs, E8.5 and E9.5 post-occipital embryo, and 9 individual TLS based on expression of 41 developmental genes measured by NanoString analysis. PC1 and PC2 represent the two components with highest percentage of explained variance. (**B**) Dot plot for 45 genes (41 test genes and 4 housekeeping genes (red font)). Dot size scales with $\log_2$ of expression. Color indicates $\log_2$ (max/min) as a proxy for range of expression. Genes are ranked from lowest range (upper left) to highest range (bottom right). (**C**) Heatmap with normalized counts of indicated *in vivo* marker genes for embryonic cell types in mouse embryonic stem cells (mESCs), the E8.5 and E9.5 post-occipital embryo, and 9 individual TLS. Pluri, pluripotency; Neuro, neuro-ectoderm; NMPs, neuromesodermal progenitors; PSM, presomitic mesoderm; Endo, endoderm; Hox, Hox genes; HK, housekeeping genes. Note i) high correlation between TLSs and the E8.5 post-occipital stage. ii) Endothelial, NMP, NE, and somitic ME genes were reproducibly expressed at similar levels across all replicates, whereas higher variation is present in the expression of PSM, endodermal, and pluripotency genes.

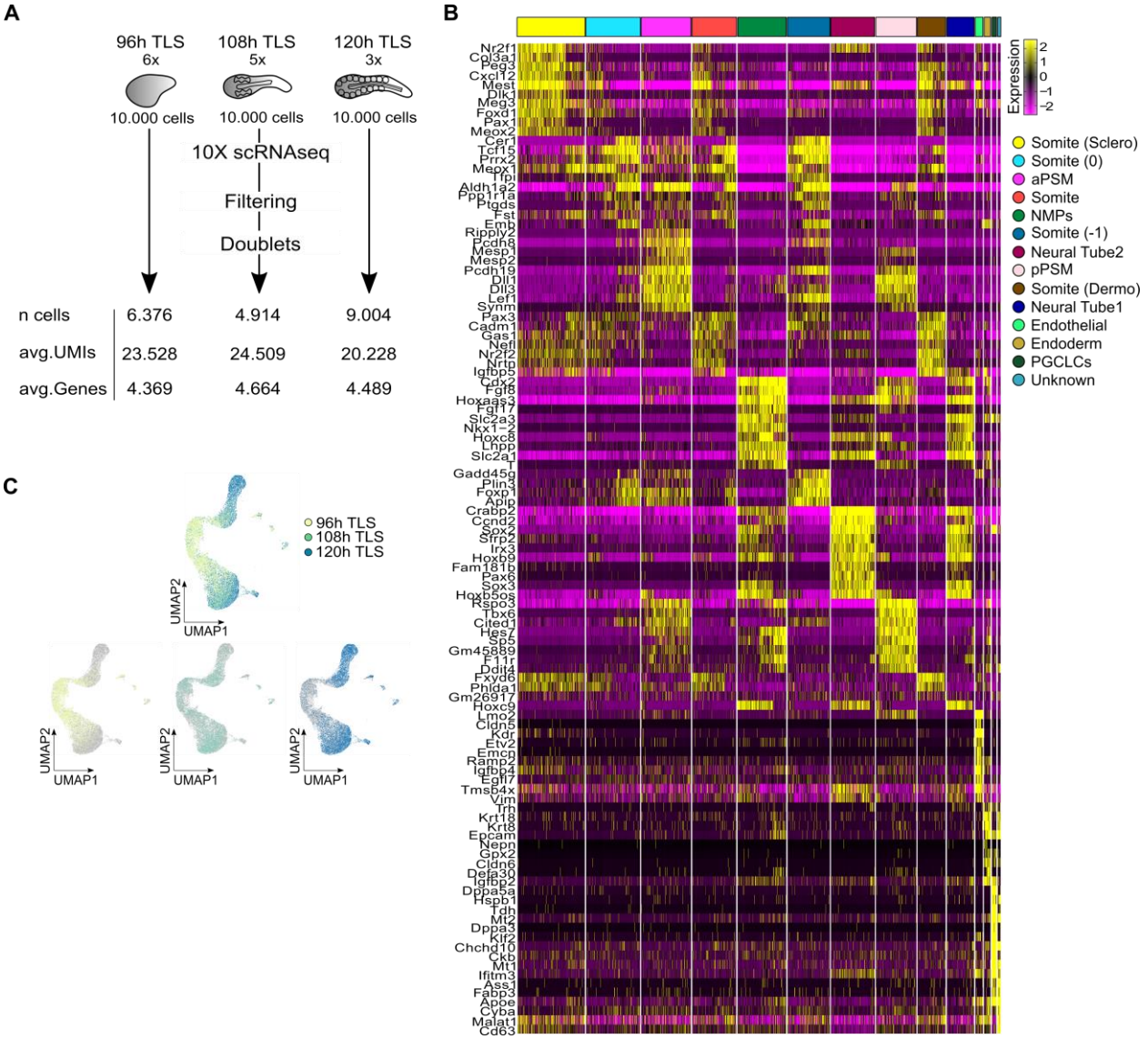

Fig. S9. Single-Cell RNA-sequencing of TLS

(A) Experimental set-up. Criteria for filtering and removal of doublets are described in Supplemental Information. (**B**) Heatmap with scaled expression of top marker genes of the 14 identified clusters. (**C**) UMAPs coloured by sampled time-points.

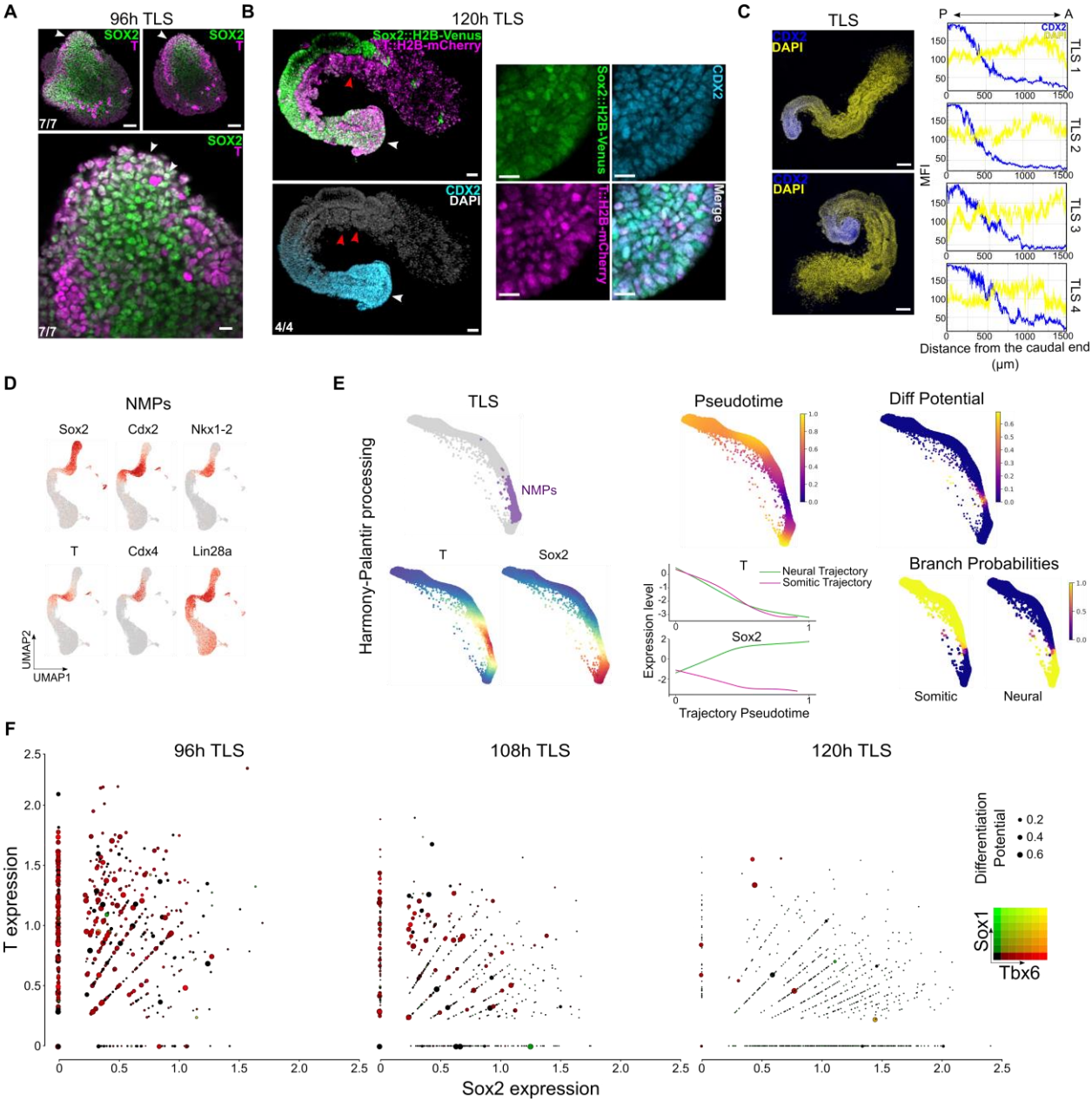

Fig. S10. Spatial allocation and analysis of differentiation dynamics of neuromesodermal progenitors in trunk-like-structures

(**A**) Confocal section showing the location of SOX2^{+}/T^{+} putative neuromesodermal progenitors (NMPs) in 96h trunk-like-structures. White arrowheads indicate SOX2^{+}/T^{+} domain/cells. The proportion of structures exhibiting the reported pattern is shown. Scale bars 50μm for upper panels, 20μm for magnifications. (**B**) 3D maximum intensity projection (whole structure) and confocal sections showing co-expression of CDX2, T::H2B-mCherry and Sox2::H2B-Venus in NMPs at the posterior end of TLS. The proportion of structures exhibiting the reported pattern is shown. Scale bar 50μm, magnification 20μm. (**C**) Immunostaining showing the localization of CDX2^{+} cells (left panel) and quantification of CDX2 and DAPI Mean Fluorescent Intensity (MFI) along the Anterior-Posterior axis of four trunk-like-structures. Scale bars 100μm. (**D**) UMAPs coloured by expression of indicated *in vivo* marker genes for NMPs. (**E**) Diffusion maps showing NMP cluster and T/Sox2 expression after Harmony processing of

single cell datasets from 96h, 108h and 120h TLS (left panel). Anti-correlation between pseudotime and differentiation potential in the same Harmony diffusion map space with the NMP cluster as an anchor point as calculated with Palantir (upper right panel). Pseudotemporal expression dynamics of T and Sox2 along the neural and somitic trajectories and their calculated branching probabilities (bottom right panel). (**F**) Time resolved scatter plots showing T/Sox2 expression in all single cells attributed to the NMP cluster. Dot size scales with differentiation potential. Dot color indicates the relative ratio of Sox1 and Tbx6 expression. Notably, Tbx6-expressing cells retain a high differentiation potential, whereas $Sox1^{+}$ cells are almost all committed.

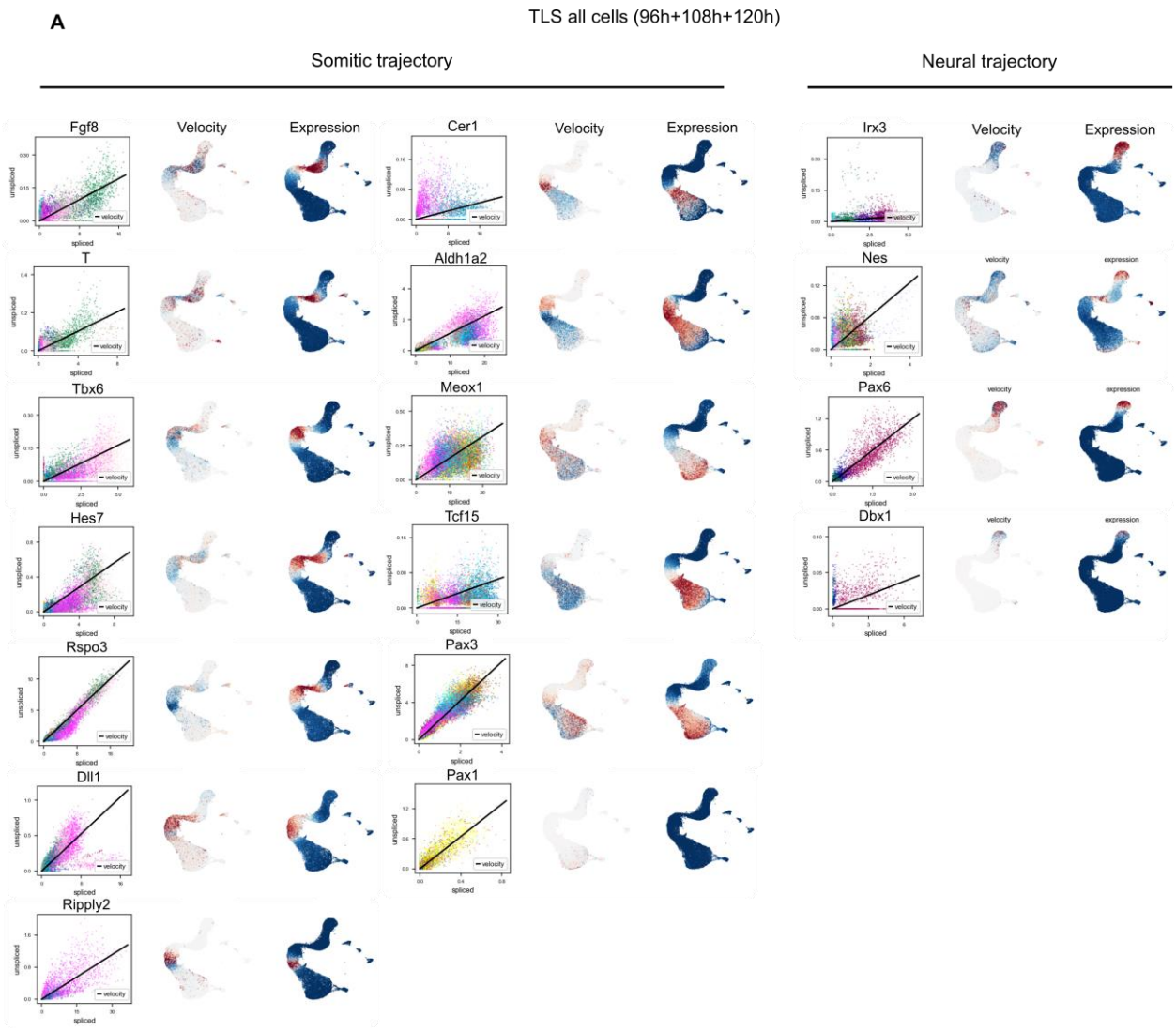

Fig. S11. RNA Velocity of *in vivo* cell type marker genes in TLS

(**A**) Velocity and expression of *in vivo* marker genes of the somitic and neural trajectory projected on the pooled (96+108+120h TLS) UMAP. Cells are coloured following the cluster colours in **Fig. 3A**.

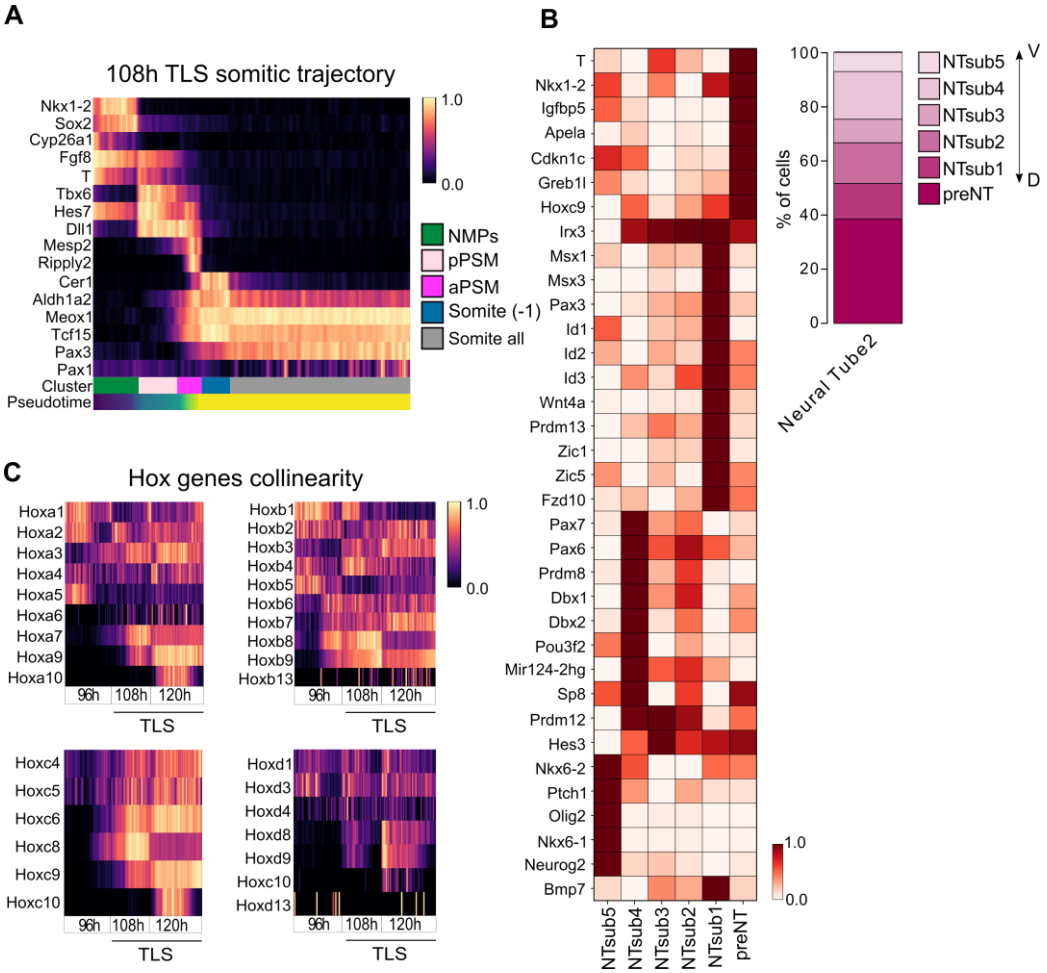

Fig. S12. Spatiotemporal progression of gene expression in TLS

(**A**) Heatmap with scaled expression of genes involved in somitogenesis in 4914 cells from 108h TLS rooted in NMPs and ordered by pseudotime. aPSM, anterior PSM, pPSM, posterior PSM. (**B**) Heatmap with scaled expression for neural lineage marker genes of each cluster that confer progenitor identity along the DV-axis *in vivo* (left panel) and stacked bar plot showing subcluster distributions. D, Dorsal; V, Ventral. (**C**) Heatmap with scaled expression of Hox genes in NMPs and their direct descendants (Neural Tube1, pPSM), ordered by pseudotime after rooting in 96h.

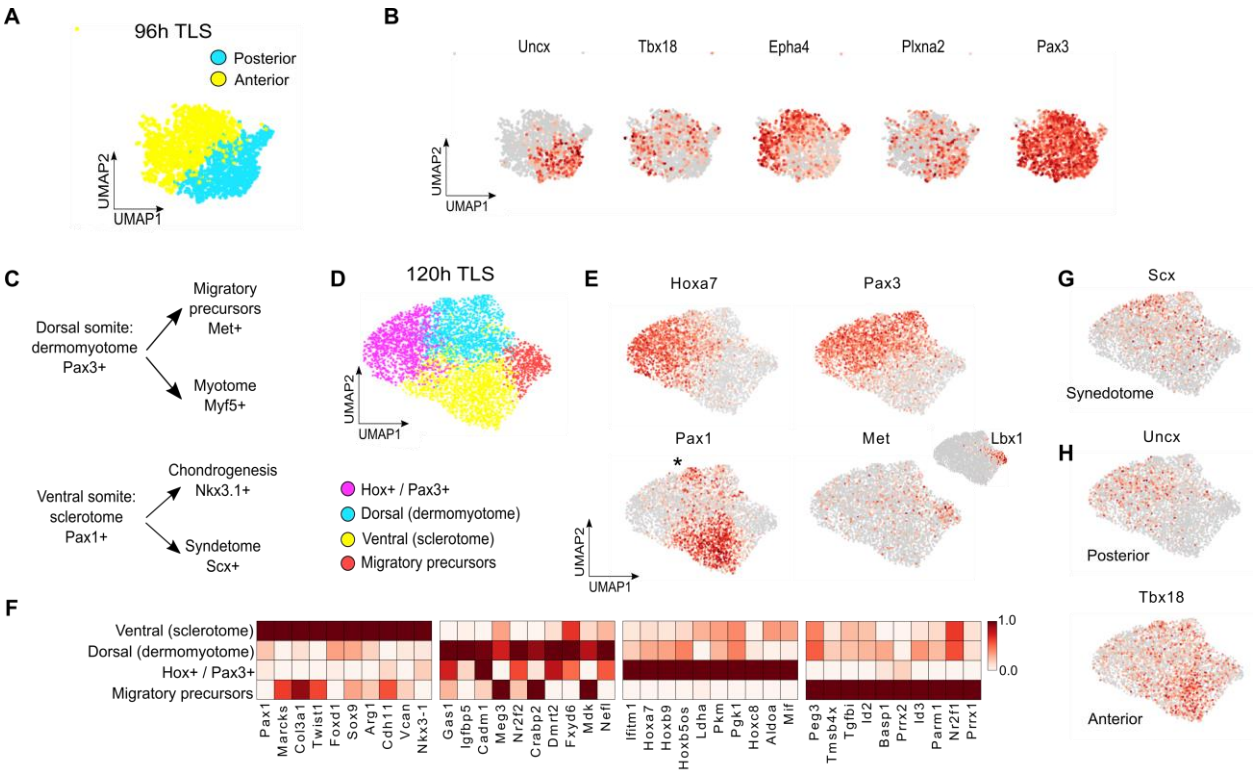

Fig. S13. Subclustering of TLS somitic cells

(**A**) UMAP representation of reclustering of 96h TLS somitic cells. (**B**) UMAP coloured by expression of indicated *in vivo* marker genes for anterior and posterior somite compartments. (**C**) *In vivo*, the dorsal somite compartment is marked by Pax3 expression and gives rise to dermamyotome and subsequently migratory muscle precursors and myotome. The ventral somite compartment is marked by Pax1 and gives rise to the sclerotome, which subsequently differentiates into chondrocytes and syndetome. (**D**) UMAP representation of reclustering of 120h TLS somitic cells. (**E**) UMAP colored by expression of indicated marker genes for dorsal and ventral somite compartments. Asterisk refers to $Pax1^{+}/Scx^{+}$ putative syndetome population. (**F**) Heatmap with scaled expression for top 10 marker genes of each cluster. (**G**) UMAP coloured by expression of *in vivo* syndetome marker Scx. Note that the Scx domain partially overlaps with a separate cluster of $Pax1^{+}$ cells (indicated with an asterisk in (**E**)). (**H**) UMAP coloured by expression of posterior (upper panel) and anterior (bottom panel) markers Uncx and Tbx18.

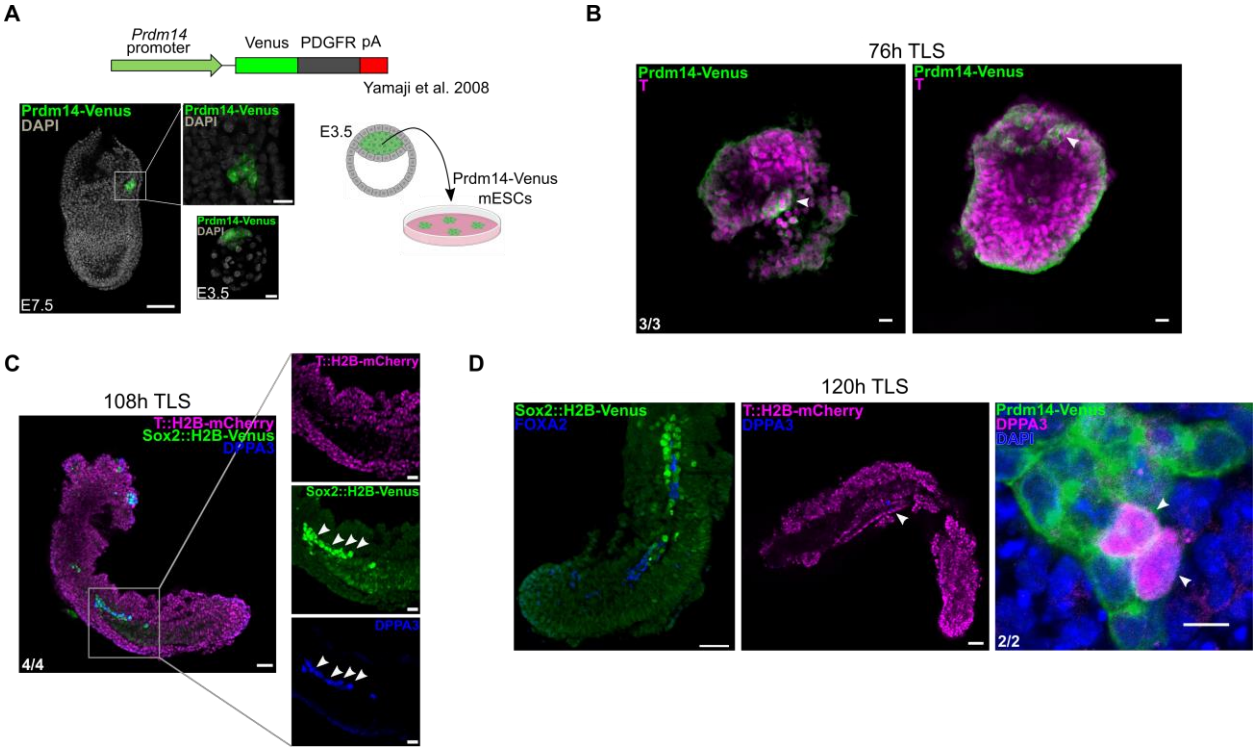

Fig. S14. Primordial Germ Cell Like Cells (PGCLCs) in TLS

(**A**) Prdm14-Venus reporter mESC line derivation. Schematic of the reporter construct (upper panel). Confocal section of an E7.5 mouse embryo and magnification of the Prmd14-Venus^{+} PGCs and Prdm14-Venus reporter expression in the E3.5 mouse blastocyst (left panel) from which mESCs were derived (right panel). PDGFR, Platelet-derived growth factor receptor; pA, polyA. Scale bar 200µm, 20µm for magnifications. (**B**) Confocal sections of 76h aggregates showing co-expression of T and Prdm14^{VE}. White arrowheads indicate T/Prdm14^{VE} double positive cells. Scale bar 20µm. (**C**) Confocal section of 108h TLS showing PGCLCs that co-express Sox2^{VE} and DPPA3. Scale bars 50µm, 25µm for magnifications. (**D**) Confocal section of 120h TLS stained for FOXA2 and Sox2^{VE} (left). Note how Sox2$^{VE\text{-}high}$ cells contact the FOXA2^{+} cells in 120h TLS. Scale bars 50µm. Confocal section of 120h TLS stained for DPPA3 showing the whole structure from **Fig. 4E** (middle). Scale bars 50µm. Confocal section of 120h TLS stained for DPPA3 and showing PGCLCs co-expressing Prdm14^{VE} and DPPA3 (right). Scale bar 10µm. In (**B-D**) the proportion of structures exhibiting the reported staining pattern is shown.

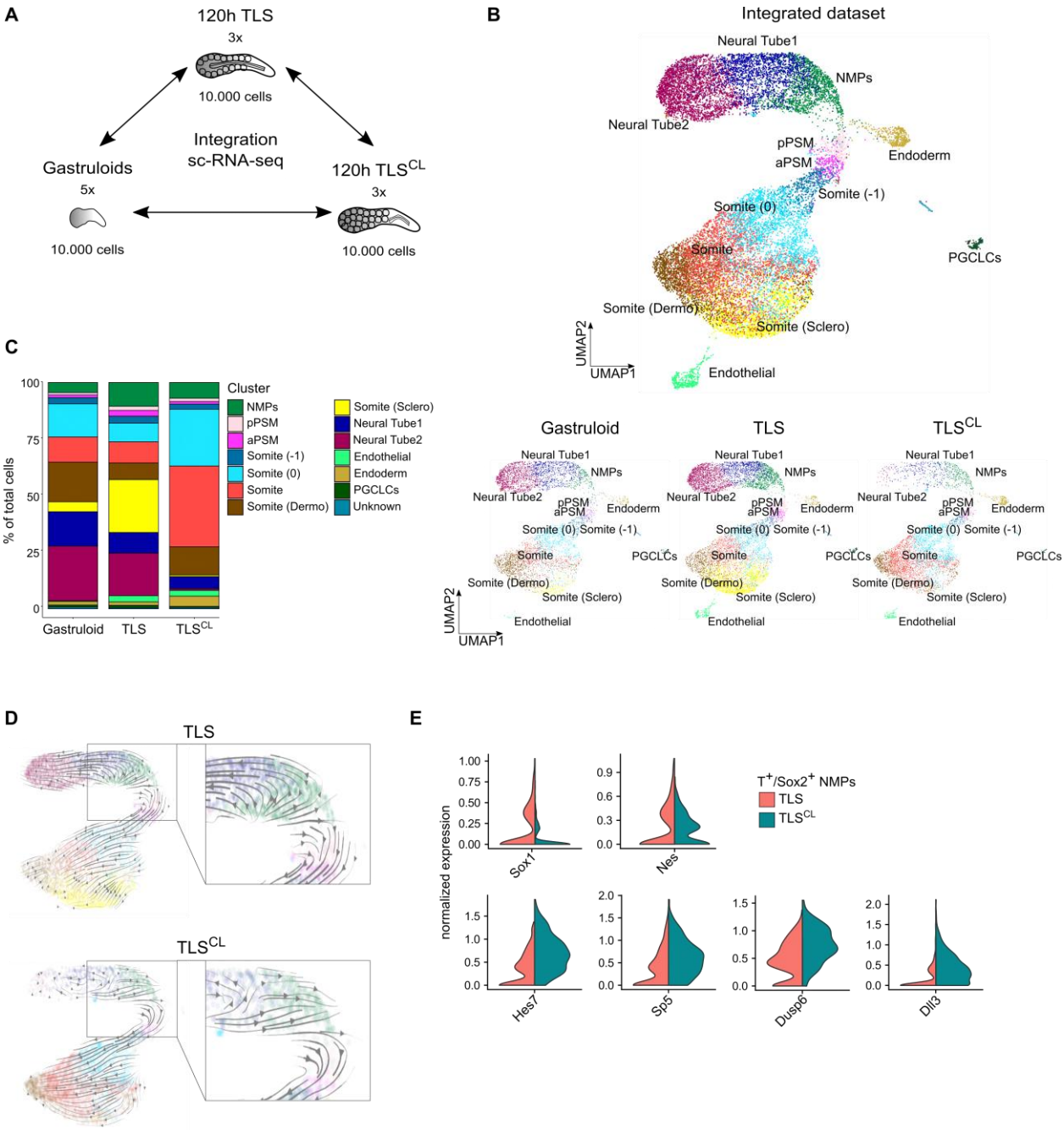

Fig. S15. Comparative sc-RNA-seq of TLS, TLS^{CL} and gastruloids

(A) Experimental set-up for comparative sc-RNA-seq. (**B**) Integrated and split UMAP of cells from gastruloids, TLS and TLS^{CL} showing differences in cluster proportions for the three protocols. Cells were assigned based on the 120h TLS clustering (see Supplemental Information for details). (**C**) Stacked bar plots showing the differential proportions of the cell states in the three protocols (gastruloids, TLS and TLS^{CL}). Note i) the drastically increased proportion of endothelial cells in TLS and TLS^{CL} compared to gastruloids; ii) the increased proportion of sclerotomal cells in TLS compared to gastruloids; iii) the almost absence of sclerotomal and more mature neural cells in TLS^{CL}. (**D**) Integrated UMAP colored by identified clusters as in (**A**) with trajectories inferred from RNA Velocity for putative NMPs and

descendants in 120h TLS (top panel) and 120h TLS^{CL} (bottom panel). Note that in TLS^{CL} putative NMPs are predicted to exclusively give rise to somitic descendants, whereas in TLS arrows point into the direction of both neural and somitic clusters indicative of true bipotency. Moreover, note that in TLS^{CL} even early neural cells appear to revert to a somitic fate. (**E**) Split violin plots showing expression of neural marker genes (top panels) and PSM marker genes (bottom panels) in T/Sox2 co-expressing putative NMPs in 120h TLS and 120h TLS^{CL}.

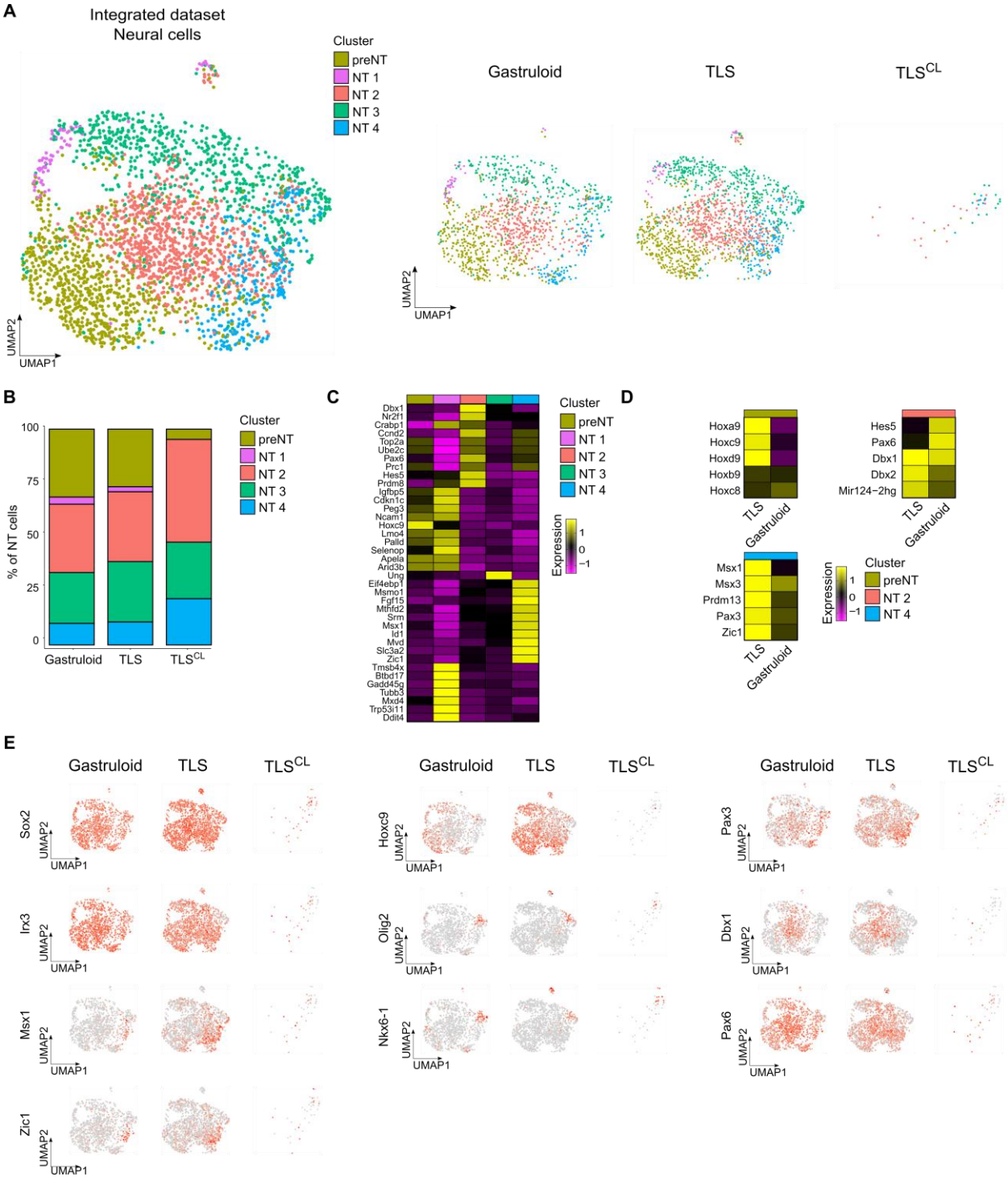

Fig. S16. Comparative sc-RNA-seq of neural cells in TLS, TLS^{CL} and gastruloids

(**A**) Integrated and split UMAP of most mature neural cells extracted from gastruloids, TLS and TLS^{CL} (main cluster "Neural Tube 2") colored by the five sub-clusters identified. Note the drastic reduction of this cell-type in TLS^{CL}. (**B**) Stacked bar plot showing the proportions of the five clusters in gastruloids, TLS and TLS^{CL}. (**C**) Heatmap of cluster averaged normalized expression of the key marker genes for each of the five clusters. (**D**) Heatmap showing pairwise comparison between TLS and gastruloids of cluster averaged normalized expression of selected genes for clusters preNT, NT2 and NT4. (**E**) UMAPs colored by the expression of the indicated marker genes of NT subpopulations and split by condition (gastruloids, TLS and TLS^{CL}). Note that, although gastruloids and TLS are overall very similar, some

important differences are revealed: i) preNT cells have a more advanced Hox code in TLS as compared to gastruloids; ii) the expression of dorsal neural tube markers (including Msx1, Msx3, Zic1, Pax3) is increased in TLS compared to gastruloids *(13)*.

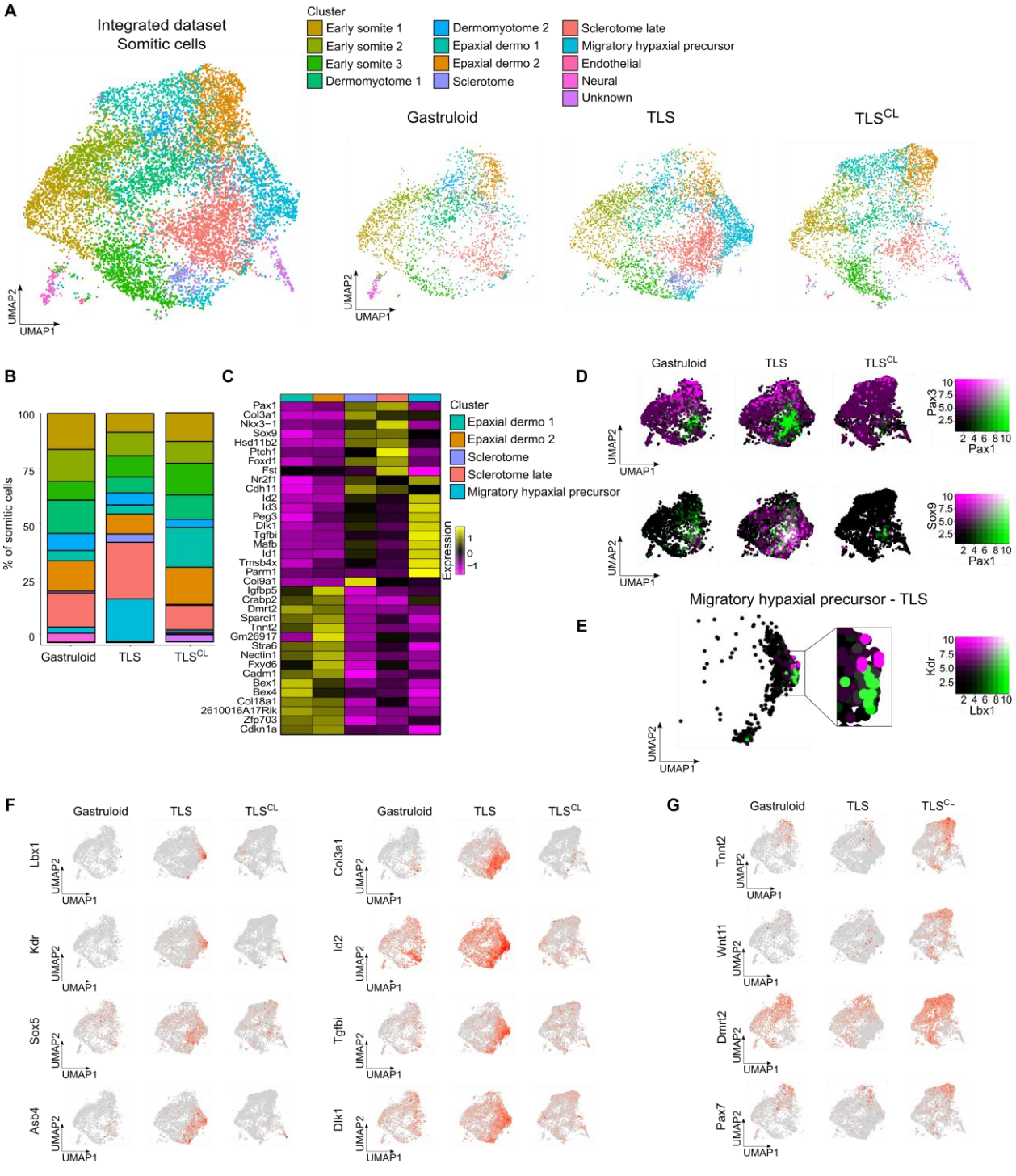

Fig. S17. Comparison of somitic cell types in TLS, TLS^{CL} and gastruloids by sc-RNA-seq

(**A**) Integrated and split UMAP of somitic cells extracted from gastruloids, TLS and TLS^{CL} colored by 13 subclusters identified. (**B**) Stacked bar plots showing the proportions of the 13 clusters in gastruloids, TLS and TLS^{CL}. Note that putative hypaxial migratory limb muscle precursors, marked by Lbx1 (see panel **C**,**E,F**) are exclusively detected in TLS *(26)*. Moreover, sclerotomal cells are enriched in TLS compared to gastruloids (see also panel **D**). Also note that WNT activation in combination with BMP inhibition in TLS^{CL} increased the fraction of Dmrt2/Wnt11/Tnnt2-expressing cells, which likely represent epaxial dermomyotome, at the expense of hypaxial dermomyotome *(14, 42)*. (**C**) Heatmap of

cluster averaged normalized expression of the key marker genes for condition-specific clusters. Note that marker genes of the migratory hypaxial precursor cluster comprise many BMP and TGFB-induced genes (Id1, Id2, Id3, Tgfbi). (**D**) Blended split UMAPs showing the expression of Pax3 and Pax1 (upper panel) or Pax1 and Sox9 (lower panel) in gastruloids, TLS and TLS^{CL}. Note that $Pax1^{+}/Sox9^{+}$ putative chondrocyte precursors are more frequent in TLS compared to gastruloids. (**E**) Blended UMAP showing two unique populations of cells expressing Kdr or Lbx1 specifically in the migratory hypaxial precursor cluster in TLS. The Kdr^{+} cells may represent dorsal somite-derived putative endothelial progenitors that derive from the same progenitor as $Lbx1^{+}$ cells (*43*). (**F**) Split UMAPs colored by the expression of the indicated cluster-specific genes in gastruloids, TLS and TLS^{CL}. (**G**) Split UMAPs colored by the expression of the indicated putative epaxial dermomyotome marker genes in gastruloids, TLS and TLS^{CL}.

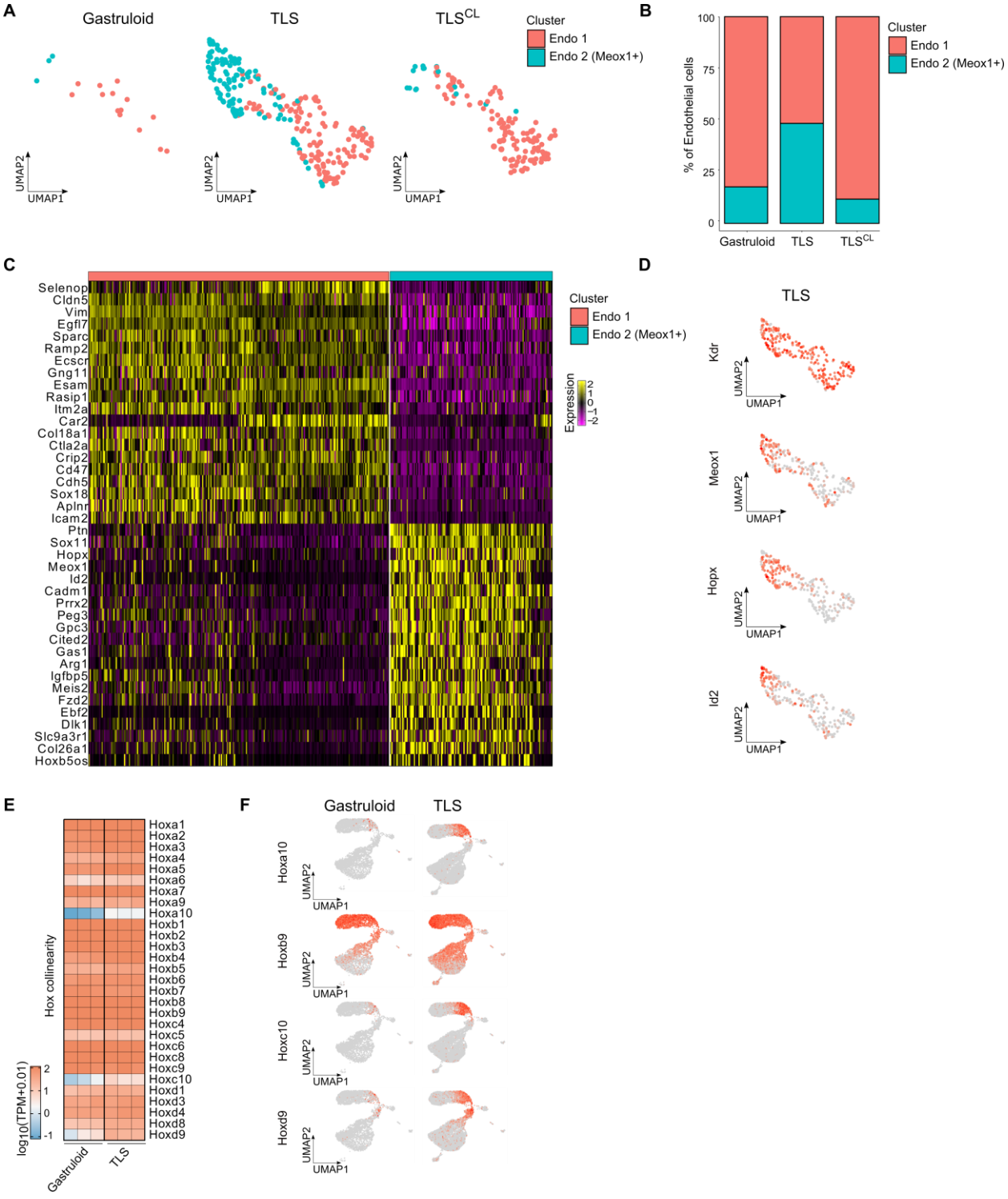

Fig. S18. Comparison of endothelial cell types and Hox progression in TLS, TLS^{CL} and gastruloids

(**A**) Split UMAP of endothelial cells extracted from gastruloids, TLS and TLS^{CL} colored by two identified endothelial sub-clusters. (**B**) Stacked bar plots showing the different proportions of the two clusters in gastruloids, TLS and TLS^{CL}. (**C**) Heatmap of normalized expression of top 20 marker genes for the two clusters. (**D**) UMAPs of endothelial cells from TLS colored by the expression of the indicated genes. Note that, whereas pan-endothelial marker Kdr is expressed in both clusters, Meox1, Hopx and

Id2 are specifically expressed in Endo 2 (Meox1^{+}) cluster. *In vivo*, this gene signature is observed in endotome cells that eventually induce haematopoietic stem cells (*44*). (**E**) Heatmap showing $\log_{10}$TPM values for the indicated Hox genes in gastruloids and TLS in bulk-RNA-seq data. Each column represents an independent biological replicate. (**F**) UMAPs colored by the expression of the indicated late Hox genes in gastruloids and TLS.

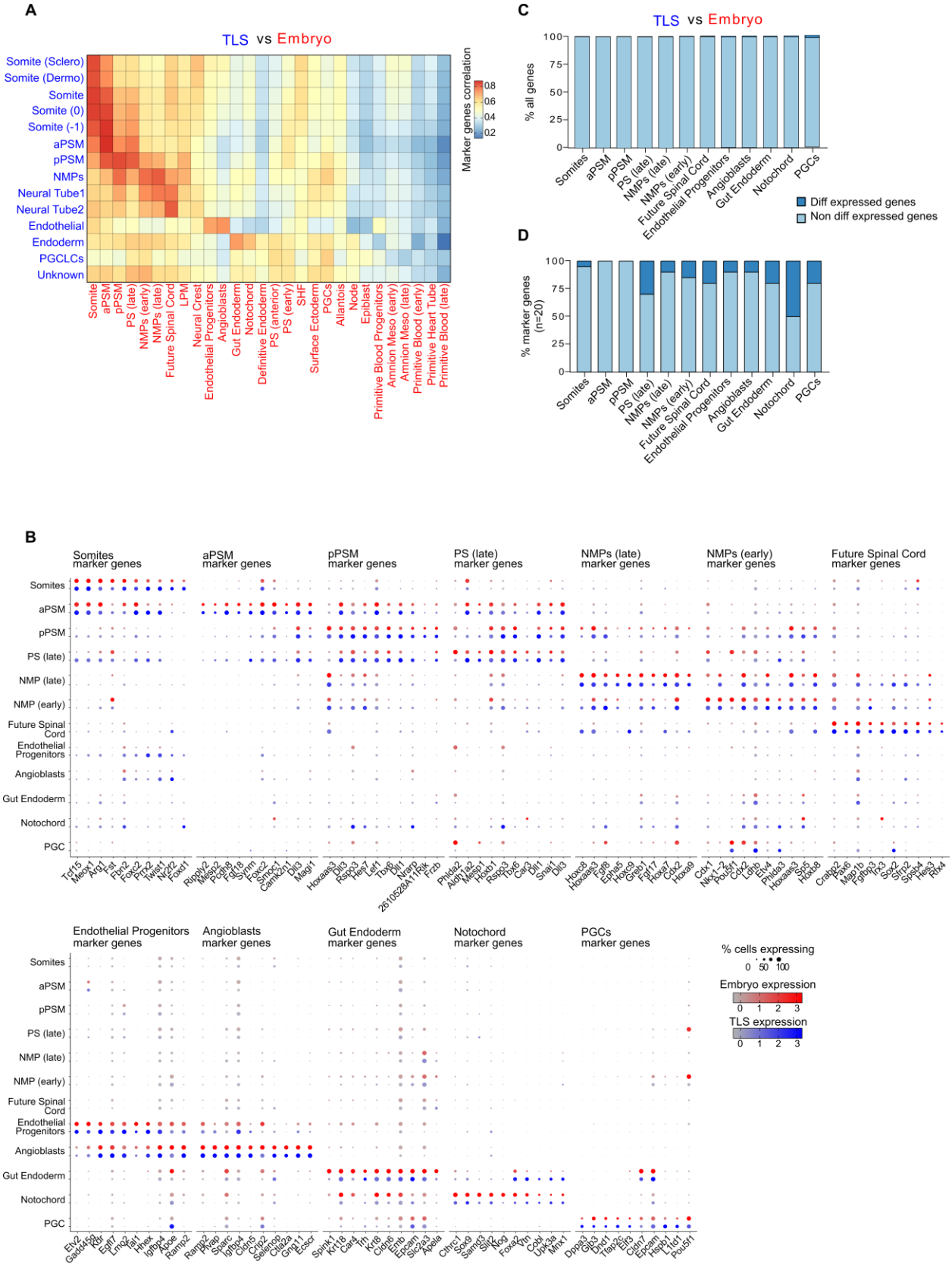
A
TLS vs Embryo
Somite (Sclero)
Somite (Dermo)
Somite
Somite (0)
Somite (-1)
aPSM
pPSM
NMPs
Neural Tube1
Neural Tube2
Endothelial
Endoderm
PGCLCs
Unknown
Somite
aPSM
pPSM
PS (late)
NMPs (early)
NMPs (late)
Future Spinal Cord
LPM
Neural Crest
Endothelial Progenitors
Angioblasts
Gut Endoderm
Notochord
Definitive Endoderm
PS (anterior)
PS (early)
SHF
Surface Ectoderm
PGCs
Allantois
Node
Epiblast
Primitive Blood Progenitors
Amnion Meso (early)
Amnion Meso (late)
Primitive Blood (early)
Primitive Heart Tube
Primitive Blood (late)
Marker genes correlation
0.8
0.6
0.4
0.2
C
TLS vs Embryo
% all genes
100
75
50
25
0
Somites
aPSM
pPSM
PS (late)
NMPs (late)
NMPs (early)
Future Spinal Cord
Endothelial Progenitors
Angioblasts
Gut Endoderm
Notochord
PGCs
Diff expressed genes
Non diff expressed genes
D
% marker genes (n=20)
100
75
50
25
0
Somites
aPSM
pPSM
PS (late)
NMPs (late)
NMPs (early)
Future Spinal Cord
Endothelial Progenitors
Angioblasts
Gut Endoderm
Notochord
PGCs
B
Somites marker genes
aPSM marker genes
pPSM marker genes
PS (late) marker genes
NMPs (late) marker genes
NMPs (early) marker genes
Future Spinal Cord marker genes
Somites
aPSM
pPSM
PS (late)
NMP (late)
NMP (early)
Future Spinal Cord
Endothelial Progenitors
Angioblasts
Gut Endoderm
Notochord
PGC
Tcf15 Meox1 Arg1 Fst Fbn2 Foxc2 Prrx2 Twist1 Nr2f2 Foxd1
Ripply2 Meox2 Pcdh8 Fgf18 Sytm Foxc2 Smoc1 Camk2n1 Dll3 Magi1
Hoxaas3 Dll3 Rspo3 Hes7 Lef1 Tbx6 Dll1 Nrarp 2610528A11Rik Frzb
Phlda2 Aldh1a2 Meis1 Hoxb1 Rspo3 Tbx6 Car3 Dll1 Snai1 Dll3
Hoxc8 Hoxaas3 Fgf8 Epha5 Hoxc9 Grem1 Fgf17 Hoxa7 Cdx2 Hoxa9
Cdx1 Nkx1-2 Pou3f1 Cdx2 Lin28b Etv4 Phlda3 Hoxaas3 Spic5 Hoxb8
Crabp2 Pax6 Map1b Fgfbp3 Tfr3 Sox2 Sfrp2 Sprr1a Hes3 Rtl4
Endothelial Progenitors marker genes
Angioblasts marker genes
Gut Endoderm marker genes
Notochord marker genes
PGCs marker genes
Somites
aPSM
pPSM
PS (late)
NMP (late)
NMP (early)
Future Spinal Cord
Endothelial Progenitors
Angioblasts
Gut Endoderm
Notochord
PGC
Etv2 Gadd45g Kdr Egfl7 Lmo2 Tal1 Hhex Igfbp4 Apoe Ramp2
Ramp2 Prrx1 Sparc Igfbp4 Cldn5 Crip2 Selenop Cldn5a Gng11 Escr
Spint1 Krt18 Car4 Tfh Krt8 Cldn6 Emb Epcam Slc2a3 Apela
Cthrc1 Sox9 Samd3 Shh Nog Foxa2 Vtn Cdh1 Upk3a Mnx1
Dppa3 Gli3 Dnd1 Tfap2c Eif3l Cldn7 Epcam Hspb11 Tdrd1 Pou5f1
% cells expressing
0 50 100
Embryo expression
0 1 2 3
TLS expression
0 1 2 3

Fig. S19. Comparison of TLS cell states with post-occipital cell states of the mouse embryo

(**A**) Heatmap showing correlation of cell states identified in TLS (blue font) and cell states identified in the post-occipital E7.5 and E8.5 embryo (red font) based on top 20 marker genes for *in vivo* clusters. (**B**) Dot plots showing the ten most conserved cell state markers for indicated cell states between TLS and the embryo. (**C**) Stacked bar plots with percentage of expressed genes (average expression in cluster > 0) that are differentially expressed in TLS as compared to the embryo (min.diff.pct > 0.25 and log_2FC > 1). (**D**) Stacked bar plots with percentage of top 20 *in vivo* cluster marker genes that are differentially expressed in TLS as compared to the embryo (min.diff.pct > 0.25 and log_2FC > 1).

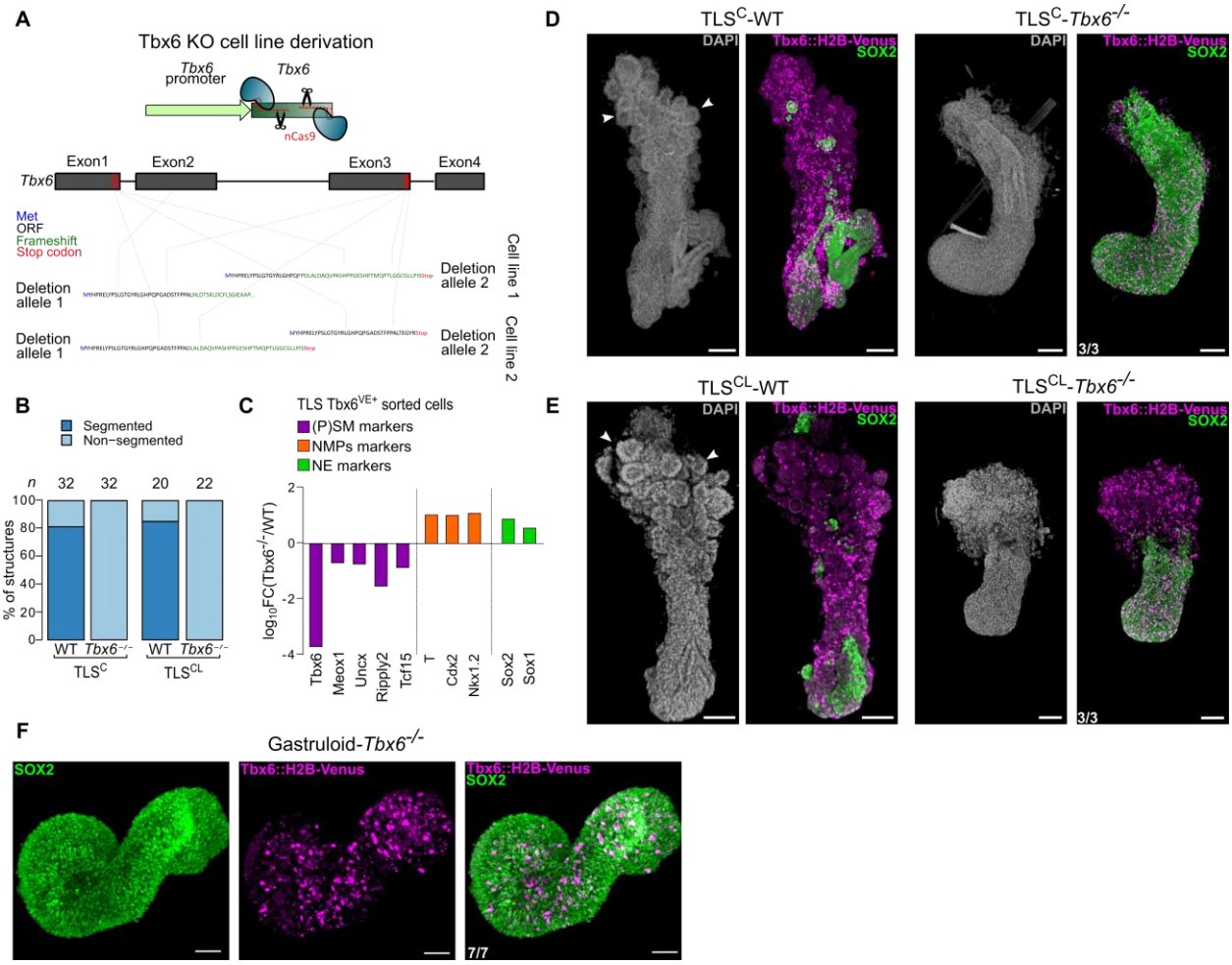

Fig. S20. Loss of segmentation and molecular phenotype in Tbx6$^{-/-}$ TLS models

(**A**) Schematic of two independent *Tbx6::H2B-Venus* (Tbx6^{VE}); *Tbx6*$^{-/-}$ mESC lines with Sanger sequencing validation. (**B**). Segmentation phenotype of TLS^{C}-WT, TLS^{C}-*Tbx6*$^{-/-}$, TLS^{CL}-WT and TLS^{CL}-*Tbx6*$^{-/-}$. Data were obtained in 3 different experiments performed with 2 independent mESC lines of each genotype. (**C**) qRT-PCR showing expression of (P)SM, NMPs and NE genes in FACS-purified Tbx6^{VE+} cells from TLS-WT and TLS-*Tbx6*$^{-/-}$. Log_{10}FC was calculated from the fold change ratio between TLS-*Tbx6*$^{-/-}$ and TLS-WT. Bars represent the mean of two biological replicates. (**D**) 3D maximum intensity projections of TLS^{C}-WT and TLS^{C}-*Tbx6*$^{-/-}$ immunostained for Tbx6^{VE} (magenta) and SOX2 (green). (**E**) 3D maximum intensity projections of TLS^{CL}-WT and TLS^{CL}-*Tbx6*$^{-/-}$ immunostained for Tbx6^{VE} (magenta) and SOX2 (green). (**F**) 3D maximum intensity projections of *Tbx6*$^{-/-}$ gastruloids- immunostained for Tbx6^{VE} (magenta) and SOX2 (green). The proportion of structures exhibiting the reported staining pattern is shown. Scale bars 100µm.

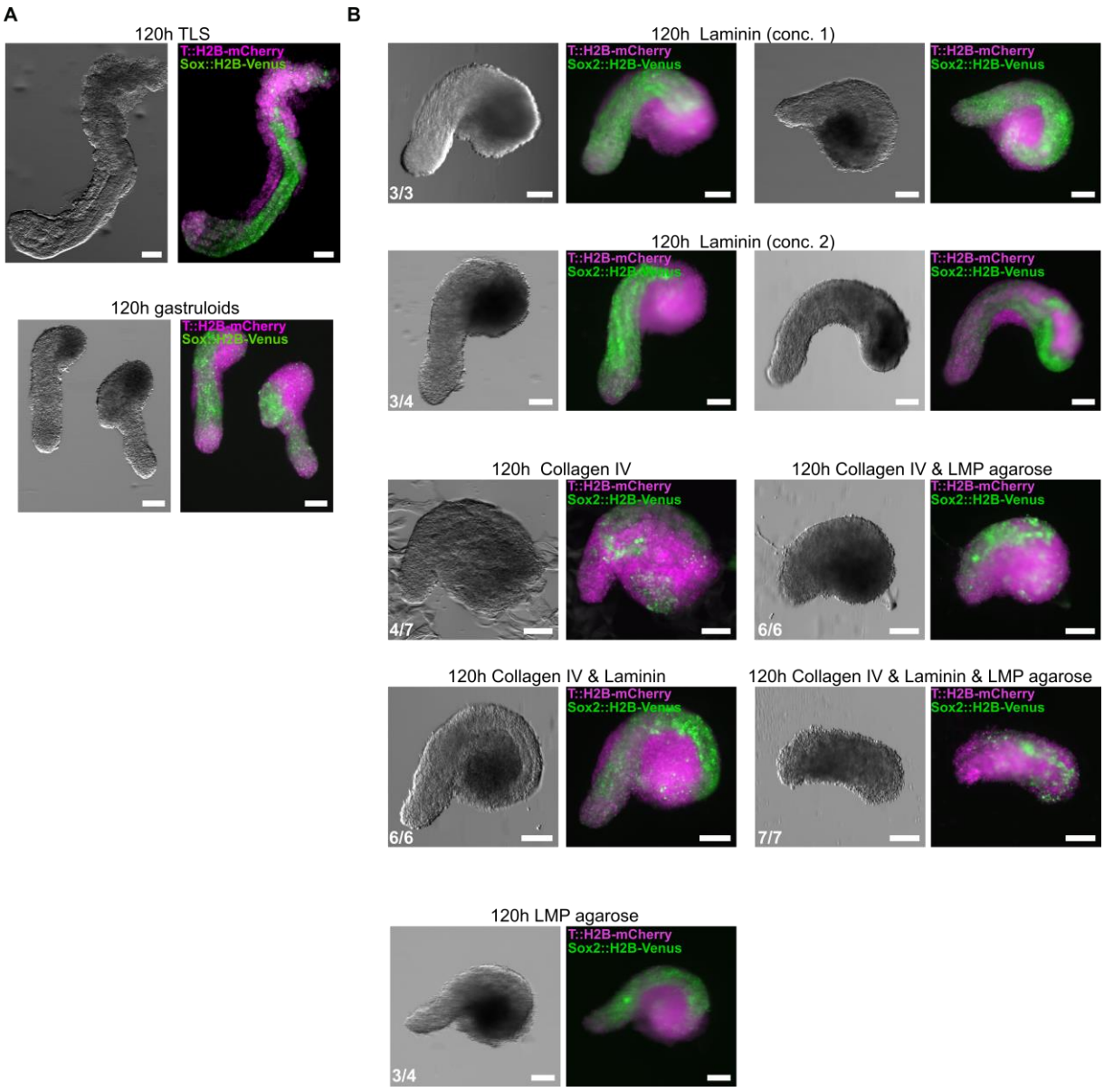

Fig. S21. Matrigel components screen to mimic TLS architecture in gastruloids

(A) Wide-field fluorescence imaging showing general tissue morphology and organization of neural ($Sox2^{VE+}$) and mesodermal (T^{mCH+}) domains in TLS and gastruloids. TLS (upper panel) shows the formation of a neural tube and somites, while gastruloids (lower panel) do not display any clear embryo-like architecture. (**B**) Adding the ECM components Laminin and Collagen IV (Col) to 96h gastruloids, alone or in combination, fails to induce embryo-like morphogenesis in gastruloids in the absence of matrigel. From top to bottom: 0.06mg/ml Laminin (conc.1), 0.12mg/ml Laminin (conc.2), 0.12 mg/ml Collagen IV, 0.12 mg/ml Collagen IV + LMP agarose, 0.048 mg/ml Collagen IV + 0.036mg/ml Laminin, 0.048 mg/ml Collagen IV + 0.036mg/ml Laminin + LMP agarose, LMP agarose. LMP agarose was prepared as a 1.5% stock solution, which was further diluted 1:20 in PBS. LMP, low melting point. For each condition the proportion of structures exhibiting the shown morphology and $T^{mCH}/Sox2^{VE}$ distribution is shown. Scale bars 100µm.

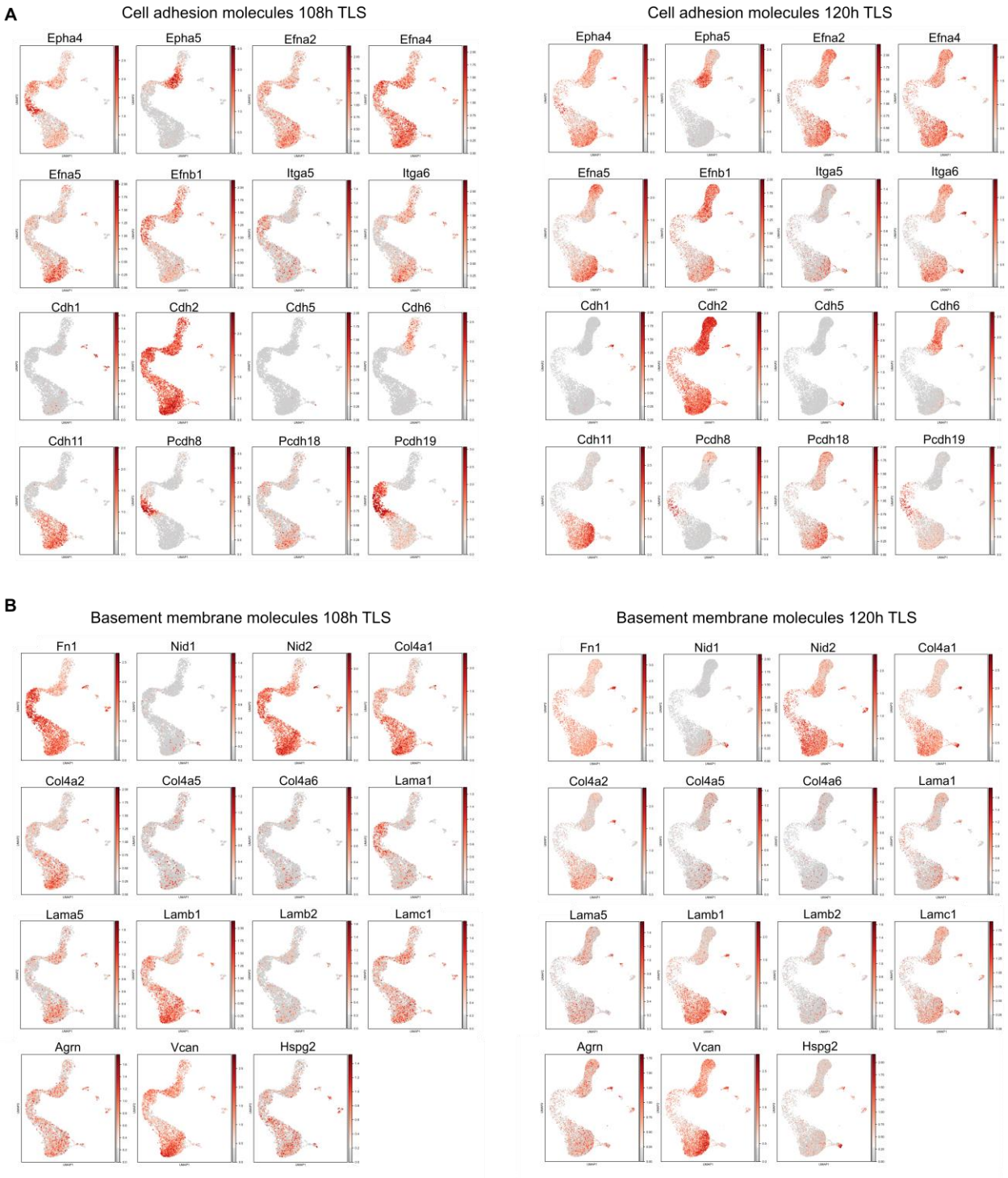

Fig. S22. Cell state specific expression of matrisome components and cell adhesion molecules in trunk-like-structures

(**A**) UMAP of 108h and 120h TLS colored by expression of indicated genes encoding cell adhesion molecules. (**B**) UMAP of 108h and120h TLS colored by expression of indicated genes encoding basement membrane components.

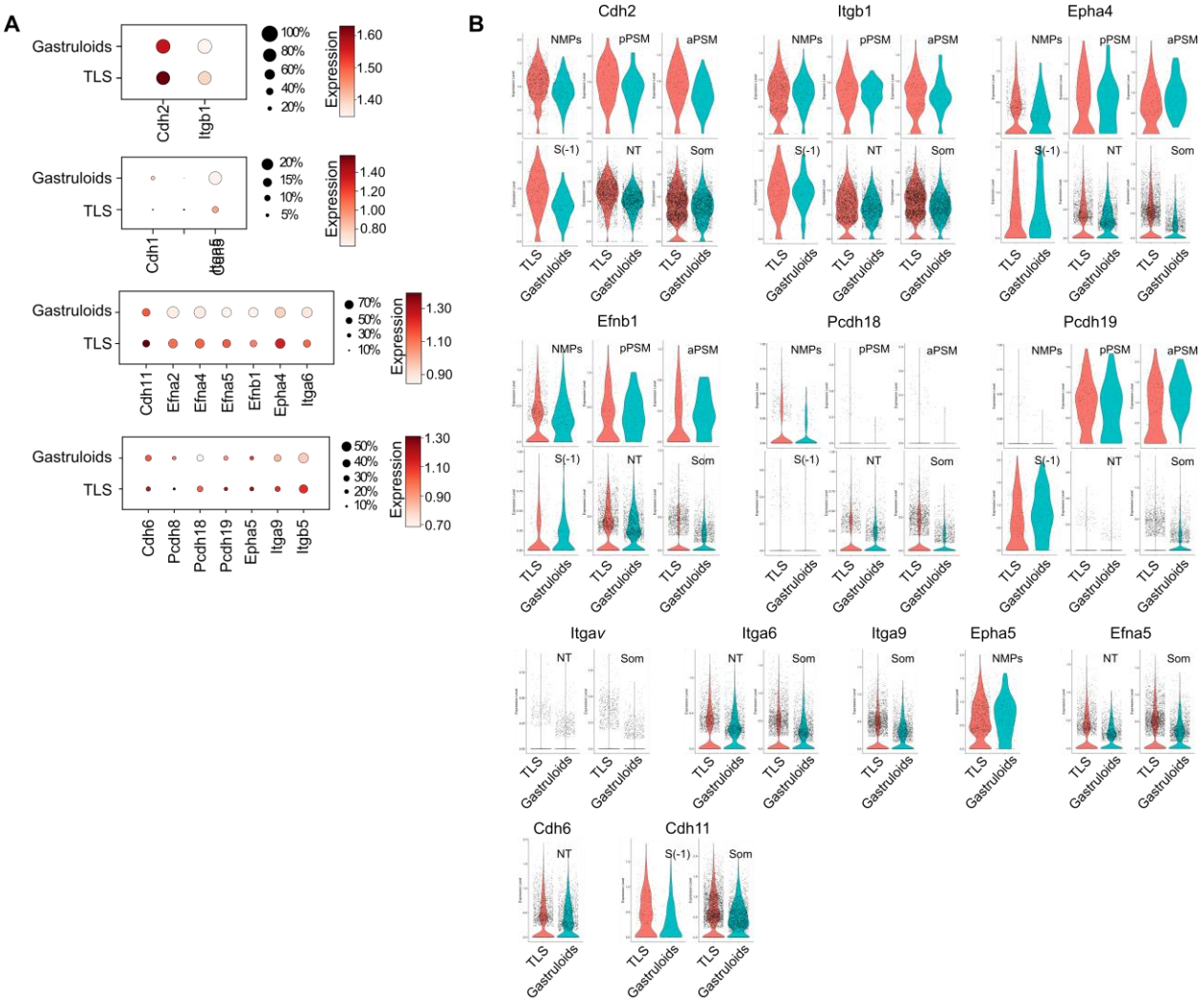

Fig. S23. Expression of cell adhesion molecules in different cell-types in TLS, TLS^{CL} and gastruloids

(A) Dotplot showing the percentage of cells expressing indicated key cell adhesion molecules and their average expression level in these cells. (**B**) Violin plots showing the expression of indicated genes in single cells of different cell states in trunk-like-structures and gastruloids. NMPs, neuromesodermal progenitors; pPSM, posterior pre-somitic mesoderm; aPSM, anterior pre-somitic mesoderm; S (-1), Somite (-1); NT, Neural Tube; Som, Somite.

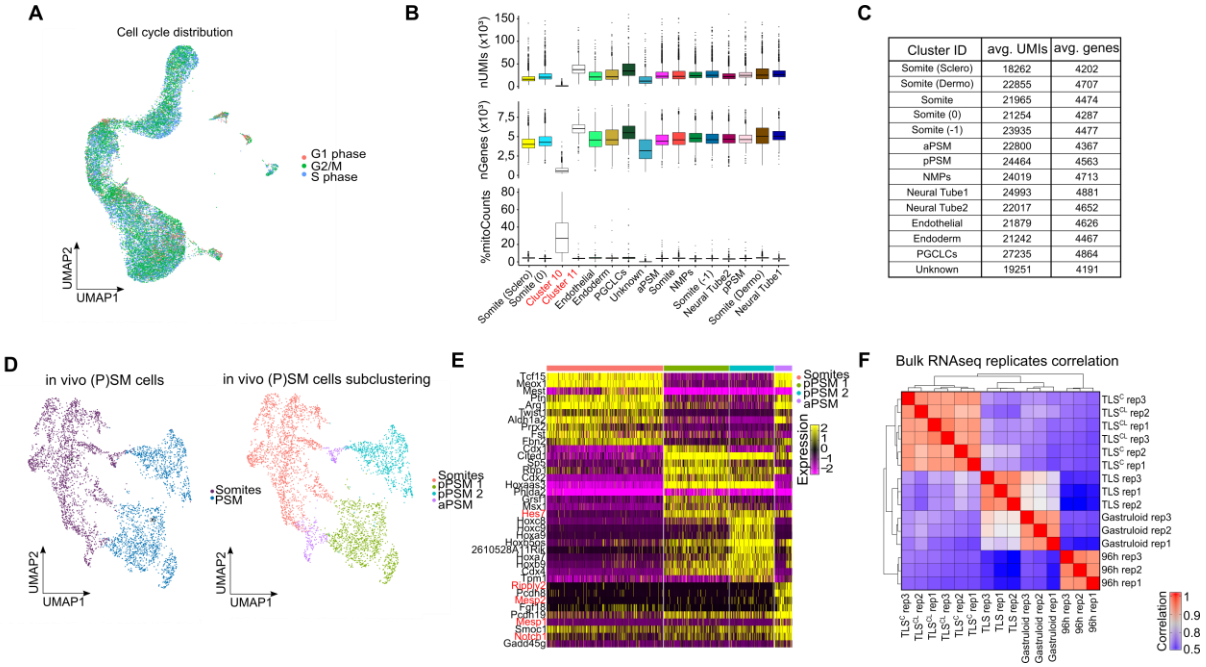

Fig. S24. Quality control

(**A**) UMAP of 96h, 108h and 120h TLS colored by predicted cell cycle phase, showing that cells do not cluster by cell cycle phase. To this end, first for each gene the cell cycle phase was estimated and subsequently this source of heterogeneity was regressed out so that the scaled data could be used for UMAP calculation without a cell cycle phase bias. (**B**) Statistics for each of the clusters before correction for number of UMIs, genes and mitochondrial fraction. Cluster 10 is a clear outlier showing a strong bias towards an extremely high mitochondrial fraction and cluster 11 was marked by a high percentage of cells with an unusual high number of UMIs and detected genes (indicative of doublets), such that after removal of these cells, only a small fraction of cells was left. Both clusters (10 and 11) were therefore excluded from further analysis. Boxes indicate interquartile range (IQR). Whiskers extends to 1.5xIQR from the hinge. Dots indicate outliers. Central line represent the median. (**C**) Table displaying the average number of UMIs and genes detected for all cells separated by cluster. The average number of UMIs and genes is homogeneous between clusters and does not show any outliers. (**D**) UMAP of *in vivo* cells from Somite and PSM cell states. UMAP coloured by initial cell state (left) compared to more stringent cluster assignment (right). The stringent cluster assignment results in four sub-clusters, that split “PSM" cluster into two equally sized clusters, while the “Somites" cluster splits into one larger cluster and a second group of cells that are located at the transition of the previously identified Somites and PSM cluster. (**E**) Heatmap of normalized expression of marker genes for the newly assigned four clusters coming from *in vivo* “PSM“ and “Somites” cells. The two PSM clusters both show strong expression of the same subset of marker genes, however, the new small subset bridging the PSM and Somite cells has a clear anterior PSM (aPSM) signature. (**F**) Correlation analysis of bulk RNA-seq, showing very high correlation between replicates of the same treatment. Only genes covered by all samples and expressed in at least one sample were considered (n = 29,963) and the pairwise Pearson correlation coefficient was calculated on the $\log_2$(TPM+1) values.

Movies S1-S2. (separate files)

3D rendering of lightsheet imaging of TLSs generated with T::H2B-mCherry/Sox2::H2B-Venus double reporter cell line. For Movie S1, sample was stained with a mCherry and Venus antibodies. Movie S2 shows reporter signals without antibody staining.

Movies S3-S5. (separate files)

Wide-field live imaging of TLS after embedding in matrigel 96h post-aggregation showing elongation, somite segmentation, and neural tube formation. **Movie S3:** 3 TLSs; **Movie S4:** 2 TLSs; **Movie S5:** 2 TLSs; Similar results were obtained in n=5 independent experiments.

Movies S6-S7. (separate files)

Wide-field live imaging of TLS^{C} after embedding in matrigel 96h post-aggregation showing elongation, somite segmentation and organization in bunches of grapes, and limited neural tube formation. **Movie S6:** 2 TLS^{C}; **Movie S7:** 2 TLS^{C}. Similar results were obtained in n=4 independent experiments.

Movies S8-S9. (separate files)

Wide-field live imaging of TLS^{CL} after embedding in matrigel 96h post-aggregation showing elongation, somite segmentation and organization in bunches of grapes, and limited neural tube formation. **Movie S6:** 2 TLS^{CL}; **Movie S7:** 2 TLS^{CL}. Similar results were obtained in n=4 independent experiments.

Movie S10. (separate file)

3D rendering of lightsheet imaging of TLS stained with Phalloidin, FN1 and DAPI.

Data S1. (separate file)

Lists of genes used to generate plots from **Fig. S1I**, **Fig. 2B,F**, **Fig. S6C,D,E** and **Fig. S7** and corresponding statistical analysis

Data S2. (separate file)

Significantly enriched processes in TLS versus Gastruloids as revealed by Gene Set Enrichment Analysis

Data S3. (separate file)

Overview of buffers, culture media, kits, and antibodies used in this study

Data S4. (separate file)

Sequences of oligos, gRNAs and *in situ hybridization* probes used in this study

Data S5. (separate file)

Morphometric quantifications of TLSs full structures and substructures and segmentation clock periods with corresponding statistical analysis

Data S6. (separate file)

Raw Ct values and quantification from quantitative PCR analysis

Data S7. (separate file)

Details of Nanostring probes used in this study and Nanostring normalized gene counts

Data S8. (separate file)

Marker genes for TLS clusters (single-cell RNA-Seq) & Marker genes for embryo clusters (single-cell RNA-Seq)

Data S9. (separate file)

Top 25 conserved cluster marker genes between TLS and embryo & Genes differentially expressed between TLS cell types and their corresponding *in vivo* cell type

Data S10. (separate file)

Expression values in TPM (Transcripts Per Million) for all genes (bulk RNA-Seq)

33. D. A. Stafford, L. J. Brunet, M. K. Khokha, A. N. Economides, R. M. Harland, Coopera-tive activity of noggin and gremlin 1 in axial skeleton development. Development. 138, 1005–1014 (2011).
34. X. M. Yang, K. Vogan, P. Gros, M. Park, Expression of the met receptor tyrosine kinase in muscle progenitor cells in somites and limbs is absent in Splotch mice. Development. 122, 2163–2171 (1996).
35. Y. J. Jiang et al., Notch signalling and the synchronization of the somite segmentation clock. Nature. 408, 475–479 (2000).
36. T. J. Cunningham et al., Retinoic acid activity in undifferentiated neural progenitors is suf-ficient to fulfill its role in restricting fgf8 expression for somitogenesis. PLoS One. 10, e0137894 (2015).
37. R. Morita et al., ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. Proc. Natl. Acad. Sci. USA. 112, 160–165 (2015).
38. W. W. Sugden et al., Endoglin controls blood vessel diameter through endothelial cell shape changes in response to haemodynamic cues. Nat. Cell Biol. 19, 653–665 (2017).
39. K. Xu et al., Blood vessel tubulogenesis requires Rasip1 regulation of GTPase signaling. Dev. Cell. 20, 526–539 (2011).
40. S. Kanno et al., Roles of two VEGF receptors, Flt-1 and KDR, in the signal transduction of VEGF effects in human vascular endothelial cells. Oncogene. 19, 2138–2146 (2000).
41. S. Gauvrit et al., HHEX is a transcriptional regulator of the VEGFC/FLT4/PROX1 signal-ing axis during vascular development. Nat. Commun. 9, 2704 (2018).
42. T. Sato, D. Rocancourt, L. Marques, S. Thorsteinsdóttir, M. Buckingham, A Pax3/Dmrt2/Myf5 regulatory cascade functions at the onset of myogenesis. PLoS Genet. 6, e1000897 (2010).
43. M. Lagha et al., Pax3:Foxc2 reciprocal repression in the somite modulates muscular versus vascular cell fate choice in multipotent progenitors. Dev. Cell. 17, 892–899 (2009).
44. P. D. Nguyen et al., Haematopoietic stem cell induction by somite-derived endothelial cells controlled by meox1. Nature. 512, 314–318 (2014).
45. G. S. Eakin, A.-K. Hadjantonakis, Production of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. Nat. Protoc. 1, 1145–1153 (2006).
46. M. Yamaji et al., Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat. Genet. 40, 1016–1022 (2008).
47. N. Nakagata, Cryopreservation of mouse spermatozoa and in vitro fertilization. Methods Mol. Biol. 693, 57–73 (2011).
48. S. H. L. George et al., Developmental and adult phenotyping directly from mutant embry-onic stem cells. Proc. Natl. Acad. Sci. USA. 104, 4455–4460 (2007).
49. J. P. Muyrers, Y. Zhang, G. Testa, A. F. Stewart, Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. 27, 1555–1557 (1999).
50. F. A. Ran et al., Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).

51. L. Beccari et al., Generating Gastruloids from Mouse Embryonic Stem Cells. Protoc exch (2018), doi:10.1038/protex.2018.094.
52. N. Otsu, A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst. Man. Cybern. 9, 62–66 (1979).
53. J. Serra, Image Analysis and Mathematical Morphology (Academic Press, Inc., USA, 1983).
54. J. Schindelin et al., Fiji: an open-source platform for biological-image analysis. Nat. Meth-ods. 9, 676–682 (2012).
55. T. Y. Zhang, C. Y. Suen, A fast parallel algorithm for thinning digital patterns. Commun ACM. 27, 236–239 (1984).
56. K. Benkrid, D. Crookes, A. Benkrid, in Proceedings of the Irish Machine Vision and Image Processing Conference (2000), pp. 51–57.
57. S. van der Walt et al., scikit-image: image processing in Python. PeerJ. 2, e453 (2014).
58. M. J. Zdilla et al., Circularity, solidity, axes of a best fit ellipse, aspect ratio, and roundness of the foramen ovale: A morphometric analysis with neurosurgical considerations. J Crani-ofac Surg. 27, 222–228 (2016).
59. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell tran-scriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420 (2018).
60. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018).
61. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to tran-sient cell states through dynamical modeling. BioRxiv (2019), doi:10.1101/820936.
62. F. A. Wolf et al., PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 20, 59 (2019).
63. S. Nowotschin et al., The emergent landscape of the mouse gut endoderm at single-cell res-olution. Nature. 569, 361–367 (2019).
64. M. Setty et al., Characterization of cell fate probabilities in single-cell data with Palantir. Nat. Biotechnol. 37, 451–460 (2019).
65. D. van Dijk et al., Recovering Gene Interactions from Single-Cell Data Using Data Diffu-sion. Cell. 174, 716–729.e27 (2018).
66. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 17, 10 (2011).
67. A. Dobin et al., STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29, 15–21 (2013).
68. M. Pertea et al., StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295 (2015).
69. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and' ' dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
70. A. Sergushichev, An algorithm for fast preranked gene set enrichment analysis using cumu-lative statistic calculation. BioRxiv (2016), doi:10.1101/060012.
71. J. Goedhart, A web app for plotting and annotating continuous data. PLoS Biol. (2020), doi: 10.1371/journal.pbio.3000581

Paper 2: *Generation of Mouse Pluripotent Stem Cell-derived Trunk-like Structures: an in vitro Model of Post-implantation Embryogenesis*

Authors:

Bolondi A, Haut L, Gassaloglu SI, Burton P, Kretzmer H, Buschow R, Meissner A, Herrmann BG, Veenvliet JV.

Published in ***BioProtocol***, 5th Jun 2021, Vol 11, Issue 11

DOI: 10.21769/BioProtoc.4042

Personal contribution:

I designed all the experiments and analysis, prepared the visualizations and figures, and wrote the manuscript. In detail, I performed needed experiments that contributed to all main figures during the project realization together with Haut L, Gassalog SI and Burton P. Initial and revised version of the manuscript, including figures, were written by me together with Veenvliet JV. Meissner A and Herrmann BG supervised the project. Buschow R and Kretzmer H helped with data analysis.

Link to the original publication:

https://doi.org/10.21769/BioProtoc.4042

bio-protocol

Generation of Mouse Pluripotent Stem Cell-derived Trunk-like Structures: An *in vitro* Model of Post-implantation Embryogenesis

Adriano Bolondi[1,2], Leah Haut[1], Seher Ipek Gassaloglu[3], Polly Burton[3], Helene Kretzmer[1], René Buschow[4], Alexander Meissner[1,2,5,6,*], Bernhard G. Herrmann[3,7,*] and Jesse V. Veenvliet[3,8,*]

[1]Dept. of Genome Regulation, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany;
[2]Institute of Chemistry and Biochemistry, Freie Universität Berlin, 14195 Berlin, Germany;
[3]Dept. of Developmental Genetics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; [4]Microscopy and Cryo-Electron Microscopy, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany;
[5]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA;
[6]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA;
[7]Institute for Medical Genetics, Charité - University Medicine Berlin, Campus Benjamin Franklin, 12203 Berlin, Germany;
[8]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany

[*]For correspondence: meissner@molgen.mpg.de; herrmann@molgen.mpg.de; veenvliet@mpi-cbg.de

Post-implantation mammalian embryogenesis involves profound molecular, cellular and morphogenetic changes. The study of these highly dynamic processes is complicated by the limited accessibility of *in utero* development. In recent years several complementary *in vitro* systems comprising self-organized assemblies of mouse embryonic stem cells, such as gastruloids, have been reported. We recently demonstrated that the morphogenetic potential of gastruloids can be further unlocked by the addition of a low percentage of Matrigel as an extra cellular matrix surrogate. This resulted in the formation of highly organized trunk-like structures (TLSs) with a neural tube that is frequently flanked by bilateral somites. Notably, development at the molecular and morphogenetic levels is highly reminiscent of the natural embryo. To facilitate access to this powerful model, here we provide a detailed step-by-step protocol that should allow any lab with access to standard cell culture techniques to implement the culture system. This will provide the user with a means to investigate early mid-gestational mouse embryogenesis at an unprecedented spatiotemporal resolution.

Keywords: Trunk-like structures, Gastrulation, Somites, Self-organization, Morphogenesis, Gastruloids, Embryoids, Organoids, *In vitro* models, Stem cells

Gastrulation and early organogenesis represent developmental events that are crucial for the successful generation of a functional body plan. In mammals, these processes start just after the embryo implants *in utero*, and within few days, a variety of morphologically and functionally diverse tissues emerge. It is currently difficult to study these highly dynamic changes *in vivo*, and *ex vivo* culture of post-implantation mouse embryos is laborious, costly, and requires rigorous training, which often renders it impractical for most laboratories. These impediments have led to extensive efforts to model post-implantation and early mid-gestational development *in vitro* using embryonic stem cells (reviewed in Shahbazi and Zernicka-Goetz 2018; Shahbazi *et al.*, 2019; Baillie-Benson *et al.*, 2020; Veenvliet and Herrmann, 2021). In particular, post-implantation development can be modelled with gastruloids, mouse or human embryonic stem cell (mESC/hESC) aggregates that self-organize (van den Brink *et al.*, 2014 and 2020; Moris *et al.*, 2020). The original mouse gastruloid culture protocol resulted in elongated structures with embryo-like expression domains similar to the post-occipital mouse embryo and with correct positioning of the three body axes, but limited morphogenesis (van den Brink *et al.*, 2014; Baillie-Johnson *et al.*, 2015; Beccari *et al.*, 2018a and 2018b; Turner *et al.*, 2017). More recent efforts have managed to introduce embryo-like morphological features by changing the cellular environment, such as the formation of somite-like structures or a heart tube (van den Brink *et al.*, 2020; Rossi *et al.*, 2021). Further advances have demonstrated that the addition of an extra-cellular matrix (ECM) surrogate to gastruloids can trigger a more embryo-like architecture with a gut tube as well as somites flanking a neural tube (Veenvliet *et al.*, 2020). We dubbed these embryonic organoids trunk-like structures (TLSs), since they resemble the core part of the trunk of an early mid-gestational embryo (~embryonic stage (E) 8.5-9). Importantly, during the timeframe of TLS induction (96-120 hour post-aggregation), the gene regulatory programs are highly similar to the developing embryo. Moreover, the segmentation clock, an oscillator driving the rhythmic deposition of somites *in vivo*, is active at an embryo-like pace in the TLS (Pourquié, 2003; Veenvliet *et al.*, 2020).

The TLS model is easy to access, track, manipulate, and scale, which makes it a powerful tool to study post-implantation and early mid-gestational mammalian development in a dish. Here, we provide a comprehensive step-by-step procedure to facilitate the generation of trunk-like structures. We also describe how to process TLSs for downstream analysis, including whole-mount immunofluorescent staining and (single cell) RNA sequencing.

Materials and Reagents

1. Pipet tips, variable volumes (Biozym, SafeSeal SurPhob VT)
2. 1.5 ml tubes (Sarstedt, catalog number: 72.706)
3. 15 ml Falcon tubes (Sarstedt, catalog number: 62.554.502)
4. 50 ml Falcon tubes (Sarstedt, catalog number: 62.547.254)
5. 6 cm cell culture plates (Sarstedt, catalog number: 83.3901.300)
6. Ultra-low attachment 96-well plates (Corning, Costar, catalog number: CLS7007)
7. 6-well cell culture plates (Corning, catalog number: 3516)
8. 10 cm cell culture plates (Corning, catalog number: 430167)
9. Luna cell counting slides (Logos Biosystems, catalog number: L12001)
10. µ-Slide 8-well glass bottom (Ibidi, catalog number: 80827)
11. Flowmi cell strainers 40 µm (Merck, catalog number: BAH136800040)
12. Bottle top vacuum filter unit (Corning, catalog number: CLS431096)
13. KnockOut DMEM (Gibco, catalog number: 10829018)
14. 100x Penicillin (5000 U/ml)-Streptomycin (5000 µg/ml) (Lonza, catalog number: DE17-603E)
15. 100x Glutamine, 200 mM (Lonza, catalog number: BE17-605E)
16. 100x Nucleosides (Sigma, catalog number: ES-008D)
17. Gibco 2-Mercaptoethanol, 55 mM solution in DPBS (Gibco, catalog number: 21985023)
18. Fetal Calf Serum (FCS), both regular (Pan Biotech, catalog number: P30-3306) and qualified and embryonic stem cell culture tested (Pan Biotech, catalog number: P30-2602)
19. TrypLE (Gibco, catalog number: 12604013) OR 0.05% Trypsin-EDTA (1x) (Gibco, catalog number: 25300-054)
20. NDiff 227 medium (Takara, catalog number: Y40002)
21. CHIR99021 InSolution (Sigma, catalog number: 361571) OR 10 mM in dimethyl sulfoxide (DMSO) (Tocris Biosciences, catalog number: 4423)
22. LDN193189 (Reprocell, catalog number: 04-0074-10)
23. DMSO (Sigma, catalog number: D2650)
24. Matrigel Growth Factor Reduced (GFR), Phenol Red-free (Corning, catalog number: 356231) – multiple lots/batches have been tested yielding similar results in terms of trunk-like-structure generation efficiency

25. Gelatin 2% solution (Sigma, catalog number: G1393)
26. DPBS, w/o $MgCl_2/CaCl_2$ (Gibco, catalog number: 14190144)
27. PBS with $MgCl_2/CaCl_2$ (Sigma, catalog number: D8662)
28. Murine Leukemia Inhibitory Factor (LIF) ESGRO™ (10^7U/ml) (Millipore, catalog number: ESG1107)
29. Trypan Blue (Bio-Rad, catalog number: 1450021)
30. UltraPure Dnase/Rnase-Free Distilled Water (Invitrogen, catalog number: 10977049)
31. Reagent Reservoirs 60 ml (Merck, catalog number: BR703411)
32. Bovine Serum Albumin powder (BSA) (Sigma, catalog number: A2153)
33. Dulbecco's Modified Eagle's Medium (DMEM) 4,500 mg/ml glucose, without sodium pyruvate (Lonza, catalog number: BE12-733F)
34. Cell culture grade water (Lonza, catalog number: BE17-724Q)
35. 0.1% Gelatin solution (see Recipes)
36. Mouse embryonic fibroblast (MEF) medium (see Recipes)
37. Mouse embryonic stem cell (mESC) medium (see Recipes)
38. PBS/0.5% BSA solution (see Recipes)

Equipment

1. Biological safety cabinet (Thermo Fisher Scientific, model: Herasafe KS12)
2. Clean horizontal laminar flow hood (Thermo Fisher Scientific, model: HeraGuard ECO)
3. Cell culture incubator (Thermo Fisher Scientific, model: Heracell Vios 160i)
4. Cell culture centrifuge (Eppendorf, model: Centrifuge 5804R)
5. Variable volume pipets and multichannel pipets (Eppendorf, model: Research® plus pipet)
6. Horizontal light source, Light ring (Nikon, P-DF LED Darkfield Unit) or other stereomicroscope stand
7. Automated cell counter (Logos biosystems, Luna automated cell counter, L10001)
8. Cell culture water bath (LAUDA Aqualine, catalog number: AL18)
9. Tissue culture vacuum pump (Vacuubrand, catalog number: 20727200)
10. Microcentrifuge (Eppendorf, model: 5424R)

Equipment set up:

1. Cell culture incubators are set to 37°C, 5% CO_2.

NOTE: We have also successfully generated TLSs at 7.5% CO_2, but routinely use 5%.

2. Cell culture water bath (set to 37°C).
3. All centrifugation steps are performed at room temperature, unless otherwise indicated.

Procedure

A. Seeding mouse embryonic fibroblasts (MEFs)

NOTES:

a. Seed MEFs at least one day prior to seeding the mESCs.

b. Pre-warm MEF medium in the water bath for at least 20 min before starting.

c. MEF plates should be used within one week of seeding.

1. Coat a 6 cm cell culture plate with 3 ml 0.1% gelatin solution.

   NOTE: Gelatin coated culture plates have to be prepared fresh on the day of seeding MEFs and cannot be stored.

2. Leave the plate at room temperature for 15 min.

   Next, thaw a vial of mitotically inactive MEFs at 37°C in the water bath.

   NOTES:

   a. Inactive MEFs are mitotically inactivated in-house using mitomycin C treatment (3 h at 37°C).

   b. You need 1.0 × 10^6 MEFs to coat a 6 cm cell culture plate. Thaw the appropriate number depending on the number of 6 cm cell culture plates needed.

3. Add the MEFs to a 15 ml Falcon tube containing 5 ml pre-warmed MEF medium.
4. Centrifuge the cells at 200 × *g* for 5 min.
5. While centrifuging, aspirate gelatin from each 6 cm plate and add 2 ml MEF medium.
6. Aspirate the supernatant from the 15 ml tube containing MEFs and resuspend the cell pellet at a concentration of 1.0 × 10^6 cells per ml.

   NOTE: Viable MEFs are counted at the time of freezing and there is no need to count them again after thawing.

7. Add 1 ml cell suspension to each prepared 6 cm plate.
8. Place the plate in the incubator and swirl the plate to ensure even distribution of cells.

B. Seeding mouse embryonic stem cells (mESCs)

NOTES:

a. Pre-warm mESC medium in the water bath for at least 20 min before starting.

b. We routinely use mESCs with an F1G4 genetic background for TLS protocol generation (George et al., 2007).

1. Thaw a vial of mESCs at 37°C in the water bath immediately before plating.

NOTE: You need 3.5 × 10^5 mESCs for a 6 cm MEF-coated plate. Thaw the appropriate number depending on the number of 6 cm cell culture plates needed.

2. Add the mESCs to a 15 ml Falcon tube containing 5 ml pre-warmed mESC medium.
3. Centrifuge the cells at 200 × *g* for 5 min.
4. While centrifuging, aspirate MEF medium from each 6 cm plate containing MEFs and add 2 ml mESC medium.
5. Aspirate the supernatant from the 15 ml tube containing mESCs and resuspend the cell pellet at a concentration of 3.5 × 10^5 cells per ml.

   NOTE: Viable mESCs are counted at the time of freezing and there is no need to count them again after thawing.
6. Add 1 ml cell suspension to each MEF coated 6 cm plate.
7. Place the plate in the incubator and swirl the plate to ensure even distribution of cells.
8. Replace the medium daily with 3 ml fresh mESC medium.

C. Passaging mESCs

NOTES:

a. Passage mESCs every 48 h at a splitting ratio of 1:8-1:10. Colony density and morphology should look similar to that shown in ***Figure 1A****. Do not let your culture overgrow (****Figure 1B****).*

b. The splitting time and ratios detailed here are optimized for the mESC lines used in Veenvliet et al., (2020). Based on the proliferation rate of the mESC line used, splitting times and ratios may need to be adjusted. This may be especially true if transgenic lines and/or mESCs with different genetic background are used.

c. Pre-warm mESC medium and TrypLE in the water bath for at least 20 min before starting.

d. Prepare MEF-coated plates one day prior to passaging mESCs.

e. Instead of TrypLE, 0.05% Trypsin-EDTA can be used.

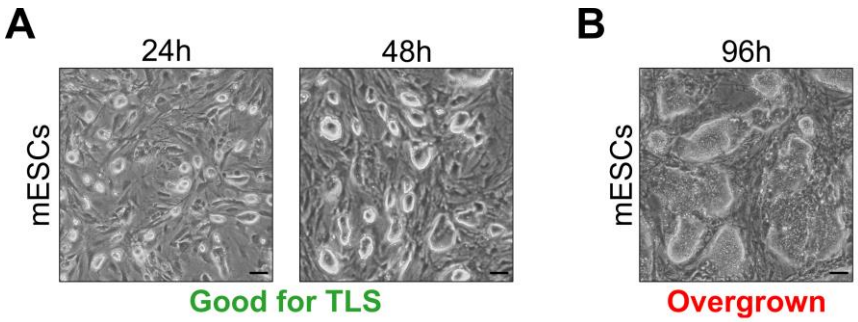

Figure 1. Optimal embryonic stem cell culture densities for successful TLS generation. A. mESC culture densities suitable for TLS generation (24 h and 48 h after seeding). **B.** mESC culture density unsuitable for TLS generation (96 h after seeding). Scale bars for all panels, 50 μm.

1. Aspirate the medium from the mESC plate and wash with 3 ml DPBS.
2. Aspirate the DPBS and add 1 ml TrypLE.
3. Ensure that the plate surface is evenly covered with TrypLE and place it in the incubator at 37°C for 5 min.
4. After 5 min, dislodge the colonies with a P1000 pipet set to 800 μl by pipetting up and down in the plate 20 times.
5. Inactivate the TrypLE by adding 1 ml mESC medium and pipet further to obtain a single cell suspension.
6. Transfer the cell suspension to a 15 ml Falcon tube and wash the plate with an additional 3 ml mESC medium to recover all cells. Transfer these cells to the same 15 ml tube.
7. Centrifuge the cells at 200 × *g* for 5 min.
8. While centrifuging, aspirate the MEF medium from the previously prepared 6 cm plate containing MEFs and add 2 ml mESC medium.
9. Aspirate the supernatant from the 15 ml tube containing mESCs and resuspend the cell pellet in 2 ml mESCs medium.
10. Add the appropriate amount of cell suspension to each MEF-coated 6 cm plate (ratio 1:8-1:10 → 200-250 μl).
11. Adjust the final volume to 3 ml.
12. Place the plate in the incubator and swirl the plate to ensure even distribution of cells.
13. Replace the medium daily with 3 ml fresh pre-warmed mESC medium.

D. Generation of trunk-like structures (TLSs)

NOTES:

a. *The input cell number for each well detailed here is optimized for the mESC lines used in Veenvliet et al., (2020). Based on the proliferation rate of the mESC line (especially for transgenic lines and/or mESCs with a different genetic background), the cell amount may need to be adjusted to reach the same efficiency reported in Veenvliet et al., (2020).*
b. *We recommend first optimizing the standard gastruloid protocol for new cell lines, using gastruloid elongation efficiency as a fast experimental readout (Cermola et al., 2021). In our experience, good gastruloid elongation efficiency (>95%) is essential to achieve a similar TLS efficiency to that reported in Veenvliet et al., (2020). A routine optimization procedure involves the seeding of 100-600 mESCs per well, with a step wise increase of 50 cells.*
c. *mESCs must be in culture for at least one passage before starting.*
d. *Pre-warm mESC medium and TrypLE in the water bath for at least 20 min before starting.*
e. *Here, we use commercially available, quality controlled NDiff 227 medium (N2B27). We and others have successfully generated gastruloids with home-made N2B27 (Beccari et al., 2018b); however, in our hands, more robust results of the gastruloid and TLS protocols are obtained with the NDiff 227 medium.*
f. *Instead of TrypLE, 0.05% Trypsin-EDTA can be used.*
g. *A schematic overview of the TLS generation protocol indicating critical timepoints is provided in* ***Figure 2.***

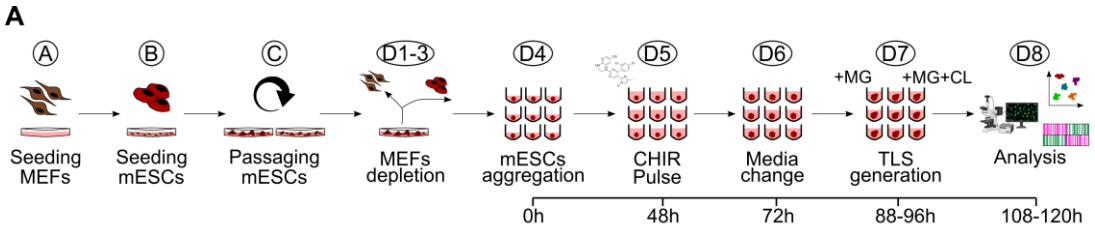

Figure 2. Schematic overview of the TLS generation protocol. A. Workflow for the generation of trunk-like structures (TLS) from seeding of MEFs to downstream analysis. MG, Matrigel; CL, CHIR+LDN; MEFs, mouse embryonic fibroblasts; mESCs, mouse embryonic stem cells.

D1. Prepare 6-well plate for MEF depletion

1. Coat three wells of a 6-well plate with 2 ml 0.1% gelatin solution for each 6 cm plate that will be used for TLS generation.
2. Incubate the 6-well plate at room temperature for 15 min.
3. Aspirate gelatin solution and add 1 ml mESC medium to each well.
4. Store plate in the incubator until use.

D2. Prepare a single cell suspension

1. Aspirate the medium from the mESC plate and wash with 3 ml DPBS.
2. Aspirate the DPBS and add 1 ml TrypLE.
3. Ensure that the plate surface is evenly covered with TrypLE and place it in the incubator at 37°C for 5 min.
4. After 5 min, dislodge the colonies with a P1000 pipet set to 800 µl by pipetting up and down in the plate 20 times.
5. Inactivate the TrypLE by adding 1 ml mESC medium and pipet-mix.
6. Transfer the cell suspension to a 15 ml Falcon tube and wash the plate with an additional 3 ml mESC medium to recover all cells. Transfer these cells to the same 15 ml tube.
7. Centrifuge the cells at 200 × *g* for 5 min.
8. Resuspend the cell pellet in 1 ml mESC medium and pipet up and down 50 times.
   NOTE: Here it is critical to achieve a proper single cell suspension to avoid losing mESCs (or retaining MEFs) during MEF depletion and to ensure the best protocol performance. We recommend checking for a proper single cell suspension under a microscope.

D3. MEF depletion

NOTE: An example how the wells with cells attached to the bottom should look like after each step of MEF depletion is provided in ***Figure 3****.*

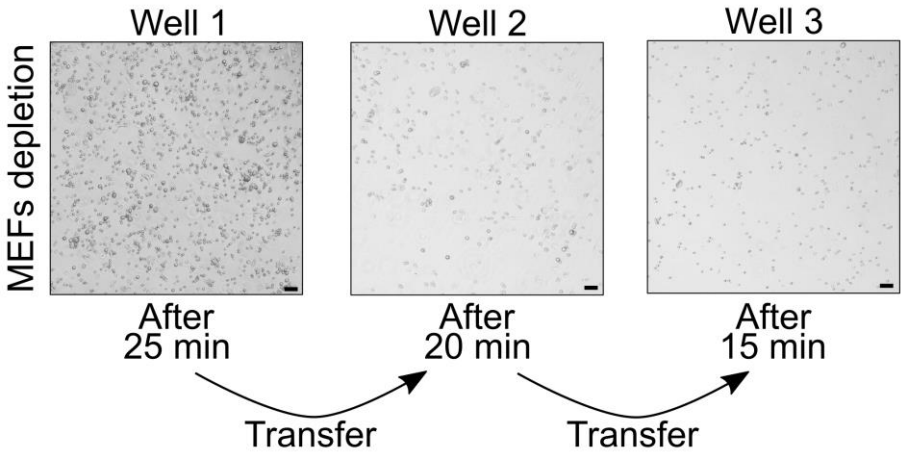

Figure 3. MEF depletion prior to mESC aggregation. MEFs attach to the 0.1% gelatin-coated wells. Scale bars, 50 μm.
NOTE: With consecutive transfers, the amount of attached cells decreases. After the third incubation MEF depletion is completed and mESCs are ready to be used for aggregation.

1. Transfer the obtained cell suspension to a well of the prepared 6-well plate (see D1).
   NOTE: Transfer the amount of mESCs present in one 6 cm plate into one well of the prepared 6-well plate. The presence of too many cells could result in decreased depletion efficiency.
2. Pipet-mix 10 times.
3. Place the plate in the incubator and swirl the plate to ensure even distribution of cells. Leave untouched for 25 min.
4. Next use a P1000 pipet set to 1 ml to carefully transfer all the cells in suspension to another well.
   NOTE: It is critical not to dislodge the MEFs, which are attached to the bottom of the wells.
5. Pipet up and down 10 times in the new well to ensure a single cell suspension.
   NOTE: Cells may clump during the incubation; therefore, it is critical to pipet once the cells are transferred to the new well. We recommend confirming under the microscope that you have obtained a proper single cell suspension.
6. Place the plate in the incubator and swirl the plate to ensure even distribution of cells. Leave untouched for 20 min.

7. Next, use a P1000 pipet set to 1 ml to carefully transfer all the cells in suspension to another well.
   NOTE: It is critical not to dislodge the MEFs, which are attached to the bottom of the wells.
8. Pipet up and down 10 times in the new well to ensure a single cell suspension.
   NOTE: Cells may clump during the incubation; therefore, it is critical to pipet once the cells are transferred to the new well. We recommend confirming under the microscope that you have obtained a proper single cell suspension.
9. Place the plate in the incubator and swirl the plate to ensure even distribution of cells. Leave untouched for 15 min.
10. During this last 15 min step, equilibrate the required amount of NDiff 227 in a 10 cm dish in the incubator for at least 20 min. Longer incubation is also possible (*e.g.*, NDiff 227 can be placed in the incubator after step 6. See **Table 1** for the volume needed as a function of the number of 96-well plates to seed).
    NOTE: NDiff 227 is light-sensitive and should be protected from (direct) light as much as possible.
11. Carefully transfer all MEF-depleted mESCs to a 15 ml Falcon tube with a P1000 pipet.
    NOTE: It is critical not to dislodge the MEFs, which are attached to the bottom of the wells.

D4. mESC aggregation (0 h)

NOTE: The first 96 h of the TLS protocol are similar to the gastruloids protocol (Baillie-Johnson et al., 2015; van den Brink et al., 2014; Beccari et al., 2018b; Anlas et al., 2021). Detailed protocols for gastruloid formation, including troubleshooting, are provided elsewhere (Baillie-Johnson et al., 2015; Beccari et al., 2018b; Anlas et al., 2021).

1. Centrifuge the cells at 200 × *g* for 5 min.
2. Resuspend the cell pellet in 5 ml PBS with $MgCl_2$/$CaCl_2$ and pipet up and down 20 times (Wash 1).
3. Centrifuge the cells at 200 × *g* for 5 min.
   NOTE: In case of low starting cell numbers steps 2 and 3 can be omitted. This may however slightly compromise protocol efficiency.
4. Resuspend the cell pellet in 5 ml pre-equilibrated NDiff 227 and pipet up and down 20 times (Wash 2).
5. Centrifuge the cells at 200 × *g* for 5 min.

6. Resuspend the cell pellet in 500 µl pre-equilibrated NDiff 227 and pipet up and down 30 times. *NOTE: It is critical to obtain a single cell suspension prior to counting and plating.*
7. For counting, prepare a 1:2 dilution of the cell suspension by adding 10 µl cell suspension to 10 µl Trypan Blue.
8. Count using the Luna automated cell counter with the following set up: Dilution factor → 2; Noise reduction → 5; Live detection sensitivity → 5; Roundness → 85%; Min cell size → 10 µm; Max cell size → 20 µm; Declustering level → High.
9. Transfer the amount of cells needed for the experiment to a new Falcon tube (see **Table 1** for the number of cells needed as a function of the number of 96-well plates to seed).
10. Add the pre-incubated NDiff 227 volume needed to bring the cell suspension to a concentration of 5.7×10^3 cells/ml (see **Table 1** for volume to add as a function of the number of 96-well plates to seed).

    NOTE: This cell concentration is optimized for an input of 200 cells/well, which has been shown to give high TLS generation efficiency for all cell lines tested (Veenvliet et al., 2020). Similar results were obtained for inputs ranging from 200 to 250 cells/well.
11. Mix the new cell suspension vigorously and transfer it to a reservoir.
12. Use a multichannel pipet to transfer 35 µl to each well of an ultra-low attachment 96-well plate. Pipet gently up and down in the reservoir between each transfer.
13. Gently tap the plate 5 times on a clean bench, transfer to the incubator, and allow undisturbed aggregation for 48 h.

    NOTE: Keeping NDiff 227 outside the incubator for longer periods of time (more than 5 min) will lead to disequilibration of the medium. Therefore, try to avoid keeping NDiff 227 or plates with freshly seeded cells in NDiff 227 out the incubator for too long. In the case of handling multiple plates, we recommend putting each plate into the incubator directly after pipetting.

Number of 96-well plates	Volume of NDiff 227 to equilibrate (ml)	Total number of cells needed	Volume of NDiff 227 required to reach 5.7×10^3 cells/ml (ml)
1	12	2.2×10^4	3.85 *minus* cell volume
2	16	4.4×10^4	7.70 *minus* cell volume
3	20	6.6×10^4	11.55 *minus* cell volume
4	24	8.8×10^4	15.40 *minus* cell volume
5	28	11×10^4	19.25 *minus* cell volume

Table 1. Cell numbers and volumes required for mESC aggregation. The amounts are calculated for an input of 200 cells/well in 35 µl. Volumes and cell numbers in column 4 are calculated for 100 samples (instead of 96) per plate *plus* a 10% excess dead volume. In column 2, the volume of NDiff 227 to equilibrate is calculated based on the amount needed for washing and counting the cells (5.5 ml per experiment, independent of the number of plates, see step D4.4 and D4.6), *plus* the amount indicated in column 4, *plus* an extra volume to account for dead space in the dish and evaporation during medium equilibration.

D5. CHIR pulse (48 h)

NOTES:

a. Start this procedure at least one hour before the end of the 48 h.

*b. Cells at 48 h should have formed one single round aggregate with a diameter measuring 214 ± 13 µm (see **Figure 4A**).*

c. CHIR99021 is light-sensitive and should be protected from (direct) light as much as possible.

d. CHIR99021 should be aliquoted in single-use aliquots in brown (light-protected) sterile tubes upon arrival and not subjected to repeated freeze-thaw cycles.

1. Equilibrate the needed amount of NDiff 227 in a 10-cm dish in the incubator for at least 20 min. (see **Table 2** for the volume needed as a function of the number of 96-well plates used in the experiment).
2. Transfer the required amount of NDiff 227 to a 50-ml Falcon tube (see **Table 2** for the volume needed as a function of the number of 96-well plates used in the experiment).
3. Add CHIR99021 to the NDiff 227 medium to obtain a final concentration of 3 µM.
4. Mix the medium vigorously and transfer it to a reservoir.

5. Use a multichannel pipet to add 150 µl CHIR99021 supplemented medium to each well of the plates containing the aggregates.
6. Gently tap the plate 10 times on a clean bench, transfer to the incubator, and allow further undisturbed development for 24 h.

   NOTES:

   a. *The tapping is critical to prevent cell aggregates from attaching to the culture plates. Ensure that the aggregates are freely moving immediately after tapping (this can be checked under the microscope).*
   b. *Take caution to avoid splashing medium on the lid while tapping the plates.*

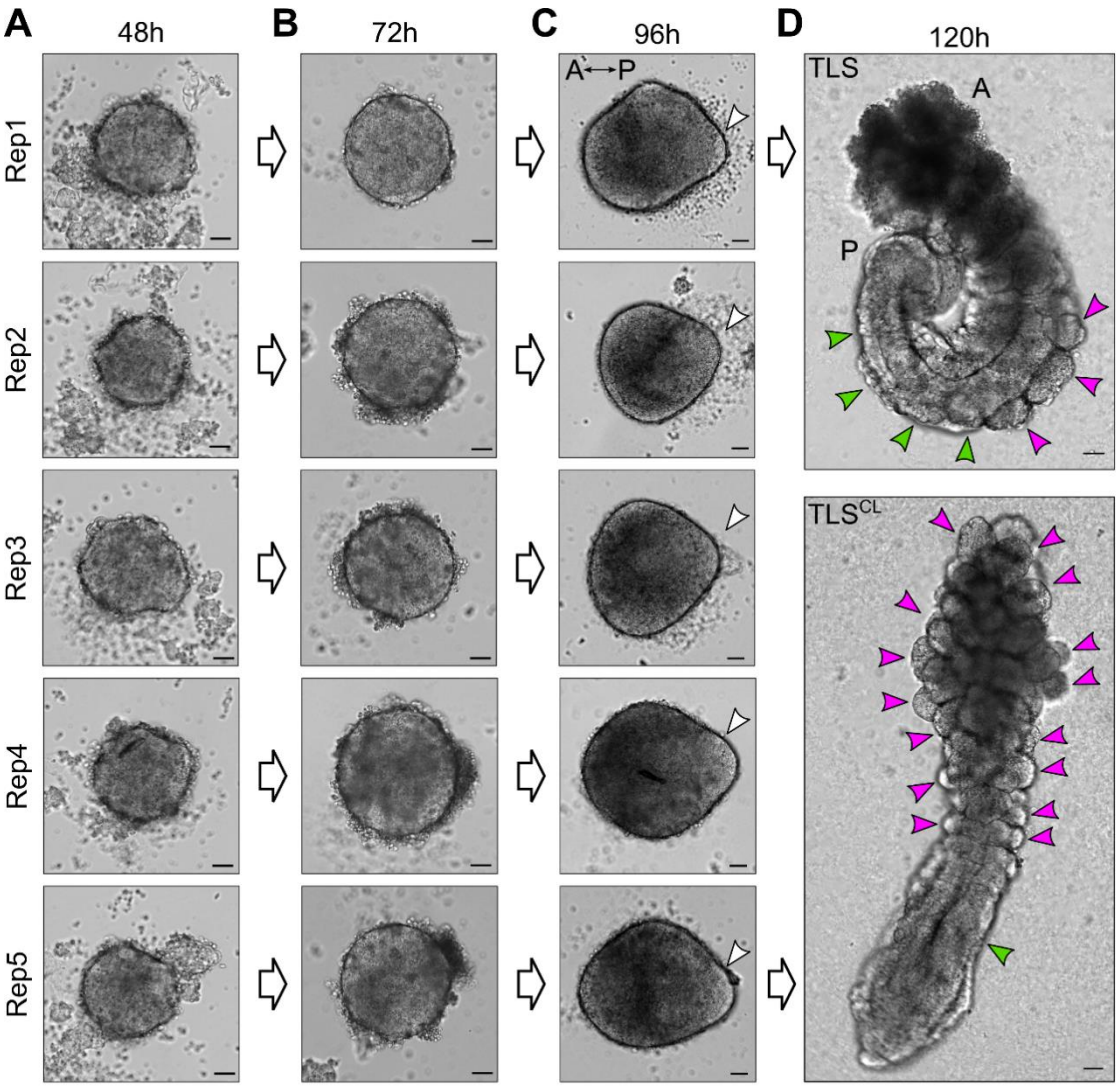

Figure 4. Examples of expected morphology of mESC-derived aggregates at several timepoints during TLS generation. A, B. mESC-derived aggregates at 48 h and 72 h after aggregation are round without clear signs of symmetry breaking. **C.** At 96 h after aggregation, the structures have clearly broken symmetry and are teardrop shaped. The white arrowheads indicate the posterior pole, where localized expression of Brachyury is expected. Note that, depending on the cell line used, aggregates may establish the teardrop-like morphology prior to 96 h. In that case, structures should be embedded in Matrigel earlier (as soon as the teardrop-like morphology is observed) to achieve optimal TLS efficiency (see main text for details). **D.** Upon addition of 5% Matrigel, the aggregates will establish an architecture reminiscent of the embryonic trunk, with somites (magenta arrowheads) flanking a neural tube (green arrowheads). Chemical modulation with a WNT agonist (5 μM CHIR99021) and BMP inhibitor (600 nM LDN193189) results in compromised neural tube development and formation of excess somites arranged like a "bunch-of-grapes" (TLS^{CL}). Scale bars for all panels, 50 μm. A, Anterior; P, Posterior.

D6. Media change (72 h)

NOTES:

a. Start this procedure at least one hour before the end of the 72 h.

*b. Aggregates at 72 h should look like the example given in **Figure 4B** and measure 244 ± 15 µm in diameter.*

c. If available, perform this step on a clean bench containing a stereoscope or a light ring to help visualize the structures and avoid loss of structures while pipetting off the old medium.

1. Equilibrate the required amount of NDiff 227 in a 10-cm dish in the incubator for at least 20 min. (see **Table 2** for the volume needed as a function of the number of 96-well plates used in the experiment).
2. Transfer the required amount of NDiff 227 to a 50-ml tube (see **Table 2** for the volume needed as a function of the number of 96-well plates used in the experiment).
3. Use a multichannel pipet to remove 150 µl from each well without disturbing the structure.

   *NOTE: CRITICAL STEP. Avoid losing structures while pipetting off the medium. This is best achieved by keeping the plate under a 30-40° angle and putting the pipet tips against the side opposite to that where the aggregate should be located. As a visual aid, a stereoscope or light ring could be used, as stated above (see **Figure 5** for a schematic of how to position the plate and tips).*

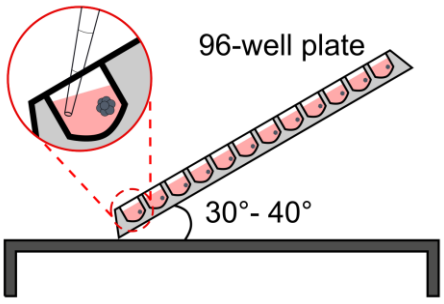

Figure 5. Schematic representation of plate positioning for media changes during the TLS generation procedure. The plate is tilted at a 30-40° angle on a clean bench, and the media is carefully aspirated with a multichannel pipet avoiding disturbance of the aggregates.

4. Pour the pre-equilibrated medium in a reservoir.
5. Use a multichannel pipet to add 150 µl medium to each well of the plates containing the aggregates.
6. Gently tap the plate 10 times on a clean bench, transfer to the incubator, and allow further undisturbed development for 16-24 h.

   NOTES:

   a. *The tapping is critical to prevent cell aggregates from attaching to the culture plates. Ensure that the aggregates are freely moving immediately after tapping (this can be checked under the microscope).*
   b. *Tale caution to avoid splashing medium on the lid while tapping the plates.*
   c. *Thaw overnight at 4°C on ice the amount of Matrigel required the following day (see **Table 2** for the volume needed as a function of the number of 96-well plates used in the experiment).*
   d. *We have used multiple Matrigel batches with comparable results.*

D7. TLS generation (88-96 h)

NOTES:

a. *Start monitoring the TLSs around 88 h after aggregation for the appearance of a "teardrop-like" shape (see **Figure 4C**). Structures should present a longer axis (421 ± 33 µm) and a shorter axis (337 ± 30 µm), with an axis ratio of 0.8 ± 0.07.*
b. *Start this procedure immediately when a "teardrop-like" shape is observed in the majority of the TLSs (or latest 96 h after aggregation) to achieve optimal TLS formation efficiency.*
c. *If available, perform this step on a clean bench containing a stereoscope or a light ring to help visualize the structures and avoid losing structures while pipetting off the old medium.*
d. *If performing chemical modulation at this step, follow the "Variant protocol: chemical modulation during TLS generation."*

1. Equilibrate the required amount of NDiff 227 in a 10-cm dish in the incubator for at least 20 min (see **Table 2** for volume needed as a function of the number of 96-well plates used in the experiment).
2. Transfer the required amount of NDiff 227 to a 50-ml Falcon tube and place it on ice (see **Table 2** for the volume required as a function of the number of 96-well plates used in the experiment).

3. Once the medium has cooled down, supplement it with the correct volume of Matrigel on ice and mix vigorously.

   NOTE: It is critical, while handling Matrigel, that every step is performed on ice to avoid clumping. It is also recommended to pre-cool the pipet tips used for the handling of 100% Matrigel by placing the box in the fridge until use; see ***Table 2*** *for the volume needed as a function of the number of 96-well plates used in the experiment.*
4. Move the Falcon tube with Matrigel-supplemented medium to room temperature.
5. Use a multichannel pipet to remove 150 µl from each well without disturbing the structure.

   NOTES:

   a. *CRITICAL STEP. Avoid losing structures while pipetting out the medium. This is best achieved by keeping the plate under a 30-40° angle and putting the pipet tips against the side opposite to that where the aggregate should be located. As a visual aid, a stereoscope or light ring could be used, as stated above (see* ***Figure 5*** *for a schematic of how to position the plate and tips).*
   b. *This step should not take more than 5 min after Matrigel-supplemented medium has been placed at room temperature. If there are multiple plates, it is advisable to remove the medium from all plates before equilibrating the Matrigel-supplemented medium at room temperature and keep the structures in the incubator.*
6. Pour the Matrigel-supplemented medium into a reservoir.
7. Use a multichannel pipet to add 150 µl Matrigel-supplemented medium in each well of the plates containing the aggregates.
8. Gently tap the plate 10 times on a clean bench and transfer to the incubator.

   NOTES:

   a. *The tapping is critical to prevent cell aggregates from attaching to the culture plates. Ensure that aggregates are freely moving immediately after tapping (this can be checked under the microscope).*
   b. *Take caution to avoid splashing medium on the lid while tapping the plates.*
   c. *If performing Live Cell Imaging after embedding, allow the TLSs to settle for 1 h in the incubator before starting imaging.*

Variant protocol: chemical modulation during TLS embedding.

Follow “D7. TLS generation (88-96 h)” for all passages except for point 3.

3’. Once the medium has cooled down, supplement it with the correct volume of Matrigel and appropriate chemical compounds for your modulations on ice and mix vigorously.

NOTES:

a. *It is critical, while handling Matrigel, that every step is performed on ice to avoid clumping; see* ***Table 2*** *for the volume needed as a function of the number of 96-well plates used in the experiment.*
b. *Add an equal volume of diluent to the control sample TLSs when performing chemical modulations.*
c. *In Veenvliet et al., (2020), we induced excess somite production and compromised neural tube development by supplementing the medium with 5 μM CHIR99021, alone (*TLS^{C}*) or in combination with 600 nM LDN193189 (*TLS^{CL}*).*

Number of 96-well plates	Volume of NDiff 227 to equilibrate (D5, D6, D7) (ml)	Volume of NDiff 227 to transfer in the 50-ml tube/s (D5, D6) (ml)**	Volume of NDiff 227 to transfer in the 50-ml tube/s (D7) (ml)**	Volume of 10 mM CHIR99021 to add for 3 µM (D5 only) (µl)	Volume of Matrigel to add for 5% final (D7 only) (ml)***
1	18	16	15.01	4.8	0.99
2	38*	32	30.03	9.6	1.97
3	52*	48	45.04	14.4	2.96
4	70*	64	60.05	19.2	3.95
5	86*	80	75.06	24	4.94

Table 2. Volumes of NDiff 227, CHIR99021, and Matrigel required during the last three steps (protocol steps D5, D6, D7) of the TLS generation protocol. In column 2, the volume of NDiff 227 to equilibrate is calculated including an excess volume (2 ml for each dish that is used for medium equilibration) to account for dead space in the dish(es) and medium evaporation during equilibration.
*For volumes higher than 30 ml, use more than one 10-cm plate to equilibrate NDiff 227 in the incubator.
**Use two 50-ml tubes if the volume exceeds 50 ml.
***Note that we calculate 5% v/v Matrigel as a function of the final volume in each well (35 µl + 150 µl = 185 µl) (as opposed to CHIR99021). This means that in the NDiff 227, the concentration of Matrigel is slightly higher than 5% (6.17%). For instance, for 1 plate of TLSs, the volume of Matrigel added is calculated as (0.05*16)*(185/150) = 0.987 ml.

EXPECTED OUTCOME:

TLSs at 120 h should look elongated with a clear anterior and posterior domain (**Figure 4D upper panel**; **Figure 6A**). Moreover, they should show clear segmentation (somites) on one or both sides of a tubular structure (neural tube). TLSs subjected to chemical modulation (TLS^{CL}) should display compromised neural tube development and formation of excess somites arranged like a “bunch-of-grapes” (**Figure 4D lower panel**; **Figure 6B**).

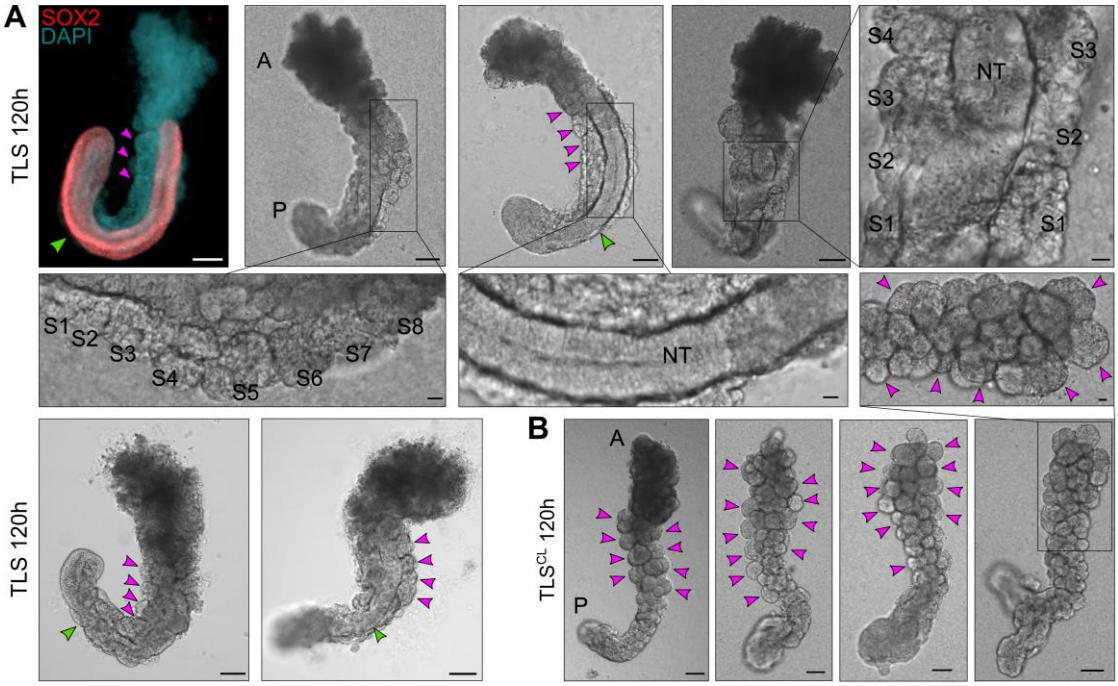

Figure 6. Expected outcome of the TLS protocol. A. Examples of trunk-like structures (TLSs) 24 h after addition of 5% Matrigel (total culture time 120 h). The left upper structure is immunostained with a SOX2 antibody and counterstained with DAPI, labeling the neural tube and nuclei, respectively. Somites are indicated with magenta arrowheads or with S1, S2, *etc.* (in magnifications); S1-S8, Somite 1-Somite 8. Neural tube (NT) is indicated with a green arrowhead. A, Anterior; P, Posterior. **B.** Expected outcome of TLSs subjected to chemical modulation (TLS^{CL}) is compromised neural tube development and formation of excess somites arranged like a "bunch-of-grapes." Somites are indicated with magenta arrowheads. Scale bars, 100 µm (whole structure) or 20 µm (magnifications). A, Anterior; P, Posterior.

D8. TLS analysis (108-120 h)

NOTES:

a. *Depending on the specific biological question, the exact time of analysis may vary (see Veenvliet et al., 2020 for time-resolved expression dynamics).*
b. *If available, perform this step on a clean bench containing a stereoscope or a light ring to help visualize the structures and avoid losing structures while processing them.*
c. *The following protocol variants are performed (D8', D8'', D8''') depending on the downstream applications.*
d. *Prepare a P200 tip box with the tip cut-off at the 50-µl mark for TLS picking (if using unmarked tips, cut approximately 9 mm off the tip).*

D8'. Staining and imaging

1. Use a P200 pipet set to 50 µl and the cut-off tips to manually pick each individual TLS that needs to be analyzed in a well of an Ibidi 8-chamber plate.
2. Add 200 µl cold PBS/0.5% BSA solution to each well containing a TLS.
3. Remove 200 µl with a pipet and perform the same 200 µl cold PBS/0.5% BSA solution wash three times.
4. Use the fixative of interest to fix TLSs in the Ibidi plate for the downstream protocol.
5. Perform the rest of the staining protocol in the Ibidi plate and image the structure with the desired microscope and settings.

   NOTE: We have so far used 4% PFA fixation for whole-mount immunofluorescence (WIFC) as well as whole-mount in situ hybridization (WISH). A detailed description of the protocols used for WIFC and WISH, including method-specific fixation times and downstream processing, is provided in the Supplemental Information of Veenvliet et al., (2020).

D8''. RNA extraction

1. Use a P200 pipet set to 50 µl and the cut-off tips to manually pick and transfer each individual TLS that needs to be analyzed to a 1.5-ml tube containing 1 ml cold PBS/0.5% BSA solution.

   NOTE: The number of TLSs that are pooled in one tube depends on the downstream application and experiment.
2. Centrifuge the TLSs at 200 × *g* for 1 min at 4°C.
3. Remove the supernatant with a P1000 pipet while being careful not to disturb the TLS pellet.
4. Wash the structures with 1 ml cold PBS/0.5% BSA solution.

   NOTE: Ensure loosening of the pellet without aspirating into the pipet tip.
5. Centrifuge the TLSs at 200 × *g* for 1 min at 4°C.
6. Remove the supernatant with a P1000 pipet while being careful not to disturb the TLS pellet.
7. Add the indicated amount of RNA lysis buffer or TRIzol depending on the desired RNA extraction strategy.

D8’’’. 10x Genomics single-cell RNA sequencing

NOTES:

a. This section explains how to process TLSs to generate a single cell suspension suitable for efficient Gel Bead-in-Emulsion (GEMs) generation. Follow the manufacturer’s instructions for every step after the single cell suspension has been generated and counted.

b. Pre-warm TrypLE in the water bath for at least 20 min before starting.

1. Use a P200 pipet set to 20 µl and the cut-off tips to manually pick each individual TLS that needs to be analyzed in a well of an ultra-low attachment 96-well plate containing 200 µl cold PBS.
2. Transfer each TLS serially five times to new wells containing 200 µl cold PBS.

   NOTES:

   a. It is CRITICAL to carry over as little volume as possible from the culture to minimize the amount of Matrigel transferred. Carrying over excess amounts of Matrigel can lead to microfluidics clogging during GEM generation.

   b. Since washing of TLSs is performed in PBS without BSA, the structures may become sticky and get stuck to the tip wall. Avoid this by pipetting a very low volume to maintain the structure at the liquid/air interphase in the tip.
3. After the five washes, transfer all the structures into a single drop of 200 µl pre-warmed TrypLE in the center of a 6-cm plate.
4. Transfer the plate to the incubator and allow cell dissociation for 25 min with pipetting every 5 min to ensure that a single cell suspension is achieved.

   NOTE: Perform the pipetting steps under a stereoscope to monitor the degree of cell dissociation and ensure no loss of material.
5. At the end of the 25 min, and after verifying correct achievement of a single cell suspension, transfer the cell suspension to a 1.5-ml tube on ice.
6. To ensure maximum cell recovery and to quench the trypsinization reaction, wash the part of the plate where the drop was located four times with 200 µl PBS/0.5% BSA solution. Add every wash to the same tube (from step 5) containing the cell suspension.
7. Filter the cell suspension using a P1000 set to 1 ml through a 40 µm Flowmi Cell Strainer in a new 1.5-ml tube on ice.
8. Centrifuge the cell suspension at 300 × *g* for 5 min at 4°C.
9. Remove 800 µl supernatant with a P1000 while being careful not to disturb the cell pellet.

NOTE: The cell pellet may be very small and barely visible, so be extremely careful during these steps.

10. Wash with 1 ml PBS/0.5% BSA solution.
11. Centrifuge the cell suspension at 300 × *g* for 5 min at 4°C.
12. Remove 800 µl supernatant with a P1000 while being careful not to disturb the cell pellet.

    NOTE: The cell pellet may be very small and barely visible, so be extremely careful during these steps.
13. Resuspend the pellet in the remaining 200 µl left in the tube.
14. Centrifuge the cell suspension at 300 × *g* for 5 min at 4°C.
15. Remove the supernatant with a P200 pipet, leaving ~42 µl in the tube.

    NOTE: Use another tube containing exactly 42 µl PBS/0.5% BSA solution as a guide to evaluate the approximate volume to leave in the tube.
16. Resuspend the cell pellet in the ~42 µl left and determine the cell suspension concentration using a manual hemocytometer (analyze a 1:5 cell suspension dilution by adding 2 µl cell suspension to 8 µl Trypan Blue).
17. Proceed with the desired amount of cells for GEM generation following the manufacturer's instructions.

Data analysis

All data and analysis needed for the development and characterization of this protocol are available in the main text or Supplemental Information of Veenvliet *et al.,* (2020).

Recipes

1. **0.1% Gelatin solution**

   Dilute sterile 2% Gelatin to 0.1% in cell culture grade water. Store at 4°C
2. **Mouse embryonic fibroblast (MEF) medium**

   NOTE: Heat inactivate the FCS for 30 min at 56°C before use.

   500 ml Dulbecco's Modified Eagle's Medium (DMEM)

   55 ml **regular** FCS (Pan Biotech; catalog number: P30-3306)

   5.5 ml 100× Glutamine

   5.5 ml 100× Penicillin/Streptomycin

Sterile filter

Store at 4°C

3. **Mouse embryonic stem cell (mESC) medium**

   NOTE: Heat inactivate the FCS for 30 min at 56°C before use.

   400 ml Knockout Dulbecco's Modified Eagle's Medium (KO-DMEM)

   75 ml **mESC tested** FCS (Pan Biotech; catalog number: P30-2602)

   5 ml 100× Glutamine

   5 ml 100× Penicillin/Streptomycin

   5 ml 100× Nucleosides

   1 ml Gibco 2-Mercaptoethanol

   Sterile filter

   Aliquot mESC medium without LIF in 40 ml portions and freeze at -20°C

   Thaw before use, then store at 4°C

   Add 1:10,000 LIF immediately before use

   Store at 4°C

   NOTE: Homemade LIF has also been successfully used; however, the right concentration has to be tested based on the purification protocol and batch concentrations.

4. **PBS/0.5% BSA solution**

   PBS with $MgCl_2$/$CaCl_2$

   0.5% BSA powder

   Prepare fresh, sterile filter, keep on ice for the procedure

   Store at 4°C

Acknowledgments

This work is based on and adapted from the method published in Veenvliet *et al.*, (2020). We are grateful for the support and feedback received to develop and characterize our *in vitro* system from present and past members of the Herrmann & Meissner laboratories, in particular Manuela Scholze-Wittler, Dennis Schifferl, Frederic Koch, Abhishek Sampath Kumar, Milena Pustet, Fabian Tobor, Simon Heimann, Lars Wittler, Stefanie Grosswendt, Zachary Smith and Atsuhiro Taguchi. The work was supported by an Alexander von Humboldt Fellowship (J.V.V.), NIH grant HG006193 (A.M.), and the Max Planck Society.

Competing interests

The authors declare no competing interests.

References

1. Anlas, K., Baillie-Benson, P., Arato, K., Turner, D. A. and Trivedi, V. (2021). Gastruloids: Embryonic Organoids from Mouse Embryonic Stem Cells to Study Patterning and Development in Early Mammalian Embryos. *Methods Mol Biol* 2258: 131-147.
2. Baillie-Benson, P., Moris, N. and Martinez Arias, A. (2020). Pluripotent stem cell models of early mammalian development. *Curr Opin Cell Biol* 66: 89-96.
3. 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*(105).
4. Beccari, L., Moris, N., Girgin, M., Turner, D. A., Baillie-Johnson, P., Cossy, A. C., Lutolf, M. P., Duboule, D. and Arias, A. M. (2018a). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* 562(7726): 272-276.
5. Beccari, L., Girgin, M., Turner, D., Baillie-Johnson, P., Cossy, A.-C., Moris, N., Lutolf, M., Duboule, D. and Martinez Arias, A. (2018b). Generating Gastruloids from Mouse Embryonic Stem Cells. *Protocol Exchange*.
6. Cermola, F., D'Aniello, C., Tatè, R., De Cesare, D., Martinez-Arias, A., Minchiotti, G. and Patriarca, E. J. J. B. (2021). Gastruloid development competence discriminates different states of pluripotency between naïve and primed. *Stem Cell Reports* 16(2): 354-369.
7. George, S. H., Gertsenstein, M., Vintersten, K., Korets-Smith, E., Murphy, J., Stevens, M. E., Haigh, J. J. and Nagy, A. (2007). Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proc Natl Acad Sci U S A* 104(11): 4455-4460.
8. Moris, N., Anlas, K., van den Brink, S. C., Alemany, A., Schroder, J., Ghimire, S., Balayo, T., van Oudenaarden, A. and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. *Nature* 582(7812): 410-415.
9. Pourquié, O. (2003). The segmentation clock: converting embryonic time into spatial pattern. *Science* 301(5631): 328-330.
10. 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(2):230-240.e6.
11. Shahbazi, M. N., Siggia, E. D. and Zernicka-Goetz, M. (2019). Self-organization of stem cells into embryos: A window on early mammalian development. *Science* 364(6444): 948-951.
12. Shahbazi, M. N. and Zernicka-Goetz, M. (2018). Deconstructing and reconstructing the

mouse and human early embryo. *Nat Cell Biol* 20(8): 878-887.

13. 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., Steventon, B., M, P. L. and Arias, A. M. (2017). Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids. *Development* 144(21): 3894-3906.
14. van den Brink, S. C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivie, J., Baillie-Johnson, P., Nichols, J., Sonnen, K. F., Martinez Arias, A. and van Oudenaarden, A. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* 582(7812): 405-409.
15. 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(22): 4231-4242.
16. Veenvliet, J. V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Guignard, L., Kumar, A. S., Pustet, M., Heimann, S., Buschow, R., Wittler, L., Timmermann, B., Meissner, A. and Herrmann, B. G. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* 370(6522).
17. Veenvliet, J. V. and Herrmann, B. G. (2021). Modeling mammalian trunk development in a dish. *Dev Biol.* 2021 474:5-15.

4 DISCUSSION

Embryonic development is a highly orchestrated process involving many cellular differentiation events that lead to the formation of a complex organism. During these processes, cells move, acquire specific functional states, and organize into complex tissues and organs to execute tightly controlled developmental programs. Studying the molecular and cellular events shaping the embryo has been the focus of developmental biologists for centuries. However, the development of the mammalian embryo *in utero* has been a barrier to acquiring a complete picture of the dynamics occurring during embryogenesis. To overcome this limitation, in the last ten years, many *in vitro* models of embryogenesis have been developed from mESCs, ranging from pre- and peri- [149, 154, 158] to post-implantation development [161-163, 169, 170, 172, 173]. Gastruloids represent the most advanced stem cell-based model in terms of developmental progression since their cells self-organize along the three main body axes and form the three embryonic germ layers [161-163, 170]. Despite these significant advances, gastruloids lack proper tissue morphogenesis and therefore are not suitable for studying the molecular and cellular programs driving the formation and interactions of functional morphological structures, such as somite or neural tube formation.

My doctoral research project focused on optimizing and characterizing a gastruloid-based model able to recapitulate the gastrulation-to-early organogenesis transition both at the molecular and morphological level. To do so, I aimed to provide a culture condition that could more faithfully recapitulate the one present *in vivo*, where the embryo develops surrounded by the basement membrane secreted by the extraembryonic tissues. Supplementing the gastruloids culture medium with a low percentage of an ECM surrogate (Matrigel) in the last 24h of development unlocks the morphogenetic potential of gastruloids. It results in the generation of morphological structures reminiscent of the ones present in the developing trunk region. The resulting structures, that we termed Trunk-Like Structures (TLS), display a high degree of trunk tissue organization, including the simultaneous development of a neural tube, somites, and a gut tube, therefore representing a fundamental improvement as compared to the gastruloid model.

4.1 TLS developmental trajectories: emergence of rare sub-populations

The molecular and cellular investigation of TLS revealed the emergence of cells with various identities and developmental potential, including the presence of somitic, neural, and endodermal cells representing the three germ layers in the trunk post-occipital region. In TLS, most cells are organized in a continuum of states that recapitulate the spatiotemporal progression of a developing embryo, with NMPs differentiating into both the neural and the paraxial mesoderm domains. NMPs display the classical gene expression signature observed *in vivo* and are characterized by the highest differentiation potential. This evidence, combined with their gradual disappearance from 96 to 120h of development, indicates they represent the major progenitor source necessary for TLS development. In accordance with this hypothesis, two independent algorithms highlighted TLS-NMPs as the root of the differentiation trajectories leading to neural tube and somitic cells [9, 174, 175].

The somitic cells in TLS are highly diverse and are characterized by distinct transcriptional identities reminiscent of DV and AP patterning events. Interestingly, small subpopulations of migratory limb muscle precursors (*Lbx1*$^{+}$*/Met*$^{+}$) and synedotome cells (*Scx*$^{+}$) emerge in the last day of TLS development, indicating a high degree of somitic maturation [176, 177]. In contrast, despite showing cellular diversification at the transcriptional level, the neural tube domain does not display the characteristic DV patterning observed *in vivo* [75, 76]. This observation is in line with the known role of SHH secreted from the notochord in the induction of the neural tube DV patterning [75]. In fact TLS, like gastruloids, do not develop a notochord structure and are therefore missing an essential signaling center for neural tube and other tissues patterning. In a recent preprint we have shown the emergence of cells with a notochord-like identity in gastruloids cultured in hypoxic conditions, paving the way to the generation of even more complex gastruloid-based *in vitro* models [178].

Another interesting cellular sub-population emerging in TLS is the endothelial cluster. Endothelial cells will contribute to the vasculature in the developing embryo and can arise from MEOX1^{+} somitic cells. Surprisingly a smaller group of endothelial cells in TLS co-express *Hopx*, *Meox1*, and *Id2*, a gene signature observed in endotome cells that eventually induce hematopoietic stem cells [179].

Finally, a group of PGCLCs gets specified from a population of PS-like/T^{+} precursor cells around 72h of TLS development. This small population, characterized by the expression of key PGC-specific marker genes, is then found close to the gut-tube-like structure, indicating a putative migratory behavior. Despite these *in vivo* like features and their molecular identity,

further work is needed to unveil these cells' exact origin and their developmental potential in TLS.

The observed complexity and maturity of TLS cells and the emergence of rare sub-populations observed in the natural embryo, including putative muscle migratory and hematopoietic precursors, indicate these structures' potential to further in the post-gastrulation/early organogenesis stages. Moreover, decoding the quantitative developmental principles leading to the emergence of these rare cell types *in vitro* will be of great interest in current and future attempts to generate pure cell populations in the context of regenerative medicine and personalized therapies.

4.2 Somitogenesis in TLS

TLS develop somites in response to Matrigel supplementation in the culture medium. Detailed molecular characterization of these morphological structures highlighted their developmental relevance, including the presence of an AP polarity within each somite, with *Uncx* expression restricted to the posterior portion. Moreover, cells forming somites in TLS display apicobasal polarity typical of epithelial tissues, with F-actin and NCAD accumulation at the apical side. In contrast, gastruloids do not show similar expression patterns and organized structural proteins deposition. These molecular and morphological features are crucial for the downstream specification of the somitic DV compartments *in vivo* and, therefore, a fundamental advance observed in TLS compared to gastruloids.

Time-resolved scRNA-seq analysis allowed the profiling of the temporal dynamics leading to somite specification during TLS development, revealing the step-wise and embryo-like differentiation process culminating in the generation of mature somitic subtypes, including dermomyotome and sclerotome. Interestingly, somite segmentation in TLS followed the "*segmentation clock*" model, with the temporally and controlled deposition of a new somite every ~2h. This evidence indicates that the appearance of structures reminiscent of somites in TLS is not the result of trans-differentiation events from mESCs, but that somites emerge in response to an oscillating molecular gradient in the PSM region at an embryo-like pace. A recent work has confirmed the presence of an active oscillator in the gastruloids PSM, indicating that the developmental potential to organize cells into somites is already present in gastruloids but that the ECM is essential for the proper morphogenetic outcome [170].

TLS, therefore, represents a powerful tool to study mammalian somitogenesis from the molecular and mechanical points of view. Recent studies have shown how mechanical models could explain somitogenesis in chicken and fish compared to the chemical gradient model implicit in the "*segmentation clock*" [180-182]. These studies have been possible thanks to the accessibility of the chicken and fish developmental progressions that allowed continuous and uninterrupted observation and data collection of the process. Given their accessibility, TLS represent a perfect playground to test such models in a mammalian context, in order to build more accurate mechano-chemical predictions to explain the complex molecular and structural interactions necessary for somitogenesis.

4.3 TLS vs Gastruloids: beyond enhanced morphology

The appearance of *in vivo*-like tissue morphogenesis in TLS represents a significant advancement to the previously reported gastruloid protocol. Nevertheless, the two models display a surprisingly comparable lineage and cell-type composition, with gastruloids themselves developing somitic, neural, endoderm, PGCLCs, and endothelial cells without the presence of tissue geometries [170]. Despite the global cellular similarities observed in TLS compared to gastruloids, comparative scRNA-seq also revealed important differences. Interestingly, mature $Sox9^{+}/Pax1^{+}$ sclerotomal cells are under-represented in gastruloids, favoring less developed somitic subtypes. Moreover, hypaxial muscle migratory precursor and endothelial-derived putative hematopoietic progenitors are virtually absent in gastruloids, indicating a lower degree of cellular maturation. The expression of later (more posterior) *Hox* genes in TLS also suggests their development into a more advanced trunk stage than gastruloids. Therefore, proper tissue morphogenesis seems to be dispensable for major cell specification events during gastrulation stages but instrumental for the specification of more mature cellular identities. Further comparative studies between gastruloids and TLS will be of great relevance since they will dissect the direct effects of ECM and tissue geometry on lineage specification events.

4.4 TLS^{C} and TLS^{CL}: modulating developmental phenotypes

Given the accessible nature of *in vitro* models of embryogenesis, the opportunity to efficiently modulate the culture conditions where the structure is developing represents one of their major advantages compared to the natural embryo. Exposure to exogenous and dominant WNT activation concomitantly with ECM embedding in TLS^{C} and TLS^{CL} results in the overproduction and accumulation of hyper-epithelialized somites in the most anterior domain of the structures at the expenses of the neural domain, a morphological phenotype never observed in mammalian embryos. Somites in TLS^{C} and TLS^{CL} do not organize into rows flanking the neural tube, like in TLS, but instead, crawl like a "*bunch of grapes*." Interestingly, a similar phenomenon has been observed if transplanting quail PS cells treated with a BMP antagonist (NOGGIN) into the extraembryonic tissue of a chick embryo [183]. This study revealed the possibility of generating somites in the absence of an active molecular oscillator, resulting in "*somites without a clock*" [183]. Surprisingly, the emergence of somitic bursts in TLS^{C} and TLS^{CL} does follow a clock-like rhythm in the paraxial region of the structure, even though the segmentation of clock-free somites in the most anterior domain is also observed. These experiments indicate an intrinsic developmental plasticity of early embryonic cells that can lead to the appearance of similar morphological structures and molecular identities following (most likely) distinct developmental routes. The molecular and mechanical dependencies of these phenomena in TLS^{C} and TLS^{CL} remain to be elucidated.

Despite their morphological similarities, TLS^{C} and TLS^{CL} show molecular differences driven by the additional inhibition of BMPs in the TLS^{CL} condition. In fact, IM and LPM marker genes are downregulated in this condition compared to both conventional TLS and TLS^{C}, in line with the known role of BMP in the mediolateral patterning of the nascent mesoderm [177]. Finally, the comparison between developmental trajectories (as inferred by RNA velocity) revealed that TLS^{CL}-NMPs are highly biased toward a mesodermal identity compared to TLS-NMPs, with little to no contribution to the neural lineage [174].

These proof-of-concept experiments have shown the great potential of applying chemical modulations to interfere with developmental trajectories *in vitro*. Similar experiments have also been recently implemented in gastruloids to influence developmental outcomes [165, 170]. This line of research represents a powerful platform to disentangle the intricate relationships between signaling molecules and developmental phenotypes in morphologically relevant cell systems in a time and dose-controlled fashion.

4.5 Genetic perturbations in TLS

TLS generated from mESCs lacking the PSM-specific TF TBX6 do not undergo somitogenesis and produce ectopic neural tubes. This morphological phenotype mirrors the one observed in *Tbx6*-KO embryos, indicating a shared defective mechanism [60]. The similarity with the embryonic phenotype is also highlighted by the molecular signature of cells contributing to the ectopic tubes in TLS, which display a neural-like identity downregulating PSM and somitic marker genes [61]. This experiment has shown that TLS represents a suitable technology for genetic ablation studies finalized to the investigation of morphological and molecular phenotypes in the mid-gestation embryo.

Recent advances in high-throughput perturbation technologies such as CRISPR screens are rapidly increasing our understanding of genetic dependencies in specific biological contexts, including developmental transitions [184-187]. Applying such screens in adult organoid systems has shown its potential in identifying genes responsible for tissue integrity and cell homeostasis [188, 189]. Their implementation in the context of *in vitro* models of embryogenesis and specifically in TLS development could elucidate the function of unknown regulators involved in early developmental transitions and tissue morphogenesis. Moreover, the unique opportunity to induce the genetic perturbations in the pool of mESCs before or after their aggregation into gastruloids/TLS could distinguish between early and late genetic dependencies necessary for lineage specification events.

4.6 The role of ECM in TLS tissue morphogenesis

The ECM is a crucial component for the induction of embryonic tissue morphologies. The morphogenetic potential of gastruloids is unlocked by using an ECM surrogate (Matrigel) in the culture medium, resulting in the generation of morphological structures reminiscent of the ones present in the developing trunk region, including somites and a neural tube. Comparing gastruloids and TLS revealed an intricate relationship between ECM, tissue geometry, and specific transcriptional responses. In fact, Matrigel embedding alone can induce the upregulation of various cell adhesion molecules in TLS, including proto(cadherins), ephrins and integrins. These results indicate that the ECM contributes mechanically and structurally to regulate tissue shapes and triggers transcriptional responses that confer robustness to the morphogenetic process. In fact, integrins, through their binding to fibronectin present in the

basement membrane, have been shown to play a critical role in somite epithelialization and neural tube closure *in vivo* [128, 129, 133]. Further investigation into the role of fibronectin in TLS revealed that this protein accumulates at the cell-matrix interface and surrounds the edges of epithelialized somites and the folded neural tube. Interestingly, fibronectin is equally expressed in gastruloids and TLS at the transcriptional level, suggesting that the accumulation of fibronectin in TLS is not simply a result of differential cellular abundance. I hypothesize that, in TLS, the ECM can retain a high concentration of cell-secreted fibronectin in the proximity of the cell surfaces, allowing its binding to integrins and the consequent cellular responses. In contrast, in gastruloids, fibronectin diffuses in the culture medium and never reaches the critical local concentration necessary for the induction of morphogenesis. Further investigations implying the use of genetic mutants or neutralizing antibodies will inform on the exact mechanism by which fibronectin triggers morphological changes in TLS.

Matrigel is a heterogeneous mixture of soluble basement membrane proteins isolated from cell lines from the Engelbreth-Holm-Swarm tumor [190]. It mainly consists of laminin and type IV collagen, but traces of other small proteins and growth factors are also present. Moreover, batch-to-batch variability does not ensure a stable and controlled composition given its cellular origin. For these reasons, pinpointing the exact ECM concentration and molecular composition to induce TLS morphogenesis remains challenging, and our attempt to supplement the different Matrigel components in isolation was inconclusive. A recent study has reported the emergence of somite-like structures in gastruloids embedded in 10% Matrigel and that the efficiency with which these sub-structures emerge is inversely correlating with the Matrigel concentration applied [170]. In contrast, in TLS, 5% matrigel is sufficient to induce embryo-like morphogenesis, indicating that the distinct architectures achieved in the two culture systems could reflect a concentration "*sweet spot*" necessary to reach adequate mechanochemical constraints [173]. In this context, 3D synthetic modular matrices, such as polymeric hydrogels supplemented with chemical components could help elucidate the precise mechanical and chemical inputs needed for TLS tissue morphogenesis [191-195]. This approach will also enable fine-tuning of biochemical (e.g., morphogen gradients) and biophysical (e.g., stiffness) parameters in isolation, contributing to the complete picture of how chemical and mechanical cues interact during embryonic development [196].

Finally, the ability to measure cell adhesion molecules' expression in individual cells via scRNA-seq allowed the curation of a cell type-specific catalog of ECM receptors expression in TLS and gastruloids. These atlases will represent handy tools to further our understanding of

how specific cell adhesion molecules contribute to proper tissue morphogenesis in a tissue-specific fashion.

4.7 Variability and reproducibility in TLS

Embryonic development is a highly robust process guaranteed by a multitude of tightly controlled layers of cellular and molecular regulation in place during gestation. TLS display a certain degree of morphological variation, developing somites in roughly half of the analyzed structures (50% segmentation efficiency). Segmented TLS are also partially heterogeneous in how somites organize; they can flank one or both sides of the central neural tube (uni- vs. bi-lateral somite formation), with these two events equally likely to occur. These observations highlight an uncoupling between final morphogenetic outcomes and early transcriptional signature, reinforcing the idea that the competence to acquire morphologically complex structures is not (uniquely) dependent on the activity of genetic programs [196].

Where does this variability come from? How can seemingly homogenous structures in the early stages of TLS development result in diverse developmental outcomes? What are the molecular and structural drivers for uni- and bi-lateral somite formation? These questions remain open and probably have multifaceted explanations. Nevertheless, the TLS platform intrinsic variability offers a new lens to investigate these complex interactions, with variation representing a powerful source of information. This variation can, in fact, be leveraged to learn aspects of developmental regulations that cannot be easily studied *in vivo*, such as the establishment of a mediolateral symmetry in the embryonic trunk.

Despite variation can be a valuable tool in stem cell biology, building a reproducible system in the context of *in vitro* embryogenesis and organogenesis is of crucial importance, especially in biomedical research. TLS display a high degree of molecular reproducibility, as demonstrated by the single-structure transcriptional profiling by Nanostring. The formation of a neural tube is also observed in all structures, while the gut tube is still an under-represented event. Moreover, a detailed morphometric analysis revealed overall robustness in TLS domains area, perimeter, diameter, and volume.

An essential aspect of the innovation of new technologies is the availability of a detailed protocol that would allow researchers worldwide to adopt and implement them efficiently. Therefore, the publication of a step-by-step protocol to generate TLS with the same efficiency

observed in our laboratories was a much-needed addendum to the original publication. In this work, I detailed all critical steps and complications present in the procedure to allow the ready implementation of our technology. Interestingly, a recent preprint has reported the application of TLS to study the maintenance and exit from naïve pluripotency, indicating successful and efficient protocol dissemination [197].

4.8 Towards human *in vitro* models of embryogenesis

Mouse *in vitro* models of embryogenesis are now widely used interdisciplinary to investigate many layers of regulations during the early stages of embryonic development. However, the implementation of similar platforms using human pluripotent stem cells will be critical for translational studies, especially in light of the inter-species developmental differences between mice and humans [145, 198]. Moreover, ethical concerns have limited accessibility to early human embryo material and constrained human embryo culture in a handful of countries under the so-called "*14-days rule*", which does not allow embryo cultures longer than two weeks. These regulations have enormously hindered our knowledge about human early developmental transitions. The recent implementation of human *in vitro* models of embryogenesis faithfully recapitulating critical stages of pre- [199-201] and post-implantation [202, 203] development overcomes these ethical limitations and, for the first time, offers a robust and high-throughput approach to understand early human development. Moreover, the ability to directly derive human induced pluripotent stem cells (hiPSCs) from patient material via cellular reprogramming may shed light on congenital and developmental diseases' molecular and cellular basis.

Concerning post-implantation development, a human gastruloid model has been recently optimized, but, as for the mouse equivalent, proper tissue morphogenesis was not observed [202]. In this protocol, no ECM component was used, therefore leaving an open route to try implementing a human TLS system able to recapitulate early human development at both the molecular and morphological levels.

4.9 TLS as a platform to reduce and replace animal experimentation

With the technical limitations in place in using human embryos for scientific purposes, researchers heavily rely on animal models to understand the principles governing mammalian development. Despite animal research being central to progressing human understanding of biological processes in basic and applied research, we should start entering an era of more sustainable research planning. To this end, the 3Rs principles of "*Replace, Reduce and Refine*" have been proposed to ensure better conditions for animals used in research [204]. Roughly 10 million animals are used each year for experimentations and testing in the European Union only (~4 million for basic research and ~2 million for human disease modeling). This amount includes ~1.6 million mice generated to create and maintain transgenic lines in Germany and UK. Moreover, the recent introduction of more strict guidelines for developmental toxicity studies by the REACH and OECD will require an increase in the number of animals needed, with REACH only requiring more than 50 million vertebrates. *In vitro* models of embryogenesis could represent a valid and complementary approach to study mammalian embryonic development. In fact, it has been estimated by the AnimalTestInfo that from 2013 to 2018, in Germany only, the implementation of these alternative models in mouse studies of early development and organogenesis resulted in the reduction of animals used (at least 103.200 mice). These numbers will increase with the recent development of more reproducible and faithful *in vitro* models. Moreover, the implementation of human models can provide more accurate and human-specific responses to genetic perturbations and toxin exposures compared to mice and drastically reduce the high costs associated with animal research, allowing more laboratories worldwide to engage in developmental studies.

In order to be suitable for this conversion, *in vitro* models have to faithfully recapitulate the developmental transitions occurring in the embryo at the molecular and morphological levels. TLS cellular states are highly similar to their embryonic counterpart, as demonstrated by time-resolved comparative scRNA-seq analysis [205]. Moreover, the emergence of mature cell types from progenitors follows an embryo-like tempo, indicating a high degree of developmental synchronization with the *in vivo* progression. The emergence of a development phenotype that mirrors the natural embryo response to a genetic perturbation (*Tbx6*) holds promises for the application of TLS in mutational studies.

Despite these essential commonalities between TLS and the *in vivo* embryo, many limitations to this system remain; TLS do not develop occipital structures, a heart tube, and

other anterior domains. Moreover, the absence of extra-embryonic tissues represents a physiological barrier to understanding the embryo-to-foster interactions during embryogenesis. A recent study has shown how some newly developed human cell systems modeling the human blastocyst stages, despite being morphologically equivalent to the embryo, display molecular and cellular divergence [206]. Therefore, *in vitro* models of embryogenesis, despite being a valid alternative to reduce the use of mice in basic and biomedical research, still represent a "*reductionist approach*" compared to the natural embryo, requiring additional *in vivo* complexity needed to replace animal experimentation completely.

4.10 TLS applications in multidisciplinary contexts

TLS represent a powerful and novel technology with the potential to be implemented interdisciplinary in various contexts (**Figure 7**).

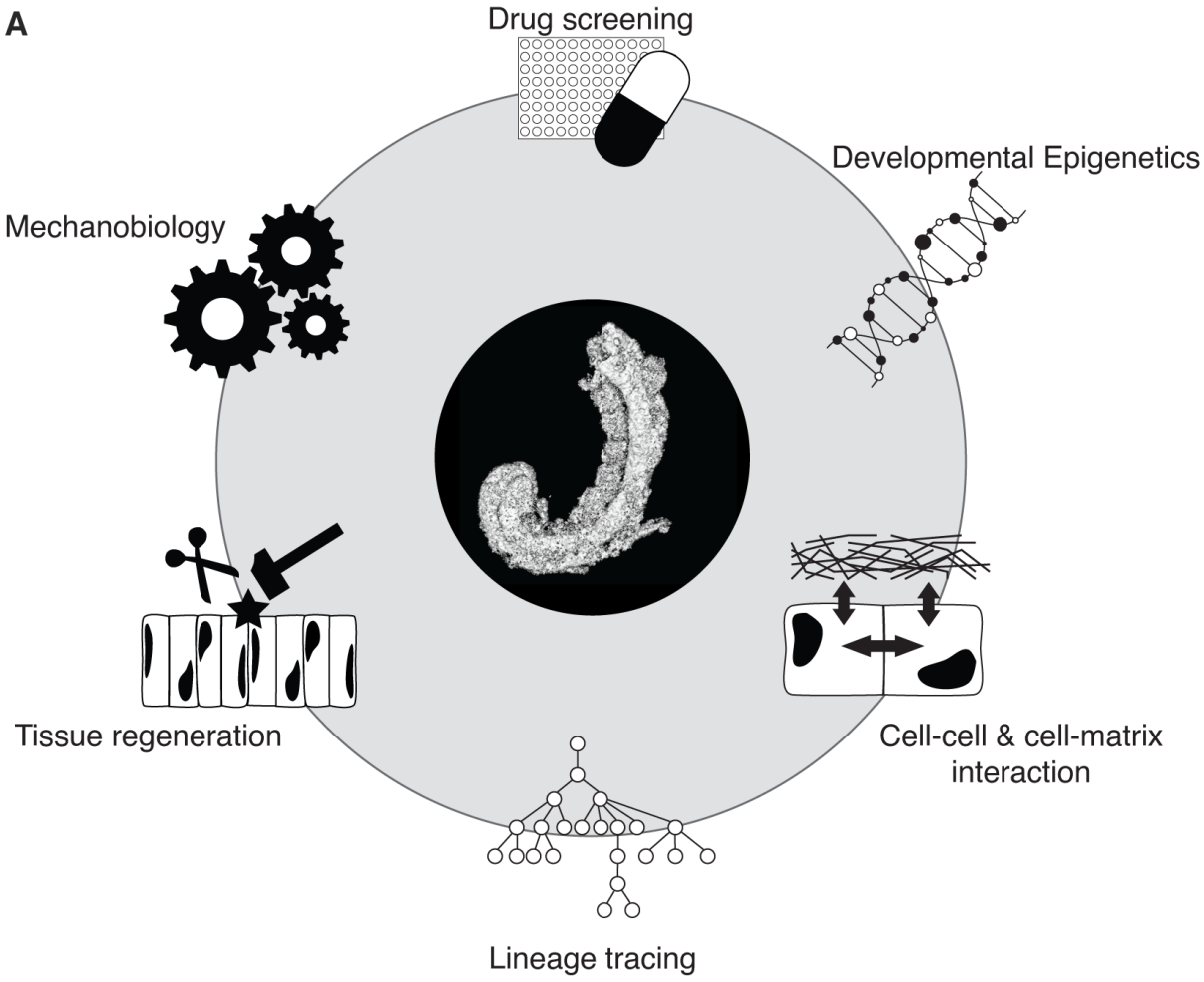

Figure 7. Future perspectives for TLS

(A) Schematic representation of future applications for TLS to study different aspects of developmental processes. TLS represent a potent tool to perform high-throughput investigations of the effect of chemical molecules in the context of early development. Given its morphogenetic potential, the model also provides a unique playground to study mechanobiology, tissue regeneration and cell-matrix interactions in early cell specification and tissue patterning. The possibility of efficiently engineering mESCs and manipulating the culture conditions are also precious features for developing lineage tracing cell systems and studying the epigenetic barriers during exit from pluripotency.

The comparison between gastruloids, TLS, and the natural embryo could inform the structural and molecular principles of how cell-cell and cell-matrix interactions drive tissue

morphogenesis and spatial domain organization during development. Moreover, the accessibility of TLS combined with their morphogenetic potential offers a unique opportunity to investigate molecular, cellular, and tissue responses to mechanical stimuli in both the context of tissue homeostasis and regeneration. Studies of this kind will contribute to understanding how different forces (e.g., tensile, compressive, shear) shape tissue morphologies and the roles these forces play in lineage allocation.

TLS also represent a fast and easily applicable alternative for lineage tracing studies *in vivo* since the engineering of stable reporter mouse lines is not time and cost-effective. The generation of mESCs reporter lines is becoming more and more accessible thanks to the development of efficient genome editing techniques. Moreover, a vast catalog of tissue-specific report mESCs already exist, and these lines could be readily implemented in the TLS protocol. Recent advances in single-cell CRISPR-based molecular recording technologies and the benchmarking of mESCs recorders also offer a more sophisticated approach that can be combined with TLS development to quantitatively measure cellular specification events to uncover genetic drivers of developmental transitions [207, 208].

Several layers of fine-tuned epigenetic regulations canalize the precise cellular transitions during embryogenesis, but disentangling the specific regulators' functions for cell types acquisition is challenging due to the early lethal phenotypes observed in mutant embryos [209, 210]. TLS could represent a valid and easy-to-manipulate alternative to study their functions during the early stages of embryonic development without the need to induce genetic perturbations. In fact, the possibility to apply chemical compound libraries that have been widely adopted in other contexts to selectively inhibit epigenetic regulators functions in high-throughput screens in TLS could provide a rapid and high-content approach to unveil their precise contributions to cell state specification events [211-214].

Finally, TLS and *in vitro* models of embryogenesis could be used for high-throughput screens to probe the teratogenic effects of drugs and substances in a cellular system that preserves an embryo-like tissue organization compared to the current models used in pre-clinical experimentations. A proof of principle study using gastruloids has demonstrated the suitability of these models in identifying harmful molecules that would affect developmental progression [215].

Given its modularity, tractability, scalability, and reproducibility, I firmly believe Trunk-Like Structures will be a powerful toolkit widely adopted by the research community to study normal and aberrant developmental principles *in vitro*.

REFERENCES

1. Zernicka-Goetz M, Morris SA, Bruce AW: Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* 2009, 10:467-477.
2. Hupalowska A, Jedrusik A, Zhu M, Bedford MT, Glover DM, Zernicka-Goetz M: CARM1 and Paraspeckles Regulate Pre-implantation Mouse Embryo Development. *Cell* 2018, 175:1902-1916 e1913.
3. Wang J, Wang L, Feng G, Wang Y, Li Y, Li X, Liu C, Jiao G, Huang C, Shi J, et al: Asymmetric Expression of LincGET Biases Cell Fate in Two-Cell Mouse Embryos. *Cell* 2018, 175:1887-1901 e1818.
4. Zhu M, Cornwall-Scoones J, Wang P, Handford CE, Na J, Thomson M, Zernicka-Goetz M: Developmental clock and mechanism of de novo polarization of the mouse embryo. *Science* 2020, 370.
5. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R: Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003, 17:126-140.
6. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S: The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003, 113:631-642.
7. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A: Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998, 95:379-391.
8. Zhang J, Tam WL, Tong GQ, Wu Q, Chan HY, Soh BS, Lou Y, Yang J, Ma Y, Chai L, et al: Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* 2006, 8:1114-1123.
9. Nowotschin S, Setty M, Kuo YY, Liu V, Garg V, Sharma R, Simon CS, Saiz N, Gardner R, Boutet SC, et al: The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* 2019, 569:361-367.
10. Koutsourakis M, Langeveld A, Patient R, Beddington R, Grosveld F: The transcription factor GATA6 is essential for early extraembryonic development. *Development* 1999, 126:723-732.
11. Morrisey EE, Tang Z, Sigrist K, Lu MM, Jiang F, Ip HS, Parmacek MS: GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* 1998, 12:3579-3590.
12. Chazaud C, Yamanaka Y, Pawson T, Rossant J: Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell* 2006, 10:615-624.
13. Bardot ES, Hadjantonakis AK: Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate. *Mech Dev* 2020, 163:103617.
14. Arnold SJ, Robertson EJ: Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol* 2009, 10:91-103.

15. Lu CC, Robertson EJ: Multiple roles for Nodal in the epiblast of the mouse embryo in the establishment of anterior-posterior patterning. *Dev Biol* 2004, 273:149-159.
16. Yamamoto M, Beppu H, Takaoka K, Meno C, Li E, Miyazono K, Hamada H: Antagonism between Smad1 and Smad2 signaling determines the site of distal visceral endoderm formation in the mouse embryo. *J Cell Biol* 2009, 184:323-334.
17. Hoshino H, Shioi G, Aizawa S: AVE protein expression and visceral endoderm cell behavior during anterior-posterior axis formation in mouse embryos: Asymmetry in OTX2 and DKK1 expression. *Dev Biol* 2015, 402:175-191.
18. Kimura-Yoshida C, Nakano H, Okamura D, Nakao K, Yonemura S, Belo JA, Aizawa S, Matsui Y, Matsuo I: Canonical Wnt signaling and its antagonist regulate anterior-posterior axis polarization by guiding cell migration in mouse visceral endoderm. *Dev Cell* 2005, 9:639-650.
19. Kemp C, Willems E, Abdo S, Lambiv L, Leyns L: Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Dev Dyn* 2005, 233:1064-1075.
20. Mohamed OA, Clarke HJ, Dufort D: Beta-catenin signaling marks the prospective site of primitive streak formation in the mouse embryo. *Dev Dyn* 2004, 231:416-424.
21. Sun X, Meyers EN, Lewandoski M, Martin GR: Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev* 1999, 13:1834-1846.
22. Winnier G, Blessing M, Labosky PA, Hogan BL: Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev* 1995, 9:2105-2116.
23. Huelsken J, Vogel R, Brinkmann V, Erdmann B, Birchmeier C, Birchmeier W: Requirement for beta-catenin in anterior-posterior axis formation in mice. *J Cell Biol* 2000, 148:567-578.
24. Liu P, Wakamiya M, Shea MJ, Albrecht U, Behringer RR, Bradley A: Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* 1999, 22:361-365.
25. Arnold SJ, Hofmann UK, Bikoff EK, Robertson EJ: Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* 2008, 135:501-511.
26. Carver EA, Jiang R, Lan Y, Oram KF, Gridley T: The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. *Mol Cell Biol* 2001, 21:8184-8188.
27. Hart AH, Hartley L, Sourris K, Stadler ES, Li R, Stanley EG, Tam PP, Elefanty AG, Robb L: Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. *Development* 2002, 129:3597-3608.
28. Rashbass P, Cooke LA, Herrmann BG, Beddington RS: A cell autonomous function of Brachyury in T/T embryonic stem cell chimaeras. *Nature* 1991, 353:348-351.
29. Lawson KA, Pedersen RA: Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse. *Development* 1987, 101:627-652.
30. Kwon GS, Viotti M, Hadjantonakis AK: The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev Cell* 2008, 15:509-520.

31. Viotti M, Nowotschin S, Hadjantonakis AK: SOX17 links gut endoderm morphogenesis and germ layer segregation. *Nat Cell Biol* 2014, 16:1146-1156.
32. Tam PP, Behringer RR: Mouse gastrulation: the formation of a mammalian body plan. *Mech Dev* 1997, 68:3-25.
33. Acampora D, Mazan S, Lallemand Y, Avantaggiato V, Maury M, Simeone A, Brulet P: Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development* 1995, 121:3279-3290.
34. Ang SL, Conlon RA, Jin O, Rossant J: Positive and negative signals from mesoderm regulate the expression of mouse Otx2 in ectoderm explants. *Development* 1994, 120:2979-2989.
35. Neijts R, Simmini S, Giuliani F, van Rooijen C, Deschamps J: Region-specific regulation of posterior axial elongation during vertebrate embryogenesis. *Dev Dyn* 2014, 243:88-98.
36. Veenvliet JV, Herrmann BG: Modeling mammalian trunk development in a dish. *Dev Biol* 2021, 474:5-15.
37. Henrique D, Abranches E, Verrier L, Storey KG: Neuromesodermal progenitors and the making of the spinal cord. *Development* 2015, 142:2864-2875.
38. Koch F, Scholze M, Wittler L, Schifferl D, Sudheer S, Grote P, Timmermann B, Macura K, Herrmann BG: Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev Cell* 2017, 42:514-526 e517.
39. Wilson V, Olivera-Martinez I, Storey KG: Stem cells, signals and vertebrate body axis extension. *Development* 2009, 136:1591-1604.
40. Nicolas JF, Mathis L, Bonnerot C, Saurin W: Evidence in the mouse for self-renewing stem cells in the formation of a segmented longitudinal structure, the myotome. *Development* 1996, 122:2933-2946.
41. Tzouanacou E, Wegener A, Wymeersch FJ, Wilson V, Nicolas JF: Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev Cell* 2009, 17:365-376.
42. Sambasivan R, Steventon B: Neuromesodermal Progenitors: A Basis for Robust Axial Patterning in Development and Evolution. *Front Cell Dev Biol* 2020, 8:607516.
43. Olivera-Martinez I, Harada H, Halley PA, Storey KG: Loss of FGF-dependent mesoderm identity and rise of endogenous retinoid signalling determine cessation of body axis elongation. *PLoS Biol* 2012, 10:e1001415.
44. Diez Del Corral R, Morales AV: The Multiple Roles of FGF Signaling in the Developing Spinal Cord. *Front Cell Dev Biol* 2017, 5:58.
45. Gouti M, Delile J, Stamataki D, Wymeersch FJ, Huang Y, Kleinjung J, Wilson V, Briscoe J: A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev Cell* 2017, 41:243-261 e247.
46. Naiche LA, Holder N, Lewandoski M: FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proc Natl Acad Sci U S A* 2011, 108:4018-4023.
47. Abu-Abed S, Dolle P, Metzger D, Wood C, MacLean G, Chambon P, Petkovich M: Developing with lethal RA levels: genetic ablation of Rarg can restore the viability of mice lacking Cyp26a1. *Development* 2003, 130:1449-1459.
48. Martin BL, Kimelman D: Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev Cell* 2012, 22:223-232.

49. Javali A, Misra A, Leonavicius K, Acharyya D, Vyas B, Sambasivan R: Co-expression of Tbx6 and Sox2 identifies a novel transient neuromesoderm progenitor cell state. *Development* 2017, 144:4522-4529.
50. Saga Y: The mechanism of somite formation in mice. *Curr Opin Genet Dev* 2012, 22:331-338.
51. Cooke J, Zeeman EC: A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J Theor Biol* 1976, 58:455-476.
52. Pourquie O: Vertebrate somitogenesis. *Annu Rev Cell Dev Biol* 2001, 17:311-350.
53. Niwa Y, Masamizu Y, Liu T, Nakayama R, Deng CX, Kageyama R: The initiation and propagation of Hes7 oscillation are cooperatively regulated by Fgf and notch signaling in the somite segmentation clock. *Dev Cell* 2007, 13:298-304.
54. Bessho Y, Hirata H, Masamizu Y, Kageyama R: Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. *Genes Dev* 2003, 17:1451-1456.
55. Dale JK, Maroto M, Dequeant ML, Malapert P, McGrew M, Pourquie O: Periodic notch inhibition by lunatic fringe underlies the chick segmentation clock. *Nature* 2003, 421:275-278.
56. Takahashi Y, Koizumi K, Takagi A, Kitajima S, Inoue T, Koseki H, Saga Y: Mesp2 initiates somite segmentation through the Notch signalling pathway. *Nat Genet* 2000, 25:390-396.
57. Takahashi J, Ohbayashi A, Oginuma M, Saito D, Mochizuki A, Saga Y, Takada S: Analysis of Ripply1/2-deficient mouse embryos reveals a mechanism underlying the rostro-caudal patterning within a somite. *Dev Biol* 2010, 342:134-145.
58. Galceran J, Farinas I, Depew MJ, Clevers H, Grosschedl R: Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. *Genes Dev* 1999, 13:709-717.
59. Takada S, Stark KL, Shea MJ, Vassileva G, McMahon JA, McMahon AP: Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev* 1994, 8:174-189.
60. Chapman DL, Papaioannou VE: Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* 1998, 391:695-697.
61. Concepcion D, Washkowitz AJ, DeSantis A, Ogea P, Yang JI, Douglas NC, Papaioannou VE: Cell lineage of timed cohorts of Tbx6-expressing cells in wild-type and Tbx6 mutant embryos. *Biol Open* 2017, 6:1065-1073.
62. Kraus F, Haenig B, Kispert A: Cloning and expression analysis of the mouse T-box gene Tbx18. *Mech Dev* 2001, 100:83-86.
63. Feller J, Schneider A, Schuster-Gossler K, Gossler A: Noncyclic Notch activity in the presomitic mesoderm demonstrates uncoupling of somite compartmentalization and boundary formation. *Genes Dev* 2008, 22:2166-2171.
64. Bussen M, Petry M, Schuster-Gossler K, Leitges M, Gossler A, Kispert A: The T-box transcription factor Tbx18 maintains the separation of anterior and posterior somite compartments. *Genes Dev* 2004, 18:1209-1221.
65. Deutsch U, Dressler GR, Gruss P: Pax 1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell* 1988, 53:617-625.

66. Fan CM, Tessier-Lavigne M: Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell* 1994, 79:1175-1186.
67. Kuan CY, Tannahill D, Cook GM, Keynes RJ: Somite polarity and segmental patterning of the peripheral nervous system. *Mech Dev* 2004, 121:1055-1068.
68. Keynes RJ, Stern CD: Mechanisms of vertebrate segmentation. *Development* 1988, 103:413-429.
69. Denetclaw WF, Jr., Christ B, Ordahl CP: Location and growth of epaxial myotome precursor cells. *Development* 1997, 124:1601-1610.
70. Reshef R, Maroto M, Lassar AB: Regulation of dorsal somitic cell fates: BMPs and Noggin control the timing and pattern of myogenic regulator expression. *Genes Dev* 1998, 12:290-303.
71. Brent AE, Schweitzer R, Tabin CJ: A somitic compartment of tendon progenitors. *Cell* 2003, 113:235-248.
72. Forlani S, Lawson KA, Deschamps J: Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development* 2003, 130:3807-3819.
73. Perantoni AO, Timofeeva O, Naillat F, Richman C, Pajni-Underwood S, Wilson C, Vainio S, Dove LF, Lewandoski M: Inactivation of FGF8 in early mesoderm reveals an essential role in kidney development. *Development* 2005, 132:3859-3871.
74. Wilde JJ, Petersen JR, Niswander L: Genetic, epigenetic, and environmental contributions to neural tube closure. *Annu Rev Genet* 2014, 48:583-611.
75. Dessaud E, McMahon AP, Briscoe J: Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* 2008, 135:2489-2503.
76. Marti E, Takada R, Bumcrot DA, Sasaki H, McMahon AP: Distribution of Sonic hedgehog peptides in the developing chick and mouse embryo. *Development* 1995, 121:2537-2547.
77. Jessell TM: Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat Rev Genet* 2000, 1:20-29.
78. Le Dreau G, Marti E: Dorsal-ventral patterning of the neural tube: a tale of three signals. *Dev Neurobiol* 2012, 72:1471-1481.
79. Tam PP, Khoo PL, Lewis SL, Bildsoe H, Wong N, Tsang TE, Gad JM, Robb L: Sequential allocation and global pattern of movement of the definitive endoderm in the mouse embryo during gastrulation. *Development* 2007, 134:251-260.
80. Zorn AM, Wells JM: Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* 2009, 25:221-251.
81. McLin VA, Rankin SA, Zorn AM: Repression of Wnt/beta-catenin signaling in the anterior endoderm is essential for liver and pancreas development. *Development* 2007, 134:2207-2217.
82. Wells JM, Melton DA: Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* 2000, 127:1563-1572.
83. Prummel KD, Nieuwenhuize S, Mosimann C: The lateral plate mesoderm. *Development* 2020, 147.

84. Ferretti E, Hadjantonakis AK: Mesoderm specification and diversification: from single cells to emergent tissues. *Curr Opin Cell Biol* 2019, 61:110-116.
85. Saxen L, Sariola H: Early organogenesis of the kidney. *Pediatr Nephrol* 1987, 1:385-392.
86. Barak H, Rosenfelder L, Schultheiss TM, Reshef R: Cell fate specification along the anterior-posterior axis of the intermediate mesoderm. *Dev Dyn* 2005, 232:901-914.
87. Dressler GR, Deutsch U, Chowdhury K, Nornes HO, Gruss P: Pax2, a new murine paired-box-containing gene and its expression in the developing excretory system. *Development* 1990, 109:787-795.
88. Chiquoine AD: The identification, origin, and migration of the primordial germ cells in the mouse embryo. *Anat Rec* 1954, 118:135-146.
89. Ginsburg M, Snow MH, McLaren A: Primordial germ cells in the mouse embryo during gastrulation. *Development* 1990, 110:521-528.
90. Saitou M, Yamaji M: Primordial germ cells in mice. *Cold Spring Harb Perspect Biol* 2012, 4.
91. Aramaki S, Hayashi K, Kurimoto K, Ohta H, Yabuta Y, Iwanari H, Mochizuki Y, Hamakubo T, Kato Y, Shirahige K, Saitou M: A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev Cell* 2013, 27:516-529.
92. Saitou M, Barton SC, Surani MA: A molecular programme for the specification of germ cell fate in mice. *Nature* 2002, 418:293-300.
93. Tam PP, Snow MH: Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. *J Embryol Exp Morphol* 1981, 64:133-147.
94. Mecham RP: Overview of extracellular matrix. *Curr Protoc Cell Biol* 2012, Chapter 10:Unit 10 11.
95. Hynes RO, Naba A: Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol* 2012, 4:a004903.
96. Engel JC, Matthias: An Overview of Extracellular Matrix Structure and Function. In; 2011
97. Mouw JK, Ou G, Weaver VM: Extracellular matrix assembly: a multiscale deconstruction. *Nat Rev Mol Cell Biol* 2014, 15:771-785.
98. Paulsson M: Basement membrane proteins: structure, assembly, and cellular interactions. *Crit Rev Biochem Mol Biol* 1992, 27:93-127.
99. Brodsky B, Persikov AV: Molecular structure of the collagen triple helix. *Adv Protein Chem* 2005, 70:301-339.
100. Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I, Birk DE: Type V collagen controls the initiation of collagen fibril assembly. *J Biol Chem* 2004, 279:53331-53337.
101. O'Leary LE, Fallas JA, Bakota EL, Kang MK, Hartgerink JD: Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nat Chem* 2011, 3:821-828.
102. Wagenseil JE, Mecham RP: Vascular extracellular matrix and arterial mechanics. *Physiol Rev* 2009, 89:957-989.
103. Kozel BA, Mecham, R.P., Rosenbloom, J.: Elastin. In *The Extracellular Matrix: an Overview Biology of Extracellular Matrix.* 2011
104. Singh P, Carraher C, Schwarzbauer JE: Assembly of fibronectin extracellular matrix. *Annu Rev Cell Dev Biol* 2010, 26:397-419.

105. Garcia AJ, Schwarzbauer JE, Boettiger D: Distinct activation states of alpha5beta1 integrin show differential binding to RGD and synergy domains of fibronectin. *Biochemistry* 2002, 41:9063-9069.
106. Schwarzbauer JE, DeSimone DW: Fibronectins, their fibrillogenesis, and in vivo functions. *Cold Spring Harb Perspect Biol* 2011, 3.
107. Ilic D, Kovacic B, Johkura K, Schlaepfer DD, Tomasevic N, Han Q, Kim JB, Howerton K, Baumbusch C, Ogiwara N, et al: FAK promotes organization of fibronectin matrix and fibrillar adhesions. *J Cell Sci* 2004, 117:177-187.
108. Xu J, Mosher, D.: Fibronectin and Other Adhesive Glycoproteins. In; 2011
109. Cui H, Freeman C, Jacobson GA, Small DH: Proteoglycans in the central nervous system: role in development, neural repair, and Alzheimer's disease. *IUBMB Life* 2013, 65:108-120.
110. Simon Davis DA, Parish CR: Heparan sulfate: a ubiquitous glycosaminoglycan with multiple roles in immunity. *Front Immunol* 2013, 4:470.
111. Bandtlow CE, Zimmermann DR: Proteoglycans in the developing brain: new conceptual insights for old proteins. *Physiol Rev* 2000, 80:1267-1290.
112. Knudson CB, Knudson W: Cartilage proteoglycans. *Semin Cell Dev Biol* 2001, 12:69-78.
113. Timpl R, Brown JC: The laminins. *Matrix Biol* 1994, 14:275-281.
114. Aumailley M, Bruckner-Tuderman L, Carter WG, Deutzmann R, Edgar D, Ekblom P, Engel J, Engvall E, Hohenester E, Jones JC, et al: A simplified laminin nomenclature. *Matrix Biol* 2005, 24:326-332.
115. Miner JH, Yurchenco PD: Laminin functions in tissue morphogenesis. *Annu Rev Cell Dev Biol* 2004, 20:255-284.
116. Aumailley M, Krieg T: Laminins: a family of diverse multifunctional molecules of basement membranes. *J Invest Dermatol* 1996, 106:209-214.
117. Yurchenco PD: Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harb Perspect Biol* 2011, 3.
118. Matsuo I, Hiramatsu R: Mechanical perspectives on the anterior-posterior axis polarization of mouse implanted embryos. *Mech Dev* 2017, 144:62-70.
119. Brown NH: Extracellular matrix in development: insights from mechanisms conserved between invertebrates and vertebrates. *Cold Spring Harb Perspect Biol* 2011, 3.
120. Avraamides CJ, Garmy-Susini B, Varner JA: Integrins in angiogenesis and lymphangiogenesis. *Nat Rev Cancer* 2008, 8:604-617.
121. Chal J, Guillot C, Pourquie O: PAPC couples the segmentation clock to somite morphogenesis by regulating N-cadherin-dependent adhesion. *Development* 2017, 144:664-676.
122. Chong SW, Jiang YJ: Off limits--integrins holding boundaries in somitogenesis. *Trends Cell Biol* 2005, 15:453-457.
123. Julich D, Cobb G, Melo AM, McMillen P, Lawton AK, Mochrie SG, Rhoades E, Holley SA: Cross-Scale Integrin Regulation Organizes ECM and Tissue Topology. *Dev Cell* 2015, 34:33-44.
124. Julich D, Mould AP, Koper E, Holley SA: Control of extracellular matrix assembly along tissue boundaries via Integrin and Eph/Ephrin signaling. *Development* 2009, 136:2913-2921.

125. Mole MA, Galea GL, Rolo A, Weberling A, Nychyk O, De Castro SC, Savery D, Fassler R, Ybot-Gonzalez P, Greene NDE, Copp AJ: Integrin-Mediated Focal Anchorage Drives Epithelial Zippering during Mouse Neural Tube Closure. *Dev Cell* 2020, 52:321-334 e326.
126. Porcionatto MA: The extracellular matrix provides directional cues for neuronal migration during cerebellar development. *Braz J Med Biol Res* 2006, 39:313-320.
127. Raz P, Lohmann CH, Turner J, Wang L, Poythress N, Blanchard C, Boyan BD, Schwartz Z: 1alpha,25(OH)2D3 regulation of integrin expression is substrate dependent. *J Biomed Mater Res A* 2004, 71:217-225.
128. George EL, Georges-Labouesse EN, Patel-King RS, Rayburn H, Hynes RO: Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. *Development* 1993, 119:1079-1091.
129. Giros A, Grgur K, Gossler A, Costell M: alpha5beta1 integrin-mediated adhesion to fibronectin is required for axis elongation and somitogenesis in mice. *PLoS One* 2011, 6:e22002.
130. Goh KL, Yang JT, Hynes RO: Mesodermal defects and cranial neural crest apoptosis in alpha5 integrin-null embryos. *Development* 1997, 124:4309-4319.
131. Yang JT, Bader BL, Kreidberg JA, Ullman-Cullere M, Trevithick JE, Hynes RO: Overlapping and independent functions of fibronectin receptor integrins in early mesodermal development. *Dev Biol* 1999, 215:264-277.
132. Yang JT, Rayburn H, Hynes RO: Embryonic mesodermal defects in alpha 5 integrin-deficient mice. *Development* 1993, 119:1093-1105.
133. Takahashi S, Leiss M, Moser M, Ohashi T, Kitao T, Heckmann D, Pfeifer A, Kessler H, Takagi J, Erickson HP, Fassler R: The RGD motif in fibronectin is essential for development but dispensable for fibril assembly. *J Cell Biol* 2007, 178:167-178.
134. De Arcangelis A, Mark M, Kreidberg J, Sorokin L, Georges-Labouesse E: Synergistic activities of alpha3 and alpha6 integrins are required during apical ectodermal ridge formation and organogenesis in the mouse. *Development* 1999, 126:3957-3968.
135. Miner JH, Cunningham J, Sanes JR: Roles for laminin in embryogenesis: exencephaly, syndactyly, and placentopathy in mice lacking the laminin alpha5 chain. *J Cell Biol* 1998, 143:1713-1723.
136. Evans MJ, Kaufman MH: Establishment in culture of pluripotential cells from mouse embryos. *Nature* 1981, 292:154-156.
137. Martin GR: Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 1981, 78:7634-7638.
138. Martin GR, Evans MJ: Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A* 1975, 72:1441-1445.
139. Pierce GB, Jr., Verney EL: An in vitro and in vivo study of differentiation in teratocarcinomas. *Cancer* 1961, 14:1017-1029.
140. Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R: The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* 1985, 87:27-45.

141. Murry CE, Keller G: Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* 2008, 132:661-680.
142. Baillie-Benson P, Moris N, Martinez Arias A: Pluripotent stem cell models of early mammalian development. *Curr Opin Cell Biol* 2020, 66:89-96.
143. Rivron N, Fu J: It takes a village to form embryo models. *Stem Cell Reports* 2021, 16:1011-1013.
144. Shahbazi MN, Siggia ED, Zernicka-Goetz M: Self-organization of stem cells into embryos: A window on early mammalian development. *Science* 2019, 364:948-951.
145. Shahbazi MN, Zernicka-Goetz M: Deconstructing and reconstructing the mouse and human early embryo. *Nat Cell Biol* 2018, 20:878-887.
146. Sozen B, Cornwall-Scoones J, Zernicka-Goetz M: The dynamics of morphogenesis in stem cell-based embryology: Novel insights for symmetry breaking. *Dev Biol* 2021, 474:82-90.
147. Marikawa Y, Tamashiro DA, Fujita TC, Alarcon VB: Aggregated P19 mouse embryonal carcinoma cells as a simple in vitro model to study the molecular regulations of mesoderm formation and axial elongation morphogenesis. *Genesis* 2009, 47:93-106.
148. ten Berge D, Koole W, Fuerer C, Fish M, Eroglu E, Nusse R: Wnt signaling mediates self-organization and axis formation in embryoid bodies. *Cell Stem Cell* 2008, 3:508-518.
149. Rivron NC, Frias-Aldeguer J, Vrij EJ, Boisset JC, Korving J, Vivie J, Truckenmuller RK, van Oudenaarden A, van Blitterswijk CA, Geijsen N: Blastocyst-like structures generated solely from stem cells. *Nature* 2018, 557:106-111.
150. Tanaka S, Kunath T, Hadjantonakis AK, Nagy A, Rossant J: Promotion of trophoblast stem cell proliferation by FGF4. *Science* 1998, 282:2072-2075.
151. Sozen B, Cox AL, De Jonghe J, Bao M, Hollfelder F, Glover DM, Zernicka-Goetz M: Self-Organization of Mouse Stem Cells into an Extended Potential Blastoid. *Dev Cell* 2019, 51:698-712 e698.
152. Yang J, Ryan DJ, Wang W, Tsang JC, Lan G, Masaki H, Gao X, Antunes L, Yu Y, Zhu Z, et al: Establishment of mouse expanded potential stem cells. *Nature* 2017, 550:393-397.
153. Bedzhov I, Zernicka-Goetz M: Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* 2014, 156:1032-1044.
154. Harrison SE, Sozen B, Christodoulou N, Kyprianou C, Zernicka-Goetz M: Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science* 2017, 356.
155. Harrison SE, Sozen B, Zernicka-Goetz M: In vitro generation of mouse polarized embryo-like structures from embryonic and trophoblast stem cells. *Nat Protoc* 2018, 13:1586-1602.
156. Kunath T, Arnaud D, Uy GD, Okamoto I, Chureau C, Yamanaka Y, Heard E, Gardner RL, Avner P, Rossant J: Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development* 2005, 132:1649-1661.
157. Niakan KK, Schrode N, Cho LT, Hadjantonakis AK: Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. *Nat Protoc* 2013, 8:1028-1041.

158. Sozen B, Amadei G, Cox A, Wang R, Na E, Czukiewska S, Chappell L, Voet T, Michel G, Jing N, et al: Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat Cell Biol* 2018, 20:979-989.
159. van den Brink SC, van Oudenaarden A: 3D gastruloids: a novel frontier in stem cell-based in vitro modeling of mammalian gastrulation. *Trends Cell Biol* 2021, 31:747-759.
160. Baillie-Johnson P, van den Brink SC, Balayo T, Turner DA, Martinez Arias A: Generation of Aggregates of Mouse Embryonic Stem Cells that Show Symmetry Breaking, Polarization and Emergent Collective Behaviour In Vitro. *J Vis Exp* 2015.
161. Turner DA, Girgin M, Alonso-Crisostomo L, Trivedi V, Baillie-Johnson P, Glodowski CR, Hayward PC, Collignon J, Gustavsen C, Serup P, et al: Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids. *Development* 2017, 144:3894-3906.
162. van den Brink SC, Baillie-Johnson P, Balayo T, Hadjantonakis AK, Nowotschin S, Turner DA, Martinez Arias A: Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* 2014, 141:4231-4242.
163. Beccari L, Moris N, Girgin M, Turner DA, Baillie-Johnson P, Cossy AC, Lutolf MP, Duboule D, Arias AM: Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* 2018, 562:272-276.
164. Izpisua-Belmonte JC, Falkenstein H, Dolle P, Renucci A, Duboule D: Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. *EMBO J* 1991, 10:2279-2289.
165. Girgin MU, Broguiere N, Mattolini L, Lutolf MP: Gastruloids generated without exogenous Wnt activation develop anterior neural tissues. *Stem Cell Reports* 2021, 16:1143-1155.
166. Girgin MU, Broguiere N, Hoehnel S, Brandenberg N, Mercier B, Arias AM, Lutolf MP: Bioengineered embryoids mimic post-implantation development in vitro. *Nat Commun* 2021, 12:5140.
167. Noémie M. L. P. Bérenger-Currias MM, Esmée Adegeest, Patrick R. van den Berg, Marleen Feliksik, Mazène Hochane, Timon Idema, Sander J. Tans, Stefan Semrau: Early neurulation recapitulated in assemblies of embryonic and extraembryonic cells. *BiorXiv* 2020.
168. Vianello S LM: In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture. *BiorXiv*.
169. Rossi G, Broguiere N, Miyamoto M, Boni A, Guiet R, Girgin M, Kelly RG, Kwon C, Lutolf MP: Capturing Cardiogenesis in Gastruloids. *Cell Stem Cell* 2021, 28:230-240 e236.
170. van den Brink SC, Alemany A, van Batenburg V, Moris N, Blotenburg M, Vivie J, Baillie-Johnson P, Nichols J, Sonnen KF, Martinez Arias A, van Oudenaarden A: Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* 2020, 582:405-409.
171. Xu PF, Borges RM, Fillatre J, de Oliveira-Melo M, Cheng T, Thisse B, Thisse C: Construction of a mammalian embryo model from stem cells organized by a morphogen signalling centre. *Nat Commun* 2021, 12:3277.

172. Bolondi A, Haut L, Gassaloglu SI, Burton P, Kretzmer H, Buschow R, Meissner A, Herrmann BG, Veenvliet JV: Generation of Mouse Pluripotent Stem Cell-derived Trunk-like Structures: An in vitro Model of Post-implantation Embryogenesis. *Bio Protoc* 2021, 11:e4042.
173. Veenvliet JV, Bolondi A, Kretzmer H, Haut L, Scholze-Wittler M, Schifferl D, Koch F, Guignard L, Kumar AS, Pustet M, et al: Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* 2020, 370.
174. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lonnerberg P, Furlan A, et al: RNA velocity of single cells. *Nature* 2018, 560:494-498.
175. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D: Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol* 2019, 37:451-460.
176. Buckingham M, Bajard L, Chang T, Daubas P, Hadchouel J, Meilhac S, Montarras D, Rocancourt D, Relaix F: The formation of skeletal muscle: from somite to limb. *J Anat* 2003, 202:59-68.
177. Chal J, Pourquie O: Making muscle: skeletal myogenesis in vivo and in vitro. *Development* 2017, 144:2104-2122.
178. Natalia López-Anguita SIG, Maximilian Stötzel, Marina Typou, Iiris Virta, Sara Hetzel, René Buschow, Burak Koksal, Derya Atilla, Ronald Maitschke-Rajasekharan, Rui Chen, Alexandra L. Mattei, Ivan Bedzhov, David Meierhofer, Alexander Meissner, Jesse V. Veenvliet, Aydan Bulut-Karslioglu: Hypoxia induces a transcriptional early primitive streak signature in pluripotent cells enhancing spontaneous elongation and lineage representation in gastruloids. *BiorXiv* 2021.
179. Nguyen PD, Hollway GE, Sonntag C, Miles LB, Hall TE, Berger S, Fernandez KJ, Gurevich DB, Cole NJ, Alaei S, et al: Haematopoietic stem cell induction by somite-derived endothelial cells controlled by meox1. *Nature* 2014, 512:314-318.
180. Adhyapok P, Piatkowska AM, Norman MJ, Clendenon SG, Stern CD, Glazier JA, Belmonte JM: A mechanical model of early somite segmentation. *iScience* 2021, 24:102317.
181. Nelemans BKA, Schmitz M, Tahir H, Merks RMH, Smit TH: Somite Division and New Boundary Formation by Mechanical Strain. *iScience* 2020, 23:100976.
182. Thomson L, Muresan L, Steventon B: The zebrafish presomitic mesoderm elongates through compaction-extension. *Cells Dev* 2021:203748.
183. Dias AS, de Almeida I, Belmonte JM, Glazier JA, Stern CD: Somites without a clock. *Science* 2014, 343:791-795.
184. Alda-Catalinas C, Bredikhin D, Hernando-Herraez I, Santos F, Kubinyecz O, Eckersley-Maslin MA, Stegle O, Reik W: A Single-Cell Transcriptomics CRISPR-Activation Screen Identifies Epigenetic Regulators of the Zygotic Genome Activation Program. *Cell Syst* 2020, 11:25-41 e29.
185. Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, Lu T, Maroc L, Norman TM, Song B, et al: High-content CRISPR screening. *Nature Reviews Methods Primers* 2022, 2.
186. Carlini V, Policarpi C, Hackett JA: Epigenetic inheritance is gated by naive pluripotency and Dppa2. *EMBO J* 2022:e108677.

187. Kuhn M, Santinha AJ, Platt RJ: Moving from in vitro to in vivo CRISPR screens. *Gene and Genome Editing* 2021, 2.
188. Michels BE, Mosa MH, Streibl BI, Zhan T, Menche C, Abou-El-Ardat K, Darvishi T, Czlonka E, Wagner S, Winter J, et al: Pooled In Vitro and In Vivo CRISPR-Cas9 Screening Identifies Tumor Suppressors in Human Colon Organoids. *Cell Stem Cell* 2020, 26:782-792 e787.
189. Ringel T, Frey N, Ringnalda F, Janjuha S, Cherkaoui S, Butz S, Srivatsa S, Pirkl M, Russo G, Villiger L, et al: Genome-Scale CRISPR Screening in Human Intestinal Organoids Identifies Drivers of TGF-beta Resistance. *Cell Stem Cell* 2020, 26:431-440 e438.
190. Kleinman HK, Martin GR: Matrigel: basement membrane matrix with biological activity. *Semin Cancer Biol* 2005, 15:378-386.
191. Ashworth JC, Morgan RL, Lis-Slimak K, Meade KA, Jones S, Spence K, Slater CE, Thompson JL, Grabowska AM, Clarke RB, et al: Preparation of a User-Defined Peptide Gel for Controlled 3D Culture Models of Cancer and Disease. *J Vis Exp* 2020.
192. Brassard JA, Lutolf MP: Engineering Stem Cell Self-organization to Build Better Organoids. *Cell Stem Cell* 2019, 24:860-876.
193. Gjorevski N, Sachs N, Manfrin A, Giger S, Bragina ME, Ordonez-Moran P, Clevers H, Lutolf MP: Designer matrices for intestinal stem cell and organoid culture. *Nature* 2016, 539:560-564.
194. Ranga A, Girgin M, Meinhardt A, Eberle D, Caiazzo M, Tanaka EM, Lutolf MP: Neural tube morphogenesis in synthetic 3D microenvironments. *Proc Natl Acad Sci U S A* 2016, 113:E6831-E6839.
195. Simunovic M, Metzger JJ, Etoc F, Yoney A, Ruzo A, Martyn I, Croft G, You DS, Brivanlou AH, Siggia ED: A 3D model of a human epiblast reveals BMP4-driven symmetry breaking. *Nat Cell Biol* 2019, 21:900-910.
196. Turing AM: The chemical basis of morphogenesis. 1953. *Bull Math Biol* 1990, 52:153-197; discussion 119-152.
197. Takuya Azami SK, Graziano Martello, Yihan Pei, Thorsten Boroviak and Jennifer Nichols: Distinct phospho-variants of STAT3 regulate naïve pluripotency and developmental pace in vivo. *BiorXiv* 2022.

198. Huch M, Koo BK: Modeling mouse and human development using organoid cultures. *Development* 2015, 142:3113-3125.
199. Kagawa H, Javali A, Khoei HH, Sommer TM, Sestini G, Novatchkova M, Scholte Op Reimer Y, Castel G, Bruneau A, Maenhoudt N, et al: Human blastoids model blastocyst development and implantation. *Nature* 2022, 601:600-605.
200. Sozen B, Jorgensen V, Weatherbee BAT, Chen S, Zhu M, Zernicka-Goetz M: Reconstructing aspects of human embryogenesis with pluripotent stem cells. *Nat Commun* 2021, 12:5550.
201. Yu L, Wei Y, Duan J, Schmitz DA, Sakurai M, Wang L, Wang K, Zhao S, Hon GC, Wu J: Blastocyst-like structures generated from human pluripotent stem cells. *Nature* 2021, 591:620-626.

202. Moris N, Anlas K, van den Brink SC, Alemany A, Schroder J, Ghimire S, Balayo T, van Oudenaarden A, Martinez Arias A: An in vitro model of early anteroposterior organization during human development. *Nature* 2020, 582:410-415.
203. Olmsted ZT, Paluh JL: Co-development of central and peripheral neurons with trunk mesendoderm in human elongating multi-lineage organized gastruloids. *Nat Commun* 2021, 12:3020.
204. Russell WB, R.: The Principles of Humane Experimental Technique. 1959.
205. Grosswendt S, Kretzmer H, Smith ZD, Kumar AS, Hetzel S, Wittler L, Klages S, Timmermann B, Mukherji S, Meissner A: Epigenetic regulator function through mouse gastrulation. *Nature* 2020, 584:102-108.
206. Cheng Zhao APR, John Paul Schell, Jere Weltner, Nicolás M. Ortega, Yi Zheng, Åsa K. Björklund, Janet Rossant, Jianping Fu, Sophie Petropoulos, Fredrik Lanner: Reprogrammed blastoids contain amnion-like cells but not trophectoderm. *BiorXiv* 2022.
207. Bowling S, Sritharan D, Osorio FG, Nguyen M, Cheung P, Rodriguez-Fraticelli A, Patel S, Yuan WC, Fujiwara Y, Li BE, et al: An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* 2020, 181:1410-1422 e1427.
208. Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B, Jost M, Quinn JJ, Yang D, Jones MG, et al: Molecular recording of mammalian embryogenesis. *Nature* 2019, 570:77-82.
209. Meissner A: Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol* 2010, 28:1079-1088.
210. Surani MA, Hayashi K, Hajkova P: Genetic and epigenetic regulators of pluripotency. *Cell* 2007, 128:747-762.
211. Ostrop J, Zwiggelaar RT, Terndrup Pedersen M, Gerbe F, Bosl K, Lindholm HT, Diez-Sanchez A, Parmar N, Radetzki S, von Kries JP, et al: A Semi-automated Organoid Screening Method Demonstrates Epigenetic Control of Intestinal Epithelial Differentiation. *Front Cell Dev Biol* 2020, 8:618552.
212. Scheer S, Ackloo S, Medina TS, Schapira M, Li F, Ward JA, Lewis AM, Northrop JP, Richardson PL, Kaniskan HU, et al: A chemical biology toolbox to study protein methyltransferases and epigenetic signaling. *Nat Commun* 2019, 10:19.
213. Wu B, Pan X, Chen X, Chen M, Shi K, Xu J, Zheng J, Niu T, Chen C, Shuai X, Liu Y: Epigenetic drug library screening identified an LSD1 inhibitor to target UTX-deficient cells for differentiation therapy. *Signal Transduct Target Ther* 2019, 4:11.
214. Wu Q, Heidenreich D, Zhou S, Ackloo S, Kramer A, Nakka K, Lima-Fernandes E, Deblois G, Duan S, Vellanki RN, et al: A chemical toolbox for the study of bromodomains and epigenetic signaling. *Nat Commun* 2019, 10:1915.
215. Mantziou V, Baillie-Benson P, Jaklin M, Kustermann S, Arias AM, Moris N: In vitro teratogenicity testing using a 3D, embryo-like gastruloid system. *Reprod Toxicol* 2021, 105:72-90.