



**Master's Program in Molecular Medicine
at the Charité - Universitätsmedizin Berlin**

Molecular recording of mouse embryonic stem cell derived Trunk-Like-Structures

Presented by

SEHER IPEK GASSALOGLU

Born on February 3rd, 1996

Supervisors: Jesse Veenvliet & Adriano Bolondi

First Evaluator: Professor Dr. Bernhard Herrmann

Second Evaluator: Professor Dr. Alexander Meissner

Completed at the Max Planck Institute for Molecular Genetics

Department of Developmental Genetics

Table of contents

1. Abstract.....	1
2. Introduction.....	2
2.1. Embryonic development in mouse	2
2.2. Gastrulation	4
2.3. Formation of gut and primordial germ cells (PGCs).....	5
2.4. Neural Tube formation	6
2.5. Trunk formation and Somitogenesis	7
2.6. Engineering Embryos in a Dish.....	9
2.7. Gastruloids.....	11
2.8. Trunk-like-structures (TLS).....	12
2.9. Lineage tracing	15
2.10. CRISPR-based barcode editing methods	16
2.11. Molecular recording of trunk-like-structures	18
3. Aim of the project.....	19
4. Materials and methods	19
4.1. Mouse Embryonic Fibroblasts (MEFs)	19
4.2. Mouse embryonic stem cell culture.....	20
4.3. Mouse embryonic stem cell lines	20
4.4. Freezing mESCs	20
4.5. Generation of trunk-like-structures (TLS)	20
4.6. Generation of Tracr TLS	21
4.7. Measurement of Cas9-GFP activation to check nucleofection and recombination efficiency	22
4.8. Analyzing the reduction of GFP+ cells	22
4.9. Measurement of Cas9-GFP activation in Tracr TLS	23

4.10.	Resampling Cas9+ cells to asses consecutive editing (also known as “ <i>Luria-Delbrück-like</i> ” assay)	23
4.11.	Phenol-Chloroform gDNA extraction.....	24
4.12.	RNA extraction and cDNA synthesis	24
4.13.	IntBC amplicon PCR and Library Preparation	25
4.14.	Single-cell RNA-seq analysis of Tracr TLS	26
4.15.	Whole-mount immunofluorescence	26
4.16.	Tissue clearing	27
4.17.	Imaging	28
4.18.	Image processing.....	28
4.19.	Image Analysis	29
4.20.	Data analysis	29
4.21.	Morphometric analysis of TLS	29
4.22.	Single-cell RNAseq computational analysis	30
4.23.	Preprocessing	30
4.24.	Quality control	30
4.25.	Cluster Assignement	30
4.26.	Pseudogenes identification.....	31
4.27.	Plotting.....	31
4.28.	FACS analysis	31
5.	Results	34
5.1.	2D Scarring dynamics of Tracr cells.....	34
5.2.	GFP- cells outgrow GFP+ cells over time.....	36
5.3.	Resampling Cas9+ cells to assess consecutive scarring activity (“ <i>Luria-Delbrück-like</i> ” assay).....	37
5.4.	3D Scarring dynamics of Tracr TLS	38
5.6.	Morphological characterization of Tracr TLS.....	40
5.7.	Scarring activity does not affect Tracr TLS morphology	42

5.8. Molecular characterization of Tracr TLS	43
5.9. Combined WNT activation and BMP inhibition results in an excess of somites phenotype in Tracr TLS.....	44
5.10. Sequential formation of somites in Tracr TLS	45
6. Discussion.....	53
6.1. Tracr TLS: molecular recording in 3D structures with an embryo-like architecture.....	55
6.2. Limitations of CRISPR-based barcode editing technology	59
6.3. Shifting to methods that will decrease the initial stress on the cells	59
6.4. Role of ECM during development	60
6.5. Concluding remarks and future perspective.....	61
7. Literature.....	62
8. Appendix.....	71
9. List of abbreviations	72
Statement of independent work	74
Acknowledgement.....	75

1. Abstract

Determination of cell fate is a crucial and complex part of embryonic development. Extensive efforts have been made to understand the mechanisms through which cell lineages give rise to differentiated cells. However, studying mammalian development is particularly challenging owing to the intricacy of the underlying decision making process. In this project, to understand the interplay between tissue morphogenesis and lineage decisions we took advantage of two remarkable systems - a highly informative molecular recorder, *tracr* cells and an organized post-occipital embryonic model, Trunk-like-structures (TLS). Since embryogenesis is inaccessible to extensive analysis *in vivo*, in this project we aimed to characterize the complexity of the *tracr* system to better understand how multiple lineage decisions are regulated in TLS. Our system demonstrated accumulation of scars in all cell types of TLS. This shows the potential of the *Tracr* system to trace individual cell states which give rise to embryo-like structures. Our molecular recording approaches will ultimately rely on single cell RNA sequencing libraries as a readout to construct lineage relationships between cells. Our *Tracr* system holds a great promise to decipher cell lineage decisions to understand developmental biology in-depth.

2. Introduction

After implantation the mammalian embryo develops within the uterus. Consequently, detailed analysis of the highly dynamic process of post-implantation embryogenesis is inaccessible to extensive analysis *in vivo*. As a result, a detailed picture of the molecular and morphological processes that eventually lead to the emergence of the three germ layers (endoderm, mesoderm, ectoderm) and the development of the three embryonic body axes (antero-posterior, dorsal-ventral, medio-lateral) still has to be painted (Solnicakrezel L. et al., 2012). To circumvent the impediments of *in vivo* research, extensive efforts have been put into the development of *in vitro* models that recapitulate different stages of mammalian embryogenesis. Recently developed stem cell models represent powerful platforms to deconstruct the dynamic processes that are inaccessible in the embryo in a controlled manner (Veenvliet, Bolondi et al., 2020, Science, in press, Zhang S. et al., 2019; Shahbazi et al., 2019; Beccari L. et al., 2018; Rivron N. et al., 2018; Sozen B. et al., 2018; Harrison S. et al., 2017; Turner D.A. et al., 2017). For example, mouse embryonic-stem cells (mESCs) can form aggregates reflecting the post-occipital embryo (*gastruloids*), but lacking proper morphogenesis (Beccari L. et al., 2018). Building on this work, Veenvliet, Bolondi et al. recently developed a novel 3D culture method for post-implantation embryogenesis that displays a high level of embryo-like organization (Veenvliet, Bolondi et al., 2020, Science, in press). To further understand the lineage relationships of differentiated cells in developing embryos, several sequencing-based lineage-tracing methods were developed (Chan, Smith et al., 2019; Spanjaard B. et al., 2018; Raj B. et al., 2018; Alemany A. et al., 2018; Kalhor R. et al., 2018). Combining novel embryoid models with such lineage-tracing methods offers an opportunity to organize the previously obtained taxonomy of cell states and state trajectories into lineage trees.

2.1. Embryonic development in mouse

After fertilization occurs in the ampulla of the oviduct by the fusion of a sperm with an egg, the first mitotic divisions known as cleavage divisions take place, as the cilia of the oviduct move the zygote towards the uterus. The entry position of the sperm from the animal axis of the ovum predicts the plane of the initial cleavage (Piotrowska & Zernicka, 2001). At the 8- or 16-cell-stage, the embryo undergoes compaction due to the expression of cell adhesion proteins such as E-cadherin, which gradually form tight

junctions between the blastomeres, forming a mulberry shape termed morula (Peyrieras et al., 1983; Fleming et al., 2001).

At the morula stage, prior to blastocyst formation, every blastomere within the morula expresses Cdx2 and Oct4 transcription factors (Niwa H. et al., 2005). At this point, the embryonic cells take the first cell fate decision between the trophoblast and inner cell mass (ICM) lineage. Eventually, descendants of external cells express Cdx2 which leads to the formation of trophoblast (trophectoderm), whereas Oct4 blocks the expression of Cdx2 in inner cells to form the pluripotent cells of the ICM. Through the process of cavitation, the morula starts developing an internal cavity. The trophoblast secretes sodium ions into the cavity, enabling the movement of water into the hollow, to form the blastocoel (Borland et al., 1977; Ekkert et al., 2004; Kawagishi et al., 2004). As the blastocoel expands, the ICM arranges on one side of the ring of trophoblast cells (Barends P.M.G et al., 1989).

Prior to implantation, a second lineage choice occurs in the ICM. Here, expression of Nanog keeps the cells in their pluripotent state whereas Gata6 leads to the formation of primitive endoderm (Ralston and Rossant 2005; Rossant 2006; Gilbert S., 2006, *Developmental biology*, p.382). On embryonic day 4.5 (E4.5), the mouse embryo is composed of three distinct cell lineages: epiblast, primitive endoderm (bilayered disc) and trophectoderm (TE) (Rivera-Pérez & Hadjantonakis, 2015) (Figure 1). Once the blastocyst hatches from the zona pellucida by secreting digestive-enzymes and signals such as Chorionic gonadotropin, it can adhere to the endometrium and make its way through the uterus lining to bury itself in the uterine wall (Srisuparp S. et al., 2001). Upon implantation at E5, the mural trophectoderm gives rise to trophoblast giant cells, whereas the remaining trophectoderm forms the ectoplacental cone and the extra-embryonic ectoderm. The trophectoderm region proximate to the ICM is known as polar trophectoderm whereas the distant set of cells is known as the mural trophectoderm (**Figure 1**) (Tam & Behringer, 1997). The blastocyst forms an egg cylinder shape as the primitive endoderm expands to line the epiblast and the inner line of mural trophectoderm to give rise to visceral endoderm and parietal endoderm respectively at E5.5 (**Figure 1**) (Tam & Behringer, 1997). Gastrulation begins at the posterior end of the embryo as the primitive streak gives rise to the trilaminar embryo from the bilaminar embryonic disc. (See section: Gastrulation)

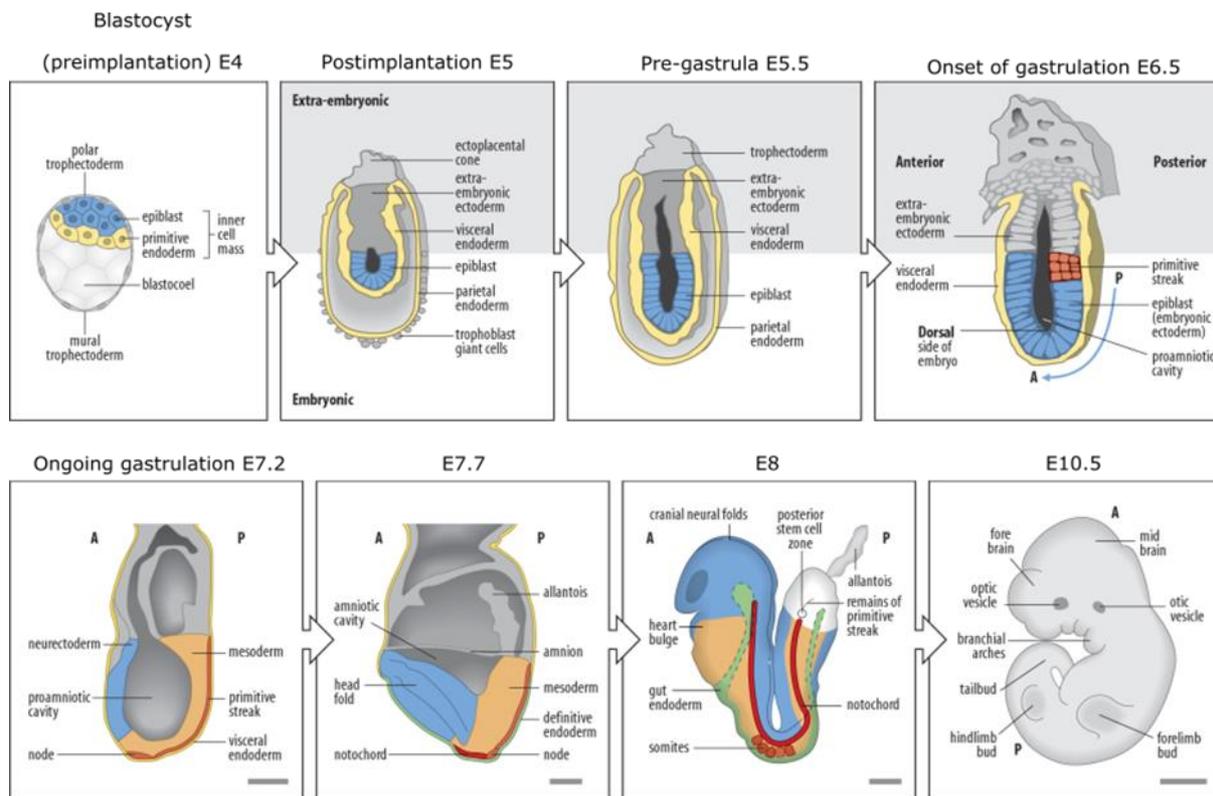


Figure 1 | Embryonic development in mouse At E4.0 the bilaminar embryonic disc becomes apparent with the formation of the epiblast and primitive endoderm. At E4.75 the embryo adheres to the endometrium. Gastrulation begins with the formation of the primitive streak at E6.5. At E6.5 – E7.5 during gastrulation the trilaminar embryo is formed from the bilaminar embryonic disc (Tam & Behringer, 1997). By E8, the embryo has developed several organ primordia such as gut endoderm, allantois, and initiated processes such as neurulation, somitogenesis and regionalisation of the heart (Massarwa & Lee Niswander, 2013). (Adapted from lecture “Molecular & Developmental Biology” (BIOL3530 with Dr. Brian E. Staveley Department of Biology Memorial University of Newfoundland http://www.mun.ca/biology/desmid/brian/BIOL3530/DEVO_03/devo_03.html 25.07.19)

2.2. Gastrulation

The signals secreted by the two extraembryonic tissues (the visceral endoderm (VE) and the trophectoderm) lead to the initiation of gastrulation at the posterior region of the embryo. The trophectoderm releases BMP (Bone Morphogenic Protein) signals to ensure that gastrulation initiates at the posterior end of the embryo while the anterior visceral endoderm secretes inhibitors to BMP, WNT and Nodal to launch the specification of the primitive streak (Siggia & Warmflash., 2018; Arnold S.J. et al., 2009; Perea-Gomez A. et al., 2002). The formation of the primitive streak defines the start of the gastrulation process. The first regionalization starts with Nodal signaling at E5 which induce proximal-distal patterning in epiblast (Arnold & Robertson, 2009). At E6.0 the anterior – posterior (A-P) axis arises by the migration of the DVE to the

prospective anterior side of the embryo (Beddington, R. S., et al, 1999). The primitive node forms at the anterior end of the primitive streak and elongates towards the posterior end of the primitive streak to form the primitive groove at E7.2. Once the primitive streak is complete, epiblast (embryonic ectoderm) cells lose E-cadherin, undergo epithelial to mesenchymal transition (EMT) and ingress as individual mesenchymal cells through the primitive streak to further give rise to the definitive endoderm and mesoderm (Burdal et al., 1993; Williams et al., 2012). The remaining epiblasts will form the definitive ectoderm, the most distal germ layer (Stern & Downs, 2012; Enders A. C. 1986). Fibroblast growth factor (FGF) and WNT are important in cell movement and cell specification. Their activation downregulates E-cadherin to allow the movement of the cells through the primitive streak and regulates transcription factors such as Brachyury (T), Tbx6 and Snail which play a role in mesodermal migration, patterning, and specification (Ciruna B. & Rossant, 2001; Gilbert S., 2006, Developmental biology, p.387). At the late gastrula stage, mesoderm is patterned mediolaterally arising from different regions of the primitive streak. The region under the influence nodal signaling of the primitive streak gives rise to the axial mesoderm, the perinodal (anterior) segment of the streak to paraxial mesoderm, the mid-segment to lateral mesoderm and the posterior segment to extraembryonic mesoderm (Pei et al., 2019; Tam & Behringer, 1997).

2.3. Formation of gut and primordial germ cells (PGCs)

During gastrulation at ~E7.25, a subset of ~40 early primordial germ cells migrate through the primitive streak and can be identified as a group of DPPA3+ cells which further migrate into the location of the posterior endoderm that forms the hindgut. Later at ~E9.5, they further migrate to the mesentery and finally to the genital ridges at ~E10.5. At this final destination, they will give rise to oocytes and spermatozoa (Saitou & Yamaji, 2012; Richardson & Lehmann, 2010).

Between E7 and E9, definitive endoderm (DE) gives rise to the gut tube, which runs all along the AP axis of the embryo. At E7.75, anterior- and caudal intestinal portal of the gut tube are formed. Between E8.0-9.5, the lateral endoderm continues to fold ventrally like a sheet of paper until it forms a tube-like-structure that is encircled by mesenchymal cells (Spence, J. R. et al., 2011) The emerging DE from the primitive streak does not completely replace the VE and by E8.75 the VE cells contain ~10% of

foregut endoderm, ~15% of midgut and ~35% of hindgut endoderm (Kwon et al., 2008).

2.4. Neural Tube formation

During the late stages of gastrulation and early neurulation, the notochord forms, which sends signals along with the primitive streak to initiate neural differentiation. These signals are received by the neuroectoderm to start neurulation, the process that causes the midline ectoderm which contains the neural precursor cells to thicken in order to form the neural plate, which will eventually fold until the edges come in contact and fuse to form the neural tube (Neuroscience, 2nd edition Editors: Dale Purves, George J Augustine et al., 2001, Chapter: The Initial Formation of the Nervous System: Gastrulation and Neurulation; Bush K.T et al., 1990). Specifically, at E7.5, the anterior neural plate forms to give rise to prospective forebrain, midbrain, hindbrain and anterior spinal cord whereas, the posterior spinal cord arises separately and is formed by the neuromesodermal progenitors (NMPs) (Henrique et al., 2015). Moreover, Sonic hedgehog (Shh) produced in the notochord induces distinct ventral identities in the overlying notochord which in later stages plays an important role in neural tube development (Kahane & Kalcheim, 2020). Neural tube formation arises by two main processes. During primary neurulation, the neural plate cells are directed to proliferate, invaginate and bud off from the surface to form a hollow tube. Whereas during secondary neurulation, the neural plate creases inward to fuse its edges (Gilbert S., 2006, Developmental biology, chapter : Formation of neural tube). Sox2 expression is controlled by two different enhancers to give rise to the posterior neural tube. The N1 enhancer drives Sox2 expression in the NMP-containing epiblast region, whereas the N2 enhancer drives Sox2 expression in ESCs and anterior epiblast. The posterior nervous system and CLE formation is determined by the epiblast population controlled by transition from N2 to N1 enhancer activity (Takemoto et al., 2006). FGF and WNT signals induce the expression of T and the Sox2 (N1) enhancer. (Takemoto et al., 2006). Thus by E8.5, NMPs choose the neural fate by downregulating T and activating SOX2, to form the neural tube which eventually gives rise to the spinal cord and new mesoderm progenitors contribute to presomitic mesoderm (PSM) (Koch et al., 2017; Henrique et al., 2015).

2.5. Trunk formation and Somitogenesis

The mammalian trunk arises from the continuous movement of cells from the primitive streak, the caudal lateral ectoderm (CLE) and the node streak border (NSB). Neuromesodermal progenitors (NMPs) are located in the CLE and NSB and transcriptionally characterized by co-expression of the pan-mesodermal marker *T* and the neural progenitor cell regulator *SOX2* (**Figure 2**) (Koch et al., 2017). As mentioned earlier, *T* controls the activation of mesodermal genes while repressing the neural lineage genes, whereas *SOX2* regulates the neural progenitors and represses mesodermal genes directly or indirectly (Koch et al., 2017; Bergsland et al., 2011). The WNT and FGF pathways form an autoregulatory feedback loop, in which *WNT3a* regulates expression of *FGF8* and *T* in the tail bud, while *FGF4* and *FGF8* regulates WNT signalling in presomitic mesoderm (Chal & Pourquié 2017; Aulehla et al., 2003, Cambrey and Wilson, 2007). *T* and *Cdx2* directly co-induce canonical WNT and FGF signalling which favours *Tbx6* (Koch et al., 2017; Gouti M. et al., 2017). The mouse PSM is composed of four consecutive transcriptional domains: the tail bud, the posterior presomitic mesoderm, the anterior presomitic mesoderm, and the forming somite (S0) (Chal et al., 2015) The paraxial mesoderm progenitors in the tail bud are exposed to the highest levels of WNT & FGF signaling and will further give rise to posterior PSM (Aulehla et al., 2003; Chal et al., 2015; Naiche et al., 2011).

At the determination front, various signaling, metabolic and transcriptional changes happen in pPSM to enter to the anterior PSM (Chal et al., 2015; Oginuma et al., 2017; Ozbudak et al., 2010). The determination front is the intersecting point between the anterior gradient of retinoic acid activity and posterior FGF & WNT expression. The somitic regions produce retinoic acid in the aPSM which is counteracted by the WNT/FGF signaling in the pPSM (Sakai et al., 2001). The process through which a patterned somite is formed is controlled by the segmentation clock, which generates pulses of Notch, FGF and WNT signaling to control the periodic production of somites (reviewed by Hubaud and Pourquié, 2014). The anterior and posterior boundaries of the future segment is determined by the bilateral stripes of *Mesp2* (Oginuma et al., 2008; Takahashi et al., 2000). Somitogenesis in mouse initiates from E8.0 and continues until E13.0 during embryonic development (**Figure 1**) (Saga, 2012; Cooke J. & Zeeman E.C., 1975) with every segment forming with a 2 hour periodicity (Kageyama R. et al., 2012, Pourquié, 2011).

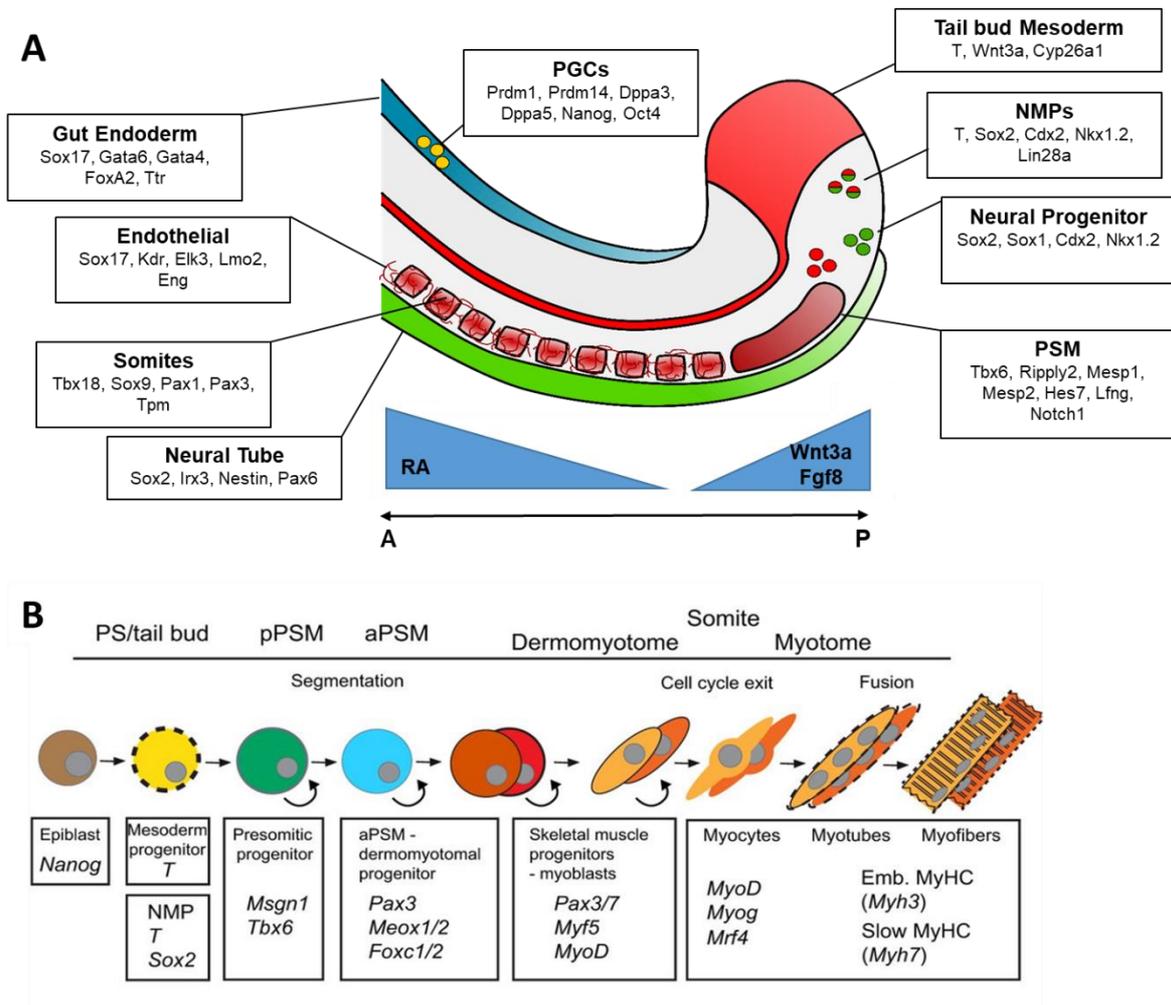


Figure 2 | Trunk formation and somitogenesis in mouse. A. By E8 the embryo gives rise to the trunk with the formation of several cell and tissue types such as somites, neural tube, gut endoderm and PGCs. B. Differentiation of the presomitic mesoderm into dermomyotome and myotome. Sequential formation of different cell types with the associated markers are shown. Figure A drawn by Adriano Bolondi using Inkscape. Figure B adapted from Chal & Pourquié 2017

As soon as a new epithelial somite is formed, it receives signals from the neighbouring tissues that drive the somite through maturation, specification and differentiation to different somite derivatives such as dermomyotome and sclerotome which give rise to the dermis, thoracic, abdominal wall and skeleton muscles, as well as the proximal rib, vertebral column, blood vessels, endothelial cells and meninges of the spinal cord, respectively (Yusuf & Brand-Saber, 2006; Christ et al., 2007).

2.6. Engineering Embryos in a Dish

Traditionally, developmental biology experiments were being carried out by direct observation of the developing embryo *in vivo* which restricts the observation of morphogenesis and growth of mammalian species. One of the first people to establish tissue culture was Wilhelm Roux who removed the medullary plate of an embryonic chicken and preserved it in a solution for two weeks, thus showing that tissues could live outside the body (discussed in Rodriguez-Hernandez C.O., 2014). Thus the “top-down” approaches were slowly established which included the deconstruction of the embryo into smaller components by experimental manipulation (Zijlstra, M., et al., 1989; Gardner R. L., 1968; Tarkowski, A. K., 1959). Emerging insights and technological advancements led scientists to develop a contrasting but complementary method comprising the pluripotent stem cells to build “synthetic” embryo-like structures, a “bottom-up approach” (Shahbazi et al., 2018).

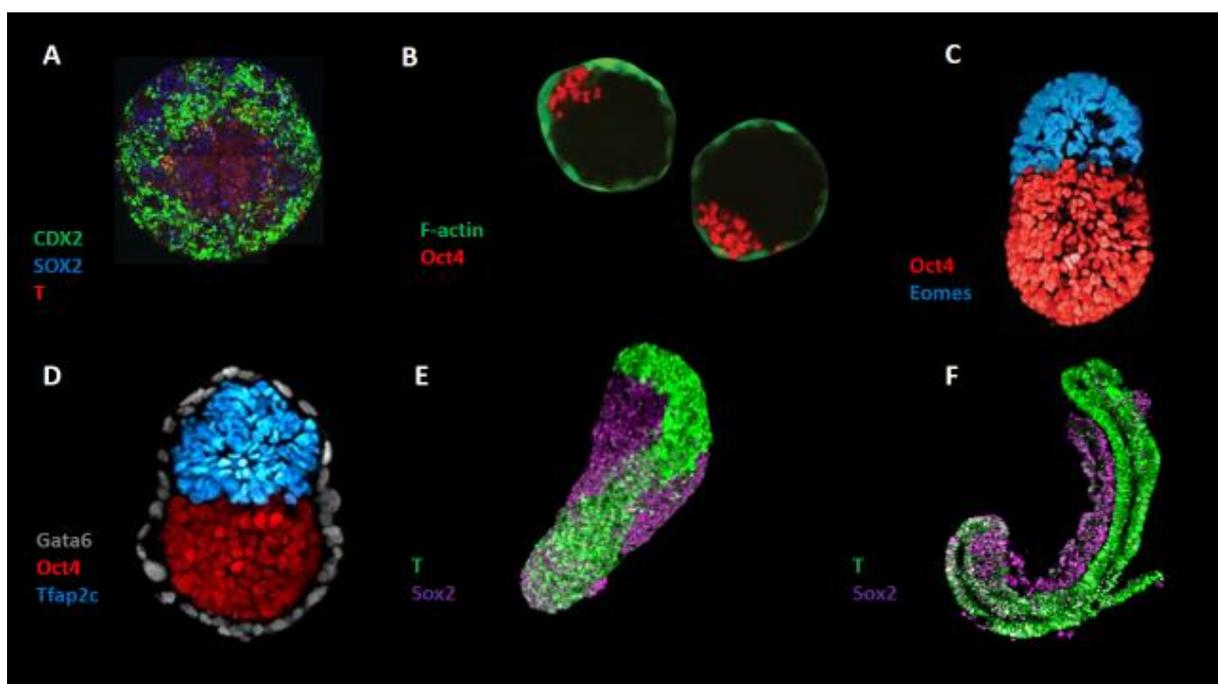


Figure 3 | Mouse embryoid models. A. Blastoids form blastocyst-like structures. B. Mouse micropatterns show cell fate specification comparable to cell states *in vivo* gastrulation E5.5 to E7.5. C. Polarized embryo-like-structure (ETS) consisting of ESCs and TSCs in a 3D scaffold recapitulates several morphological processes of E4.75 to E6.5 embryos. D. Gastrulating embryo-like structure (ETX) are composed of ESCs, TSCs and XEN cells. E. ESC-derived gastruloids exhibit symmetry breaking, germ layer formation, axial organization and elongation. F. Trunk-like-structures (TLS) display proper segmentation and neural tube-like structure. Cdx2 marks trophectoderm; T marks mesoderm; SOX2 marks neuroectoderm; F-actin marks filamentous actin; Eomes marks

TSC/trophectoderm/extraembryonic ectoderm cells; Oct4 marks pluripotent epiblast cells; CDH1 marks epithelial cells; CDH2 marks mesenchymal cells; TSCs, trophoblast stem cells. **Adapted from Morgani et al., 2019; Rivron R. C. et al., 2018; Harrison, Sozen et al., 2017; Sozen et al., 2018; Veenvliet, Bolondi et al., 2020, Science, in press; Shahbazi et al., 2019**

Over the last decade, multiple models of different embryonic stages have been developed, with various degrees of molecular and morphological complexity.

Several mouse stem cell models mimicking pre- and post-implantation stages were developed over the years, among which Morgani and colleagues (2018) treated mouse ESC micropattern system with different signals such as FGF, Nodal, BMP and WNT to initiate cell fate specification comparable to cell states *in vivo* gastrulation E5.5 to E7.5. Treatment with these signals directed mouse epiblast-like cells to undergo an epithelial-to-mesenchymal transition to form a multilayered ring to radially pattern posterior mesoderm fates. Micropatterns may be important to understand the cell dynamics at gastrulation as subpopulations can be observed on a spatial level (**Figure 3A**) Blastoids recapitulate E 3.5 blastocysts are generated by the addition of TSCs on the non-adherent ESC aggregates after 24h, to enable the organization and engulfment of TSCs around the ESCs (**Figure 3B**) (Rivron R. C. et al., 2018;. Moreover, Harrison et al. (2017) established a system by combining mESCs and Trophoblast-stem cells (TSC) in Matrigel, a 3D extracellular matrix (ECM) proxy that serves as a surrogate for the basement membrane normally provided by the primitive endoderm, and necessary for epiblast polarization and lumenogenesis *in vivo*. This model, that is reminiscent of the E5.5 embryo, demonstrated a characteristic cylindrical architecture with polarization and lumen formation in the ESC compartment followed by the formation of cavity in the TSC compartment, which then join to form the pro-amniotic cavity in response to Nodal signalling (**Figure 3C**) (Harrison et al., 2017). However, no anterior structures and EMT were observed. Hence, the same lab hypothesized that the crosstalk between the two tissues remains inadequate to properly recapitulate embryogenesis *in vitro* and developed a new model (ETX) composed of ESCs, TSCs and extraembryonic endoderm stem cells (XEN) essential for the morphogenesis of the embryo. XEN cells replace Matrigel and induce the structures to form an embryo-like architecture and display EMT (**Figure 3D**) (Sozen et al., 2018). Interestingly, other models suggest that the extra-embryonic cells are dispensable for the establishment of the body plan. For example, assemblies of mESCs (gastruloids) are able to establish anteroposterior polarity and elongation of

the posterior region in the absence of extra-embryonic tissues (**Figure 3E**) (Van den Brink et al. 2014; Turner et al., 2017) (See section: Gastruloids). The model was adapted from a study that observed the ability of the P19 embryonal carcinoma cells to form polarized, elongated structures during differentiation (Marikawa et al., 2009). Although gastruloids mimic gene expression domains similar to E8 embryo, associated morphology is missing. Therefore, Trunk-like-structures were established which reflect the embryo at a morphological and molecular level (**Figure 3F**) (Van den Brink et al., 2020; Veenvliet, Bolondi et al., 2020, Science, in press) (See section: Trunk-like-structures)

2.7. Gastruloids

Gastruloids are one of the models that have facilitated the study of post-implantation development by recapitulating stage E5.5-E9.5 of the mouse embryo and exhibiting similar processes, including symmetry breaking, germ layer formation, axial organisation and elongation (**Figure 4A**) (Beccari et al., 2018; Van den Brink et al., 2014). These structures are generated from embryonic stem cells solely and the absence of BMP4 indeed indicates the absence of extra-embryonic cells (Turner D.A et al., 2016). *In vivo*, the E5.5 embryo forms a symmetric cup-shaped epithelium with 150-300 epiblasts cells embedded in visceral endoderm and attached to the extra embryonic ectoderm on the proximal end (Munoz-Descalzo S. et al., 2015; Davidson K.C., 2015). This possibly explains why, for most ESC lines, formation of gastruloids requires the aggregation of 200-300 cells to induce embryo-like growth and elongation. By providing the right conditions at the right time, the spatiotemporal behaviour of the embryo can be recapitulated. CHIR99021 (Chiron), a constitutive activator of the WNT pathway through GSK-3 inhibition, activates T and together with WNT and FGF signaling maintain the NMPs via a positive feedback mechanism, is added at 48h post aggregation leading up to robust induction of a T expressing pole at the posterior end of the structure and axial elongation (Van den Brink et al., 2014). Interactions between Nodal and WNT signalling promote an intrinsic symmetry breaking event that results in the expression of T (Turner D.A et al., 2016). T induces the mesodermal marker Tbx6 and towards the 120h timepoint, gastruloids express different sets of genes in a spatially organized fashion resembling the post-occipital embryo such as SOX17 for the endoderm, T & TBX6 for mesoderm and SOX1 & SOX2 for neuroectoderm (Van

den Brink et al., 2014). In depth transcriptome analysis demonstrated a large set of genes present in gastruloids consistent with the locations of these tissues in embryos, such as: cardiac, endothelial and head mesenchymal cells in the anterior domains; sensory neuron precursors and olfactory-receptor genes, definitive endoderm, heart markers, early pre-somitic genes and regulators of segmentation clock (Van den Brink et al., 2020). Although the regulators of the segmentation clock are present, gastruloids lack proper somite formation (Van den Brink et al., 2020, 2014; Baillie-Johnson, P., et al. 2015).

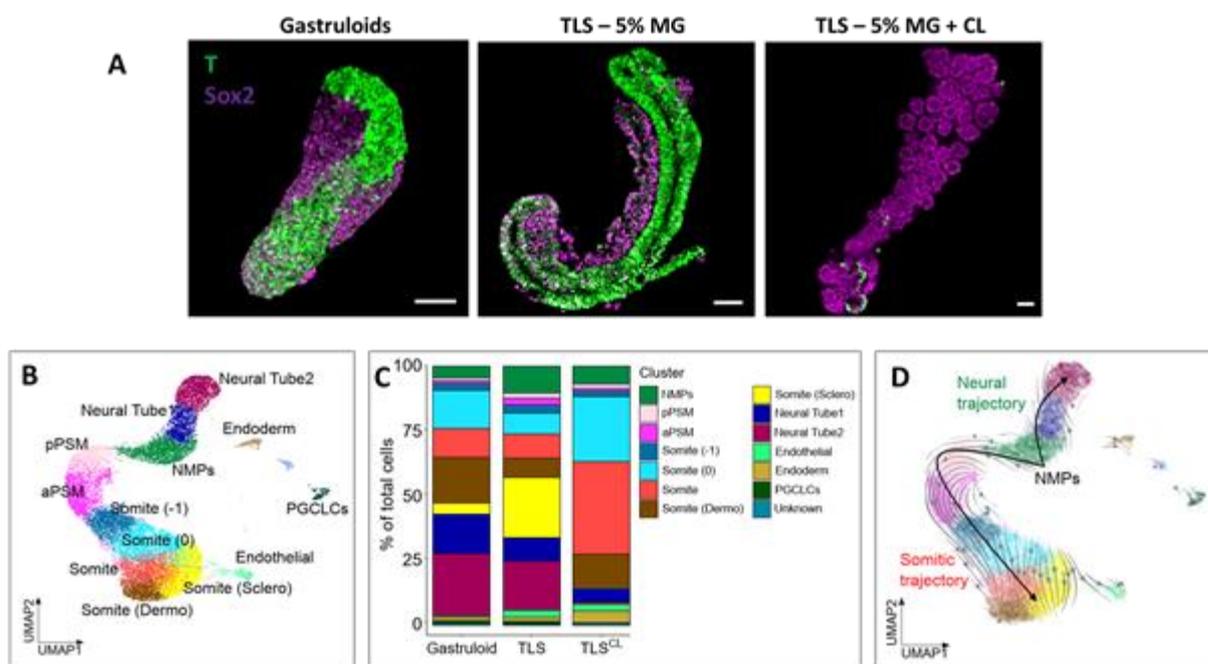


Figure 4 | TLS display high levels of embryo-like organization, reflecting the embryo at a morphological and molecular level. A. Gastruloids exhibit germ layer formation and axial organisation but fail to develop neural tube and somites. Embedding TLS in 5% MG induces segmentation and neural tube formation. 5% MG + CL shifts the NMP lineage choice towards the paraxial mesoderm, resulting in excessive somite formation. B. TLS comprise 14 different cell states. C. Comparison of the molecular profile of Gastruloids, TLS, TLS-CL. D. RNA Velocity analysis revealed neural and somitic trajectories rooted in the NMPs, Adapted from Veenliet, Bolondi et al., 2020, Science, in press. MG – matrigel; CL – Chi + LDN

2.8. Trunk-like-structures (TLS)

Matrigel is a surrogate of ECM, formed by a mixture of gelatinous proteins such as laminin, collagen IV and enactin. (Kleinman H. K. et al 1982; Chris S. Hughes C. S. et al., 2010). It was first composed by Kleinman and colleagues (1982) when they

extracted type IV collagen, laminin, and heparan sulfate proteoglycan from Engelbreth-Holm-Swarm mouse sarcoma and normal cells. Immunoassay studies have revealed that Matrigel contains several growth factors such as insulin-like growth factor 1, nerve growth factor, Transforming growth factor beta, basic fibroblast growth factor, platelet-derived growth factor, and epidermal growth factor (Vukicevic, S. et al., 1992). Apart from mechanical and structural support, ECM promotes cell growth, migration and cell-cell communication (Bonnans et al., 2014). Depending on the matrix type and presence of growth factors, using extracellular matrices can promote differentiation towards a specific lineage. For instance, individual components of ECM such as laminin-1 or collagen I have no impact on the growth or morphology of the cells. In contrast, matrigel can induce formation of glandular- and tubular-like structures (Bonnans et al., 2014). As mentioned in the earlier section, embedding mESC aggregates and trophoblast stem cells in Matrgel give rise to (TSCs) Polarized embryo-like structures (ETS) that recapitulate peri- and early post-implantation stages. Importantly, addition of XEN stem cells on ETS showed that matrigel can substitute the basement membrane secreted by the XEN-derived visceral endoderm (VE) *in vivo* (Bedzhov and Zernicka-Goetz, 2014; Sozen et al., 2018). Since gastruloids lack somitogenesis, Veenvliet, Bolondi et al. (2020) embedded 96h gastruloids in 5% matrigel which resulted in formation of somites and a neural tube-like structure. Chemical modulation of TLS with WNT activator Chiron alone or in combination with LDN (BMP signaling inhibitor) resulted in an excess of somites and improved physical separation of the segments. (**Figure 4A**) (Veenvliet, Bolondi et al., 2020, Science, in press). Generating these model further revealed that Matrigel upregulates integrins, transmembrane receptors mediating cell adhesion to the ECM *in vivo*. (Veenvliet, Bolondi et al., 2020, Science, in press). Notably, van den Brink et al. (2020) showed around the same time that 10% matrigel results in the conversion of gastruloids into a “string-of-somites”. However, in this paper neural tube and gut formation was not reported, and segments were not well-defined, but instead appeared as small ‘indentations’ anteriorly to the differentiation front. As shown by RNA velocity analysis, neural and somitic trajectories rooted in the NMPs as shown by RNA velocity analysis (Figure 4D) replacement of presomitic mesoderm and early NMP-like cells at 96h by more mature neural, somatic and late NMP-like cells at 120h. Detailed molecular analysis by single-cell RNA-seq (scRNA-seq) revealed the presence of 14 cell clusters.. Early & late neural-tube cells, as well as posterior and anterior PSM, somitic cells and somite derivatives like sclerotome and dermomyotome

were detected. In addition, endodermal, endothelial cells and primordial germ cell-like cells (PGCLCs) were also observed in TLS (**Figure 4B, C, D**).

Through spatial allocation of selected developmental marker genes it was shown that TLS display proper AP polarity during segmentation as shown by a stripy pattern of *Uncx* + posterior and *Tbx18*+ anterior somite domains (Veevliet, Bolondi et al., 2020, Science, in press). Immunofluorescence staining further revealed the alternating expression of *Uncx4.1* and *Tbx18* (Van den Brink et al., 2020) Moreover, formation of a gut-like-structure in a subset of the TLS was demonstrated through *FOXA2* and *SOX17* immunolocalization. Finally, *DPPA3*+ Primordial-germ-like cells were detected in the gut-like-structure. In addition, scRNA-seq revealed the presence of dorsal and ventral somite formation. Strikingly, as opposed to gastruloids, somitic and neural cells display apical-basal polarity in TLS with *NCAD* and F-actin (Phalloidin) accumulating at the apical surface (Veevliet, Bolondi et al., 2020, Science, in press). In addition, the segmentation clock confirmed by the expression of *Lfng* controls the sequential formation of somites in TLS (Van den Brink et al., 2020; Veevliet, Bolondi et al., 2020, Science, in press).

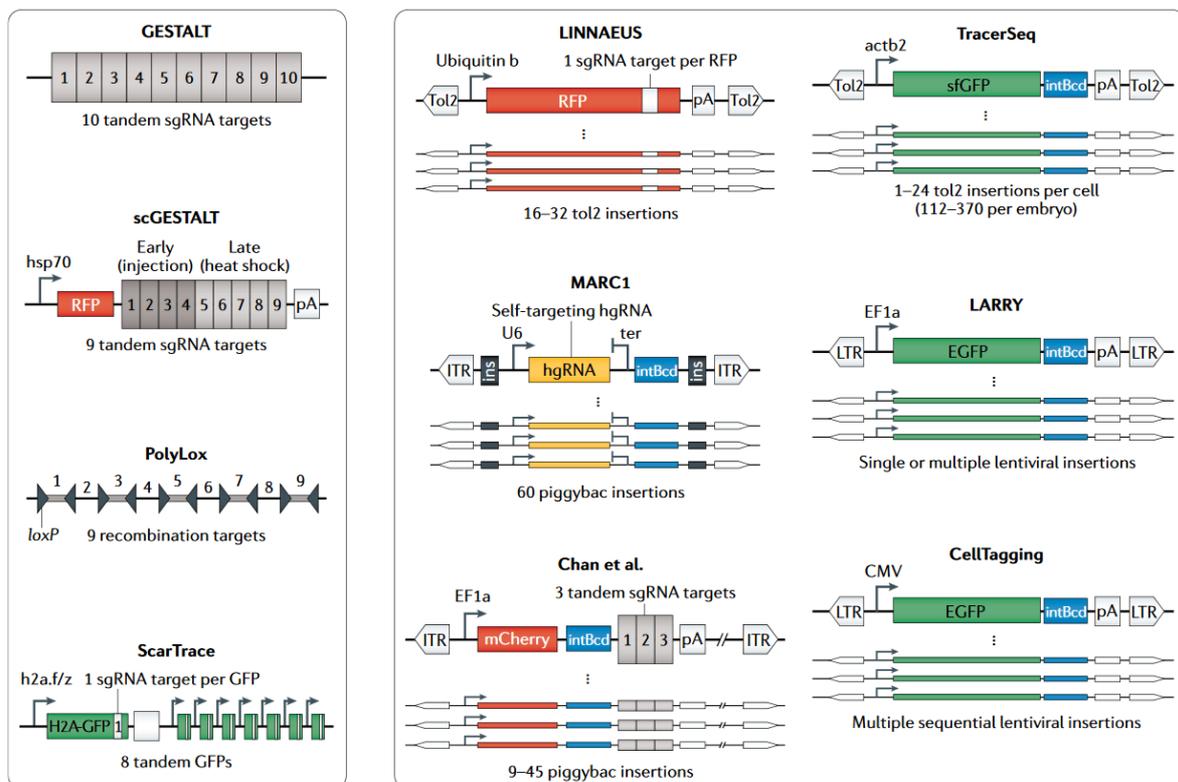


Figure 5 | Sequencing based lineage-tracing methods. GESTALT, genome editing of synthetic target arrays for lineage tracing; scGESTALT, single- cell GESTALT; LINNAEUS, lineage tracing by nuclease-activated editing of ubiquitous sequences; MARC1, mouse for actively recording cells 1; LARRY, lineage and RNA recovery; [Adapted from Wagner & Klein, 2020](#)

2.9. Lineage tracing

Currently, a vast amount of research is ongoing to understand the molecular mechanisms that control the differentiation of the cells from the zygote stage to a full developed organism to reveal the order of the events cells choose to get to their final identities. The recent advances in single cell RNA sequencing have revolutionized the transcriptome studies and by the recent application of lineage-tracing methods, mapping cell lineage trajectories have relatively become easier.

Lineage tracing experiments have existed since 19th century where scientists observed the changes in transparent invertebrate embryos by light microscopy. Ascidians were among the first organisms whose map was annotated, where Conklin, E.G., 1905 predicted cell fates by observing unsegmented eggs of three different genera of ascidians and their cleavage sequences up to the tadpole stage. Around the same time a zoologist named Whitman studied the development in leech and came to a conclusion that the fate of the cells from the earlier cleavages were distinct to a specific fate in later developmental states (Whitman, C.O., 1878). Moreover, various studies were done on *C.elegans* in the same context until in 1983, Sulston and colleagues were able to trace the entire embryonic cell lineage of *C.elegans* from zygote to newly hatched larvae stage using a light microscope and videotape recorder Sulston et al., 1983.

All the organisms mentioned above were studied by direct observation owing to their determinate cleavage patterns, the complexity of the vertebrates limits their investigation by direct observation. Thus, lineage tracing brought more advanced molecular techniques into the picture. In the beginning of the 20th century, molecular dyes and radioactive tracers were being used to label the cells These reporter systems can be integrated into the genome and the clonal populations are tracked down by activating the fluorescent marker by a drug, transfection or permanently by recombination, which then passes the marker to the daughter cells (Kretzschmar K. et al., 2012).

Most recent advances in high-throughput sequencing have reformed the former lineage tracing experiments. The use of DNA barcodes combined with sequencing technology allows for the recording of every single division in an organism by making changes in a specific target-gene, whole genome or mitochondrial genome. Moreover, recent methods have preferred the use of ‘cumulative barcoding’ in which every cell contains more than a single barcode that allows the tracking of nested clones within the population rather than ‘one-step barcoding’ where every cell contains a single barcode. In addition, these barcodes can be stored in tandem arrays or distributed arrays by several transgenic strategies. General methods used to generate barcode libraries involve: 1) random recombination, 2) cell tagging by lentiviruses, 3) transposases, 4) *in vivo* editing of target sites by CRISPR-Cas9 technology (Wagner D. E. et al., 2020).

In recombinase technology, the target DNA flanked by two recognition sites is excised or inverted by the enzymatic reaction of recombinases, creating huge diversity by rearranging numerous recognition sites in a barcode. PolyLox is a model that takes advantage of this system by using ten LoxP sites where the unique barcodes lie between nine intervening sequences. This system weakens the Cre enzyme to mediate incomplete Cre-mediated recombination to avoid the reaction of Cre on the overall Cre-Loxp sites, thus leading to the generation of 1.8 million unique barcodes (Wei P. et al., 2017). An alternative robust strategy to build up barcode libraries is the use of lentiviruses to mark cells by at least 3 – 4 tags and pool cells at different time points for transcriptome analysis. LARRY (lineage and RNA recovery) (Weinreb C. et al., 2020) and CellTagging technology are among these methods (CellTagging protocol Kong W. et al., 2019; Guo C., et al., 2019). Likewise, Tol2 transposon system can also be used to integrate barcodes just as seen in Tracer-seq (transposon-based barcoding approach) (Wagner D.E et al., 2018).

2.10. CRISPR-based barcode editing methods

An approach which is also used in this project is CRISPR-based barcode editing technology which introduces genomic barcodes to the cells. The system is induced to activate and direct Cas9 nuclease to the barcode region. Next, the Cas9-gRNA complex induces double-stranded DNA breaks (DSB) which in return activates a

molecular repair mechanism, like non-homologous end joining repair (NHEJ) that results in heritable, traceable insertions or deletions (indels). These indels may be referred to as “edits” or “scars”.

GESTALT (genome editing of synthetic target arrays for lineage tracing) was the first method to use the CRISPR-based barcode editing which used a system of nine to twelve Cas9 targets in a compact locus, each made up of 300 bp approximately (McKenna A. et al., 2016). This system was advanced by integrating the barcodes into constitutively transcribed regions, enabling their sequences to be read in the mRNA format which reformed its name to scGESTALT (single-cell GESTALT) (Raj B. et al., 2018). ScarTracr is another system that used eight tandem green fluorescent protein (GFP) sequences engineered into zebrafish; and Cas9 is injected to the zygote leading to the activation of scarring process (Alemany A. et al., 2018). A very similar method is LINNAEUS (lineage tracing by nuclease-activated editing of ubiquitous sequences) which targets cumulative RFP transgenes located at different genomic loci within the zebrafish by injecting Cas9 and sgRNA for RFP into a one-cell-stage embryo to initiate scar formation. Successful scarring process is confirmed by the loss of RFP signal. (B Spanjaard et al., 2018). To avoid the information loss by double-stranded breaks, various strategies have been developed. For instance, a CRISPR-editing method uses the enzyme nickase Cas9 (nCas9) that when fused with cytidine deaminase enzyme, substitutes C:G pairs to T:A in the 4-8 nucleotide region nearby the protospacer adjacent motif (PAM) without inducing double-stranded breaks (Hwang B. et al, 2019; Komor A.C et al., 2016). Another version of the classical CRISPR-editing system is MARC1 (Mouse for Actively Recording Cells 1) which uses homing guide RNA (hgRNA) in combination with Cas9 nuclease and guides the complex to its own hgRNA sequence repeatedly to introduce scars. A mouse that contains 60 different hgRNA sequences on different locations of its genome was crossed with a mouse that expressed the Cas9 protein to produce a developmentally barcoded mouse (Kalhor R. et al., 2018). Finally, Chan, Smith et al. (2019) studied both the cell lineage and cellular state of mice from the fertilization stage until the adulthood by using a high diversity traceable barcodes (molecular recorder). In this project, we took advantage of an adapted version of this molecular recorder to investigate cell lineages in TLS.

2.11. Molecular recording of trunk-like-structures

To investigate the relationship between cell lineage, trajectory and tissue morphogenesis, 060C11TSB8 Tracr cells (hereafter, referred to it as “Tracr cells”) were used in this project, which is a CRISPR/Cas9 based molecular recorder that consists of a target-site cassette and a three-guide cassette (**Figure 6**). The molecular recording strategy was adapted from Chan, Smith et al., 2019. The target-site cassette consists of a mCherry tag and a unique integration barcode (intBC) along with three target sites for the guide RNA (gRNA) which was delivered into the cell by PiggyBac Transposase expression vector and 10 unique barcode copies were integrated into different 3' Untranslated regions (UTRs) within the genome.

Rosa26 locus consists of LoxP sites flanking 3 poliA signals, Cas9, self-cleaving peptide T2A, GFP and three independent guide-RNA sequences controlled by three independent promoters (**Figure 6B**). The guide for site 1 (Bao_ade2) is controlled by the mU6 promoter, site 2 (Bri1) by hU6 promoter, and site 3 (whB) by bU6-2 promoter (**Figure 6C**). The system activates Cas9 activity upon pCAG-Cre-plamid nucleofection. Upon Cre-Lox recombination, the triple polyA tail signal between the LoxP sites is deleted, which allows the transcription of Cas9-T2A-GFP. Activation of Cas9 can be visually monitored by the GFP signal.

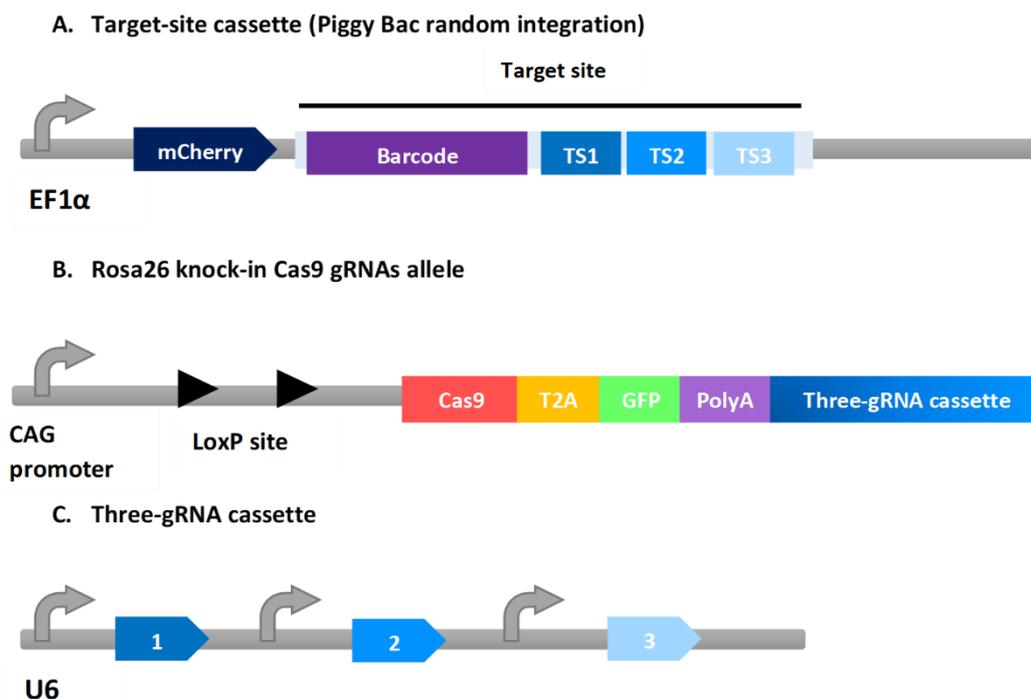


Figure 6 | Molecular recorder A. Target-site cassette consists of a mCherry tag, a unique integration barcode and three target sites for the guide RNA (gRNA). B. Rosa26 knock in with Cas9 gRNA alleles contain the LoxP site, Cas9, T2A, GFP, PolyA signals and gRNA cassette. C. Three-guide RNA are controlled by three independent promoters U6 mouse, U6 human and U6 bovine. **Adapted from Chan, Smith et al., 2019**

3. Aim of the project

The fundamental aim of this project is to understand the interplay between tissue morphogenesis and lineage decisions, i.e. “shape-to-gene feedback”. For this reason, a unique method of CRISPR/Cas9-based molecular recording that allows for retrospective lineage tracing of individual cells upon endpoint single-cell RNA-sequencing is used on the novel 3D model, trunk-like-structures, to understand the canonical lineage relationships of differentiated cells in developing embryo.

The aim of our work was divided into three micro-sections, firstly, to check if our system – Tracr cells were capable of creating scars over time that will enable lineage tracing. Secondly, to check whether Tracr TLS were competent to form canonical TLS that can recapitulate morphology and give rise to all cell states, without giving rise to any flaws in the model. Lastly, to construct lineage trees of the mammalian development using the Tracr TLS.

4. Materials and methods

4.1. Mouse Embryonic Fibroblasts (MEFs)

Mouse Embryonic Fibroblasts (MEFs) were seeded at least one day prior to the seeding of mouse Embryonic Stem Cells (mESCs). Briefly, 6 cm dishes (Corning 430166) were coated with 0,1% gelatine and incubated at room temperature for 30'. After incubation, gelatine was aspirated and dishes were washed with 2 ml PBS. Meanwhile, 6 million MEFs were thawed in the water bath for a minute and immediately transferred into 5 ml pre-warmed MEF medium (all media compositions can be found in **Table 3**). Cells were centrifuged for 5' at 1000 rpm, the supernatant was discarded, the cell pellet was re-suspended in 6 ml MEF medium and 1×10^6 cells transferred into each 6 cm dishes.

4.2. Mouse embryonic stem cell culture

mESCs were maintained on MEF coated 6 cm dishes and ES medium was refreshed daily to preserve their pluripotency (All media compositions can be found in **Table 3**). Cells were grown at 37°C with 5% CO₂ concentration. For cell passaging, cells were split with a ratio between 1:8 and 1:20 every alternative day depending on their confluency. Briefly, medium was aspirated, dishes washed once with PBS and 1 ml of pre-warmed trypsin was added. Cells were incubated for 5-10' at 37°C and then transferred into a falcon containing 3 ml ES medium in order to neutralize the trypsin. Cells were then centrifuged for 5' at 1000 rpm, the supernatant was discarded and the cell pellet was re-suspended in 3 ml ES medium. After pipetting several times to obtain a single cell suspension, around 1×10^6 cells were plated into each MEF coated dish.

4.3. Mouse embryonic stem cell lines

The mESCs lines was readily available upon the beginning of this internship. To investigate the relationship between cell lineages, trajectories and tissue morphogenesis, 060C11TSB8 Tracr cell line was used, which is a multi-purpose molecular recorder.

4.4. Freezing mESCs

To generate mESCs frozen stocks, cells were washed and harvested by trypsinization as mentioned under "Mouse embryonic stem cell culture". After centrifugation, cells were re-suspended in 500 µl 20% FCS-ES medium (All media compositions can be found under **Table 3**). Additionally 500 µl Freezing medium was added and mixed well. Cells were then transferred into cryovials and placed in a Styrofoam box at -80 °C. Cryovials were transferred into liquid nitrogen the next day for long term storage.

4.5. Generation of trunk-like-structures (TLS)

Pure mESCs are required for TLS generation, thus MEFs were depleted using gelatine coated dishes by serial transferring. Briefly, cells were trypsinized for 5' and re-suspended in 2 ml ES medium (see section: **Mouse embryonic stem cell culture**). Next, the re-suspended cells were incubated (37°C) for 25', 20' and 15' respectively

on each of the three gelatinized wells of a 6-well-plate. During each transfer, cells were re-suspended vigorously to obtain a single cell suspension. Once the mESCs were feeder-freed, they were washed once with 5 ml TPBS ($\text{MgCl}_2+\text{NaCl}$) and once with 5 ml N2B27 (NDiff27) (Takara) (All media compositions can be found under **Table 3**). It is important to note that N2B27 was equilibrated in the incubator (37°C) for 20 minutes prior to use at every step to make sure the medium was adjusted to incubator temperature and pH. Cells were then re-suspended in 250 μl of N2B27 after centrifugation for 5' at 1000 rpm. 550 – 600 cells were plated in 30 μl N2B27 medium in each well of a 96-well round bottom low attachment plate (Cellstar 96-well suspension culture plate or Costar 7007 ultra-low attachment 96-well plate). Cells were incubated for 48h to allow aggregation and then pulsed with 3 μM CHIR99021 (Chi) in 150 μl of N2B27. Twenty-four hours later, the medium was replaced by 150 μl of pre-incubated N2B27 medium. Between 92 and 96h after aggregation, gastruloids were checked under the microscope for a teardrop-like-shape where one pole has slightly begun to elongate. Once this characteristic shape was observed, gastruloids were then embedded in 5% Growth-Factor-Reduced Matrigel (Corning 356231). To this end, fresh N2B27 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 Matrigel 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 150 μl of old medium was removed from the aggregates. New medium with Matrigel (150 μl) was then added, and the cultures were returned to the incubator and further cultured at 37°C . TLS cultures were allowed to settle for at least 30 minutes before proceeding to further experimentation (e.g. live imaging). For the TLS-CL condition, 5 μM Chi and 600 nM LDN (CL) were supplemented to the medium.

4.6. Generation of Tracr TLS

Tracr TLS were generated as canonical TLS (see section: **Generation of trunk-like-structures (TLS)**). Once, the cells were feeder-freed, they were transferred into a 15 ml falcon tube, centrifuged at 1000 rpm for 5' and re-suspended in 500 μl ES medium. For nucleofection, Lonza Amaxa nucleofection kit (Cat # V4XP-3024) was used according to the manufacturer's instructions. For each nucleofection reaction, 1 million Tracr cells were counted, washed with 10 ml PBS, centrifuged and the supernatant

was discarded. Hereby, all the washing steps include centrifugation at 1000 rpm for 5' and elimination of the supernatant. A DNA mix was prepared for each condition (pCAG-Cre and pUC19-mock) by adding 18 μ l of supplement solution to 82 μ l of primary solution, and 10 μ g of pCAG-Cre plasmid (Cas9+ cells) or 5 μ g of pUC19 mock plasmid (Cas9- cells). Then, the 100 μ l DNA mix was added to the cells, mixed well by pipetting, transferred into aluminum cuvettes and placed into the Nucleofector™ Device and mESCs nucleofection program (CG104) was used. Cells were then transferred into 10 ml N2B27 medium, centrifuged and re-suspended in 1 ml N2B27 medium. Cells were counted again and diluted to 550 cells in 30 μ l N2B27 medium for the Cas9+ cells and 250 cells in 30 μ l N2B27 medium for Cas9- cells. Cells were plated as mentioned in the protocol above. The remaining cells were plated in a 6-well-plate to check nucleofection and recombination efficiency. At 48h, 72h, 96h and 120h, the cells were treated with the same conditions in (see section: **Generation of trunk-like-structures (TLS)**).

4.7. Measurement of Cas9-GFP activation to check nucleofection and recombination efficiency

To ensure Cas9 activation in pCAG-Cre transfected cells, Tracr cells were feeder-freed and nucleofected with pCAG-Cre and pUC19-mock plasmids (see section: **Generation of Tracr TLS**). A 6-well-plate was covered with 2 ml gelatin, incubated for 5 mins, aspirated and washed with 2 ml PBS. Cas9 + and Cas9- cells were equally divided and plated in the 6-well-plate. At 0h cells were directly re-suspended in 200 μ l FACS buffer. Next, cells were transferred into a FACS tube with a 35 μ m mesh and BD FACSCelesta™ Flow Cytometer machine was used for the measurement.

For the consecutive days, cells were trypsinized with 100 μ l TrypLE, incubated for 10', and re-suspended in re-suspended in 200 μ l FACS buffer as mentioned above and the amount of GFP+ and GFP- cells were measured daily to establish transfection and Cre-mediated recombination efficiency.

4.8. Analyzing the reduction of GFP+ cells

Silencing experiment: Nucleofected cells were plated and grown in a 6 cm MEF coated dish. The next day, GFP+ cells were sorted at 24h and re-suspended in ES media. Cells were replated in 12-well-plate coated with gelatin and media was changed daily.

At 120h, cells were re-suspended in FACS buffer (see section: Measurement of Cas9-GFP activation to check nucleofection and recombination efficiency) and GFP+ cells were measured by BD FACS Aria II.

Differentiation experiment: Nucleofected Cas9+ cells were plated equally in a 12-well-plate and treated with ES + LIF and N2B27 media. Cells were trypsinized and amount of GFP+ cells were measured daily from 0h-120h (see section: Measurement of Cas9-GFP activation to check nucleofection and recombination efficiency) to analyze the reduction in GFP+ cells between both the treatments.

4.9. Measurement of Cas9-GFP activation in Tracr TLS

To measure the Cas9-GFP activity in TLS, between 115 -120h, the structures were manually picked and transferred to a U-bottom-96-well plate filled with 150 μ l of PBS-BSA (5%) using a p200 pipette with the tip cut-off at the 50 μ l mark and the pipette set to 20 μ l to avoid large amounts of Matrigel transfer. After all the structures were transferred, the PBS-BSA was removed with a multi-channel-pipette set to 150 μ l and 50 μ l of TrypLE were added to each well and the 96-well-plate was incubated at 37°C for 15'. Next, 100 μ l of FACS buffer was added to each well and the structures were pipetted up and down 20 times to obtain a single cell suspension. Percentage of Cas9-GFP+ cells in each Tracr TLS were measured with BD FACS Celesta™ Flow Cytometer using the HTS mode (plate reader).

4.10. Resampling Cas9+ cells to assess consecutive editing (also known as “*Luria-Delbrück-like*” assay)

Tracr cells were depleted from MEFs and nucleofected in the same way described in (see section: **Generation of Tracr TLS**). After nucleofection, 800k cells were plated in a MEF coated dish while, while 200k cells were equally divided into two tubes for future RNA and genomic DNA (gDNA) isolation. In particular, for RNA samples, cells were re-suspended in 350 μ l RLT buffer + 1% β -Mercaptoethanol and snap frozen, whereas for gDNA extraction, cells were centrifuged for 5' at 1000 rpm, the supernatant was removed and the pellet snap frozen. From 0h – 120h, the cells were daily washed once with PBS, trypsinized, half of them were plated back into a new MEF coated dish, while

the rest was re-suspended in FACS buffer, centrifuged for 5' at 1000 rpm, finally, the supernatant was aspirated and cells were re-suspended in 1 ml FACS buffer. BD FACS Aria II was used to sort 60k-80k GFP+ cells for DNA extraction and frozen as a dry pellet; whereas for RNA extraction, 60k-80k cells were sorted in RLT buffer and frozen as described previously. (See Phenol-Chloroform gDNA extraction and, RNA extraction and cDNA synthesis

4.11. Phenol-Chloroform gDNA extraction

To yield high-purity gDNA, phenol-chloroform extraction was performed. Briefly, 200 µl cell lysis buffer (see table 3 for buffer composition) were added to the frozen cell pellets and the samples were incubated at 55°C with moderate shaking overnight to allow the enzymatic digestion of proteins and RNA. The next day, 200 µl of phenol:chloroform:isoamyl acid (25:24:1) (Thermo Scientific™ Cat. # Thermo ref: 15593-031) were added on the samples, vortexed for 2-3' and centrifuged for 10' at 13,000 rpm at RT to obtain two separate phases where the lipids and cellular debris accumulate in the organic phase while the isolated DNA in the aqueous phase. The upper aqueous phase was transferred carefully into new tubes and ethanol precipitation was performed by adding 1/10 volume of 5M NaCl, 1 µl glycogen and 2.5 volume of ethanol to the samples. The tubes were incubated at -80°C for 1 hour and -20°C for another hour. Next, samples were centrifuged at 13000 rpm for an hour at 4°C. The supernatant was removed by slowly inverting the tubes and drying them on a paper towel. Precipitated DNA was washed with 1 ml 70% ethanol followed by centrifugation for an hour at 4°C. The supernatant was removed as described above and the pellets were air dried for 15 mins at RT. Finally, the pellets were eluted in 12 µl EB buffer.

4.12. RNA extraction and cDNA synthesis

RNA extraction was performed using the RNeasy Micro Plus Kit (*Qiagen Cat. # 74034*) according to the manufacturer's instructions and the samples were eluted in 12 µl RNase free water. RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific™) was used following the manufacturer's instructions using oligo dT to enrich for the polyA+ RNA fraction.

RNA extraction and reverse transcription were done for single TLS and the *Luria-Delbrück-like* experiments. For single TLS and *Luria-Delbrück-like* experiments 12 µl and 11 µl RNA were used for reverse transcription experiment respectively.

4.13. IntBC amplicon PCR and Library Preparation

To amplify the intBCs in individual TLS, MangoMix™ Bioline reaction mix was used. The reaction consisted of 10 µl Mango Mix, 1 µl F primer, 1 µl R primer 10µl cDNA and 3 µl H₂O. (For conditions see Experiment: **Amplification of target-site-cassette, Table 3**)

Whereas, to amplify the intBCs from cDNA and gDNA in *Luria-Delbrück-like* experiment NEBNext® Ultra™ II Q5® Master Mix and a total of 100 µl reaction was prepared. This reaction was then divided into into 4 wells to preserve complexity and avoid amplification biases. The reaction consisted of 50 µl 2X Q5 Master Mix, 5 µl F primer, 5 µl R primer (Table 2.), 5 µl cDNA/ 5 µl gDNA and 35 µl H₂O. (For conditions see Experiment: **Amplification of library, Table 3**)

Prior to use for both experiments, AmpureXP magnetic beads (Beckmann coulture) were equilibrated at RT on a rotator for 30'. The 4 wells belonging to the same sample were pooled back into 1 one well of a PCR strip (final volume 100ul) and 0,55x (55ul) magnetic beads were added on each sample to capture fragments bigger than 450 bp. Samples were pipetted up and down 10 times, incubated for 5 minutes at RT then put on a magnetic rack for 3 minutes. Since intBC amplicons are in the range of 300-400 bp, the supernatant was kept and transferred into a new well and 0.8x (25ul) magnetic beads cleanup was performed. The samples were pipetted up and down 10 times, incubated for 5 mins and placed on a magnetic rack again for 3 mins. Since the beads bound fragments of 200 bp or bigger, the supernatant was discarded and the beads were washed with 200 µl 75% ethanol 2 times for 20 seconds. The supernatant was removed and the tubes were quickly spun down, placed back on the magnetic rack and the residual ethanol was removed with 10 µl pipette. The tubes were removed from the magnet, and amplicons were eluted from the beads using 30 µl water. A second PCR reaction was performed to prepare the samples for Illumina sequencing using NEBNext® Ultra™ II Q5® Master Mix and NEBNext® Multiplex Oligos for Illumina® (Cat. no E7500S). The reaction conditions were the same as mentioned above except 30 µl of the amplicons as input material were used. After amplification,

the 4 reactions for each sample were pooled into a PCR tube and 0.8x magnetic bead cleanup was applied on each sample to get rid of remaining primers, as described above. The samples were eluted in 15 µl nuclease free water. The concentrations of the amplicon libraries were measured using Qubit™ dsDNA HS Assaykit (Cat. # Thermo Q32851) according to the manufacturer's instructions. To assess the overall scarring of single TLS and correct library generation for *Luria-Delbrück experiment*, D1000 ScreenTape assay was used in Agilent 4200 TapeStation system (G2991AA) according to manufacturer's instructions.

4.14. Single-cell RNA-seq analysis of Tracr TLS

120h Tracr TLS were generated as described above. Two individual Tracr TLS were selected based on their morphology (somatic segmentation together with neural tube formation) and used for this protocol. Briefly, the two structures 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. Tracr TLSs were then dissociated in 200µl TrypLE Express (Gibco) for 15 minutes 25 minutes at 37°C, with pipetting at 5 minutes intervals. The cell suspension was filtered using Scienceware Flowmi Cell Strainers, 40µm. Cells were then washed twice with 1ml 1xPBS/0.4%BSA with centrifugation steps performed for 5 minutes at 1200rpm in low DNA bind 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 (individual Tracr TLS). 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. One third of the double stranded cDNA obtain during the library preparation was stored for future amplification and sequencing of the target site cassette. Libraries were sequenced with a minimum of 230 million paired end reads according to parameters described in the manual.

4.15. Whole-mount immunofluorescence

TLS were picked and transferred using a p200 pipette with the tip cut-off at the 50 μ l mark, into a 96-well-U-bottom plate and/or to ibidi 8-well glass-bottom plates (ibidi 80827). Next, structures were washed twice with 150 μ l PBS⁺⁺ BSA, twice with 150 μ l PBS and fixed with 4% PFA for 75 minutes at 4°C on a rocking stage. PFA was removed and the structures were washed 3x with PBS, permeabilized 3x with PBST (see table 3) and blocked overnight at 4°C using 5% fetal calf serum/PBST (blocking solution). Note: all the washes were done with 10' time interval and incubated on a rocker at all times. The next day, the structures were stained with the primary antibody for 48-72h at 4°C (**Table 1.**), then washed 3x with blocking solution, 2x with PBST and incubated in blocking solution o/n at 4°C. Next, the secondary antibody incubation was performed for 24 h at 4°C. TLS were then washed 3x with blocking solution, 2x with PBST and 1x with PBS. Finally the structures were incubated in DAPI (0,02%) o/n, washed 1x with PBS the next day and stored in PBS at 4°C. All the antibodies compositions can be found under **Table 1**

4.16. Tissue clearing

Structures were cleared with RIMS to provide good optical clarity for fluorescent imaging. To this end, the stained structures were washed 2x with PBS for 10', fixed with 4% PFA for 20', washed 3x with 0,1M phosphate buffer and finally the samples were incubated in RIMS solution. (All media compositions can be found under **Table 3**)

Table 1. List of primary and secondary antibodies used during the experiment

<i>Primary antibody</i>	Species	Dilution (in blocking solution)	Company
Sox17	Mouse	1:250	Abcam
Sox2	Goat	1:200	Abcam
<i>T</i>	Rabbit	1:200	Cell Signaling Technology
Cdx2	Goat	1:200	Biogenex

Secondary antibody	Wavelength [nm]	Dilution (in blocking solution)	Company
Donkey-anti-goat	546	1:250	<i>Invitrogen by Thermo Fisher Scientific</i>

Donkey-anti-rabbit	647	1:250	<i>Invitrogen by Thermo Fisher Scientific</i>
DAPI (in PBST-X)	461	1:5000	<i>Roche Diagnostics GmbH</i>

4.17. Imaging

Brightfield imaging was performed using the Zeiss imaging platform Celldiscoverer 7, running under ZEN blue version 3.1 to live image the gastruloids and TLS in 96-well-plates between 96 – 120 hours and also, to image the fixed TLS using the appropriate filters to view the fluorescent reporters and antibodies. Structures were imaged using a 5x/0.35 Plan-Apochromat-Objective with a 1x post magnification and 1x1 camera binning (Axiocam 506) resulting in a lateral resolution (xy) of 0.916 $\mu\text{m}/\text{pixel}$. As 96-well-plates neither allow hardware focusing, nor does the brightfield imaging allows a robust software focusing strategy, thus, each focusing z-position had to be adjusted individually. Prior to the experiment, all 96 imaged positions were manually adjusted in xyz. Typical field of view xy-sizes were 2.52 x 2.02 mm. To ensure a proper representation of the entire structure, experiments were set up in z-Stacks. Typically 13 slices with a z distance of 19.56 μm were captured, resulting in a full z-depth of 235 μm . For kinetic measurements 3D time course were acquired, over a time period of 12-24 h in 30 min intervals. All images were acquired under full environmental control including 100 % humidity, 5 % (v/v) CO₂ at 37°C.

4.18. Image processing

The processing/analysis consist of three major ZEN blue modules namely “Image Analysis”, “Intellesis” and “Open Application Development” as well as basic functions in the ZEN blue desk version. Before images were analyzed, single z-plane were extracted from individual scenes. In a second step, images were binned by averaging 4 pixel in x and y direction. The processed images were then, pumped into a ZEN image analysis pipeline were a previously trained algorithm (ZEN Intellesis), which classified each individual pixel as foreground (TLS) and/or background (image background). The algorithm generated a probability map which was implemented in ZEN image analysis module to segment regions of interest. All of the following steps were written in short custom macros in the ZEN open application development platform and ran over specified data sets/experiments. The underlying decisions, thresholds

filter were for all analyzed on identical experiments enabling a real side by side comparison of data sets from individual experimental iterations and/or replicates.

4.19. Image Analysis

The analysis module used the probability map and assigned a region of interest (TLS) from a value confidence of at least 60 %. The resulting regions of interest were filtered by an area, min/max area ranging from 25000-750000 μm^2 . Features of interest were written into text files and stored by the routine geometric features (area, perimeter, roundness, circularity, feret max, feret min, feret ratio), intensity features (compactness, convexity, radius, sd, mean, sum, range, max/min bright intensities, intensity pixel count) , and meta data features (image document name, ID, image index time, image relative time, parent ID, image scene). Only a subset of these parameters were plotted in this work (perimeter, area, circularity).

4.20. Data analysis

An R script was developed to analyze the features of interest. To this end, the correlation between Cas9-GFP+ cells of each structure and changes in geometric features of TLS at 120h were analyzed using “ggpubr” package. In addition, the time lapse changes in geometric features of Cas9+ cells and Cas9- cells were compared. The plots for both the analysis were created using the “ggplot2” package. Due to the wrong masking of the structures by the ZEN blue module image analysis program, some elongated structures that resemble “the snail shell/ fetal-position” was recognized as a sphere and scored as 0, which indicates circularity (in contrary, 1 indicates elongation). For this reason, we filtered the snail-shell-shaped structures that was considered to be as non-elongated by the program, to remove any biases.

4.21. Morphometric analysis of TLS

TLS were scored at 120h based on displayed features such as axes of elongation, segmentation and somite bilaterality. The number of structures were quantified for each condition: Cas9+ & Cas9-, and scored as “elongated” if they developed an axis, otherwise scored as “non-elongated”. Next, the structures were scored as “one axis”, if they exhibit a single axis, and “multiple axes” if they displayed two or more axes.

Moreover, TLS were defined as “segmented” when at least three neighboring segments had developed along the antero-posterior axis. To clarify, segments were interpreted as substructures that displayed i) indentations and ii) opposite curvatures at segment borders. Finally, if a structure showed segments on one side, it was defined as “unilateral” and if displayed on both sides, as “bilateral”. Structures that grew out of focus or out of scene were excluded from being analyzed for any of the categories.

4.22. Single-cell RNAseq computational analysis

Single-cell RNAseq computational analysis for the two Tracr TLS has been performed by Dr. Helene Kretzmer.

4.23. 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.

4.24. Quality control

The initial quality control was performed using scanpy. 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).

4.25. Cluster Assignment

Each cell of the Tracr TLS was then clustered with respect to TLS cell states (Veenliet, Bolondi et al., 2020, Science, in press). 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 (Tracr TLS, 120h TLS from Veenliet, Bolondi et al., 2020, Science, in press), 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 inte-grated dataset, a PCA was run to then calculate a joint UMAP (dims = 1:10).

4.26. Pseudogenes identification

Two artificial chromosomes consisting respectively of the barcode cassette and Cas9-GFP sequences were added to the reference genomes and corresponding gene annotations added to the “gene annotation file”. Processing with Cell Ranger was run as before using this genome reference and gene annotation to obtain read counts for all genes including the barcode cassette and Cas9-GFP. This allowed to map these two transgene expression in the sampled single cells.

4.27. Plotting

Plots were generated with GraphPad Prism 7, R 3.6.0 and R 3.6.1, Python, and PlotTwist.

4.28. FACS analysis

FlowJo version 10.6.2 was used to analyze the obtained FCS files. To obtain the desired population hierarchy, gatings were applied on the events in the following order: FSCA vs SSCA to obtain all the cells-of-interest and eliminate the outlier cells and debris, FSCA vs FSCW to eliminate duplets, SSCA vs SSCW to obtain single cells furthermore. Next, to view the mCherry and the GFP population, PE-CF594 and Alexa Fluor 488 channels were used.

The threshold was set to 1000 cells and the table was created using the following features: Well ID, Tube name, Freq. of parent and number of events.

Table 2. PCR conditions and primers used during the experiments

Experiment	Primer	Primer sequence	PCR program
Amplification of target-site-cassette	Forward primer	GAATCCAGCTAGCTGTGCAGC	96 °C 2 s
	Reverse primer	GTTAACACCTGCAGGAGCGATG	96 °C 10 s 56 °C 30 s 30 cycles 72 °C 30 s 72 °C 5 s 4 °C ∞
Library Preparation (Two runs - amplification of target site + preparation for Illumina sequencing)	Forward primer	TCTACACTCTTCCCTACACGACGCTC TTCCGATCTGAATCCA	98 °C 30 s
	Reverse primer	GCTAGCTGTGCAGC	98 °C 10 s 65 °C 30 s 12 cycles 72 °C 30 s
	Illumina Universal primer	GTGACTGGAGTTCAGACGTGTGCTCT TCCGATCTGTAAACAC CTGCAGGAGCGATG	72 °C 5 mins 4 °C ∞
	Illumina Indexed primers	AATGATACGGCGACCACCGAGATCTA CACTCTTCC CTACACGACGCTCTTCCGATC-s-T	
		CAAGCAGAAGACGGCATAACGAGATT GNNNNNNGTACTGGAG TTCAGACGTGTGCTCTTCCGATC-s-T	

Table 3. Media and solution compositions.

Media / Solution	Composition and Products
-------------------------	---------------------------------

ES cell medium (15 % FCS, 1000 U/ml LIF)	400 ml Knockout Dulbecco's Modified Eagle's Medium (DMEM) (4500 mg/ml glucose, w/o sodium pyruvate) (<i>Gibco</i>) 75 ml ES cell tested FCS 5 ml 100x L-glutamine (200 nM) (Lonza #BE17-605E) 5 ml 100x penicillin (5000 U/ml)/ streptomycin (5000 µg/ml) 5 ml 100x non-essential amino acids (Gibco #11140-35) 1 ml 500x β-mercaptoethanol (5mM, 1000x Invitrogen) 5 ml 100x nucleosides (Chemicon) [add 4 µl Murine Leukemia Inhibitory Factor (LIF) (10 ⁷ U/ml) fresh to each 40 ml aliquot]
ES medium containing 20 % FCS	ES medium + another 5 % FCS
FACs buffer	100 ml PBS + 2% FCS + 5 mM EDTA
Freezing Medium	15 ml Knockout DMEM (4500 mg/ml glucose, with sodium pyruvate) 5 ml FCS 5 ml DMSO
Gelatine (0.1 %)	20 ml 2 % gelatine in 380 ml cell culture water
MEF / mouse embryonic fibroblast) medium (10 % FCS)	500 ml regular Dulbecco's Modified Eagle's Medium (DMEM) (4500 mg/ml glucose, w/o sodium pyruvate) (<i>Gibco</i>) 55 ml regular fetal calf serum FCS 5.5 ml 100x L-glutamine (200 mM) (Lonza#BE 12-605E) 5.5 ml 100x penicillin (5000 U/ml)/ streptomycin (5000 µg/ml)
N2B27	NDiff 227 (Supplemented with N2 and B-27) (<i>Takara</i>)
PBS with MgCl₂/NaCl	Dulbecco's Phosphate Buffered Saline with MgCl ₂ and NaCl (<i>Sigma</i>)
PBS without MgCl₂/NaCl	Dulbecco's Phosphate Buffered Saline without MgCl ₂ and NaCl (<i>Gibco</i>)
PBS/BSA	PBS + 0.5% BSA
PBST-X	0.5 % Triton-X in PBS
Phosphate buffer (0.02M)	PB, 0.025M NaH ₂ PO ₄ + 0.075M Na ₂ HPO ₄ , pH 7.4
RIMS	40 g of RIMS133% w/v Histodenz (Sigma-Aldrich D2158) in 30 ml 0.02M PB + 0.01% Tween
Tryple	Trypsin-EDTA (0.05 %)

Trypsin	Trypsin-EDTA (0.05 %), phenol red (<i>Gibco</i>)
----------------	--

5. Results

5.1. 2D Scarring dynamics of Tracr cells

CRISPR-based barcode editing technologies have facilitated the construction of extensive transcriptional atlases of differentiated cell types (reviewed by Wagner & Klein, 2020; Kester L. & Van Oudenaarden, 2018). Directing Cas9 nuclease to a specific integrated DNA barcode within a locus creates indels (scars). The descendants of the barcoded cell will carry the same scars as the mother cell and this feature allows to precisely trace cell populations over time and differentiation (Jao et al., 2013). Given the potential of such a technology, we employed a multi-purpose molecular recorder that consist of a target-site cassette and a three-guide cassette (optimized from Chan, Smith et al., 2019 protocol, **(Figure 6)** to trace cell lineages during TLS development. The target-site cassette consists of an mCherry fluorescent protein harboring in its 3'UTRa unique integration barcode (intBC) followed by three target sites for the guide RNAs (gRNA) **(Figure 6A)**. Rosa26 locus consists of LoxP sites flanking 3 poliA signals, Cas9, self-cleaving peptide T2A, GFP and three independent guide-RNA sequences controlled by three independent promoters **(Figure 6B, 6C)**.

To make the system compatible with TLS generation from mouse embryonic stem cells, we established an inducible system which activates Cas9 activity upon pCAG-Cre-plamid nucleofection. Upon Cre-Lox recombination, the triple polyA tail signal between the LoxP sites is deleted, which allows the transcription of Cas9-T2A-GFP. Activation of Cas9 can be visually monitored by the GFP signal.

To assess the nucleofection and recombination efficiency, Tracr cells were subjected to flow cytometry analysis at 0h, 24h and 120h to monitor the GFP signal **(Figure 7A)**. Immediately after nucleofection (t = 0h) nearly no cells express Cas9-GFP, which was expected given the temporal delay in activation of the CRISPR/Cas9 system. Importantly, intBC expression, as shown by mCherry activity, was detected in almost all cells **(Figure 7B)**. As anticipated, GFP+ cells were observed in a large proportion

of the population at 24h, with 87.4% of the total cells expressing GFP signal. Surprisingly, the number of GFP+ cells decreased to 73.7% at 120h, whereas we had expected an increase in the Cas9-GFP+ cells over time (**Figure 7B**)

In conclusion, the flow cytometry data suggested that the Tracr ESC line was a functional molecular recording, since barcodes were recovered in every cell and concomitant Cas9 activation was detected upon nucleofection of Cre, with a large fraction of cells being Cas9+ at 24h. However, the unexpected decrease of Cas9+ cells at 120h prompted us to test two possible scenarios that could explain this phenomenon: outgrowth competition of non-transfected Cas9- cells or endogenous silencing of Cas9-GFP transgene.

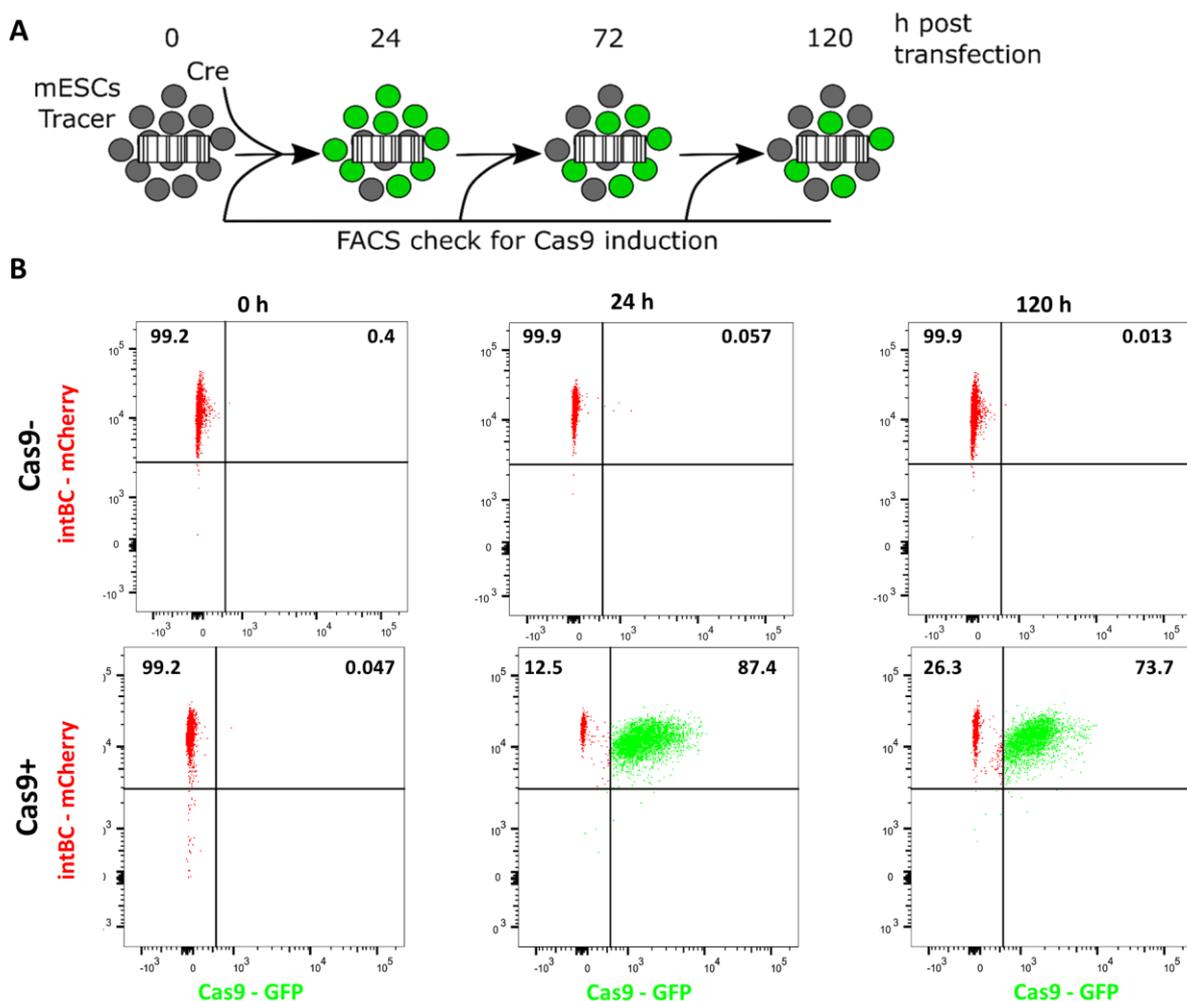


Figure 7 | Nucleofection and recombination efficiency analysis of Tracr cells. A. Schematic overview: Tracr cells were nucleofected with pCAG-Cre plasmid and analyzed at 0, 24, and 120h. B. Flow cytometry analysis of Cas9+ and Cas9- cells to detect intBC - mCherry and Cas9 - GFP signals. intBC were recovered in all cells and Cas9 nuclease was successfully activated. Schematics drawn by Adriano Bolondi using Inkscape.

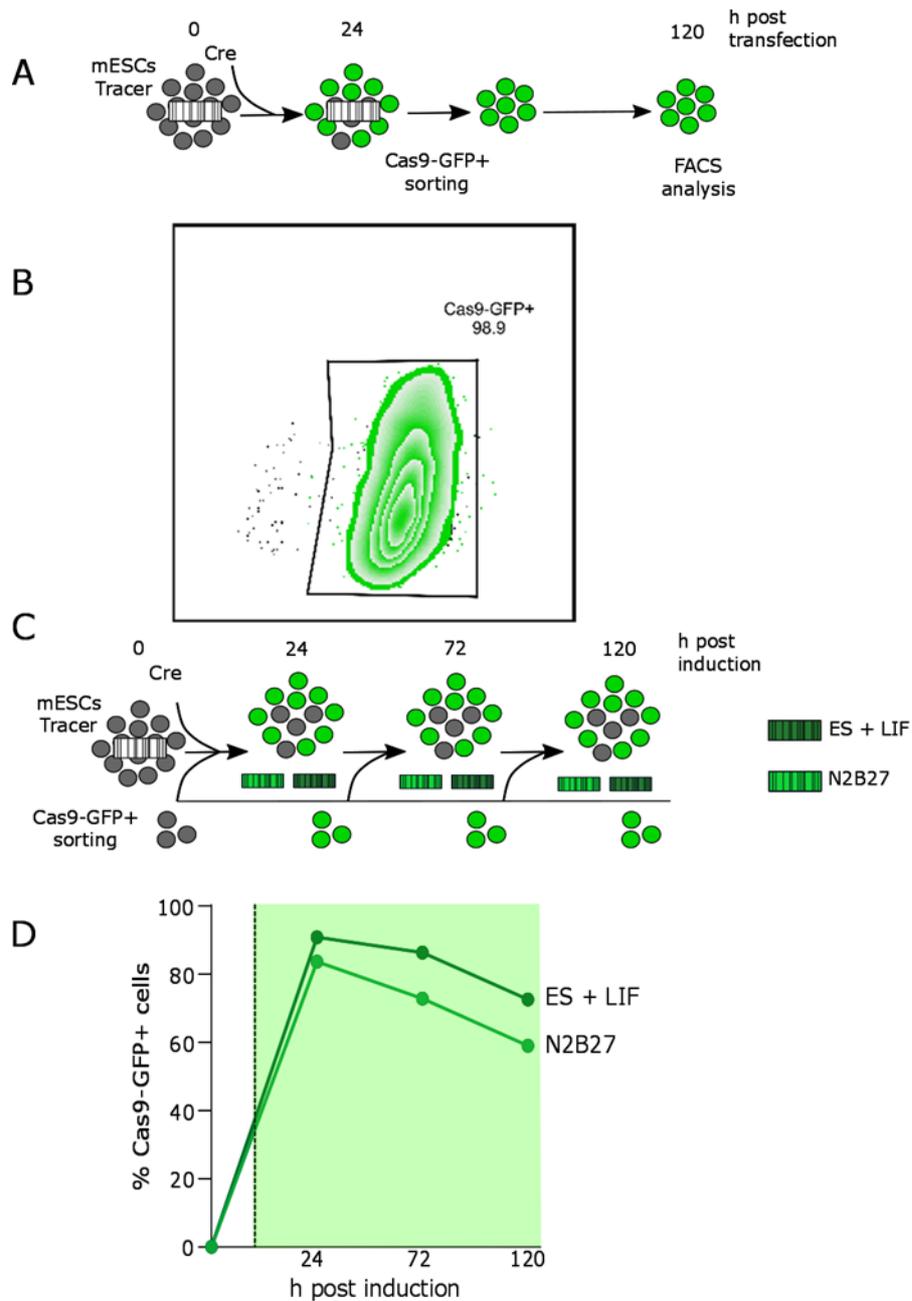


Figure 8 | Analyzing the reduction of GFP+ cells at 120h. A. Schematic overview: Tracr cells were nucleofected with pCAG-Cre plasmid; GFP+ cells were sorted after 24h, plated back on the dish and further grown until 120h; at 120h cells were analyzed for GFP expression. B. Intensity graph of Cas9-GFP+ cells at 120h. Almost all Tracr cells express GFP signal rejecting the silencing hypothesis. C. Cas9+ cells were grown with ES+LIF or N2B27 media after nucleofection. The two population were analyzed at 0, 24, 72, 120h after nucleofection to assess the amount of GFP+ cells D. Dark green line indicates ES + LIF treatment; Light green line indicates N2B27 treatment; GFP+ cells treated with differentiation medium show 20% decrease in GFP+ cells.

5.2. GFP- cells outgrow GFP+ cells over time

To distinguish between these two scenarios, GFP+ cells were FACS-purified at 24h and replated (**Figure 8A**). At 120h, cells were subjected to flow cytometry and no

reduction was observed in GFP+ cells which reasonably led us to reject the silencing hypothesis (**Figure 8B**).

Since differentiation processes can induce stress to cells as observed by the downregulation of ESC markers due to an increase in ROS levels during pluripotency-exits (Ae-Ri Ji, et al., 2010), we next tested whether creation of scars impacted the reduction of GFP+ cells in the overall population during differentiation compared to pluripotency. For this reason Cre nucleofected Tracr cells were then grown under two media conditions: ES medium and N2B27 differentiation medium, and subjected to flow cytometry analysis at 0, 24, 72, 120h to monitor the GFP signal (**Figure 8C**). Flow cytometry results showed nearly a 20% decrease in GFP+ cells in the population treated with differentiation medium over time in comparison to ES + LIF treated cells (**Figure 8 D**). The stress combination caused by DSB and differentiation process slows down the cell cycle further, and could thus lead to the GFP- fit cells outgrowing the Cas9-GFP+ cells.

5.3. Resampling Cas9+ cells to assess consecutive scarring activity (“*Luria-Delbrück-like*” assay)

It is essential to check if the intBCs accumulate scars over time and even more important to assess if these scars can be related to each other for lineage tree construction. Thus, to test if consecutive scarring activity over time was taking place, we designed an assay inspired by the classical *Luria-Delbrück experiment* (Luria & Delbrück, 1943). With 24h intervals, a subset of the GFP+ cells were FACS-purified from the same starting population of Cas9+ cells and pooled for gDNA and RNA extraction while the remaining non-purified, mixture of Cas9-GFP+ and Cas-GFP- cells were re-plated to allow the further accumulation of scars in the intBC over the next days (**Figure 9A**). The reason why we sorted GFP+ cells is to sequence only the possibly scarred-cells. Next, intBC regions were amplified from gDNA and cDNA in order to measure the degree of editing/scarring using a screentape assay and intBC-amplicon sequencing.

The screentape assay demonstrated continuity of the scarring process with deletions in a range of 1-50 nucleotides detected in all the samples sorted from 24h to 120h (**Figure 9B**) both at the gDNA and cDNA level. We conclude that Tracr cells accumulate scars on their intBCs over

time and that these profiles increase in complexity over time, having the potential to be used to trace cell lineages.

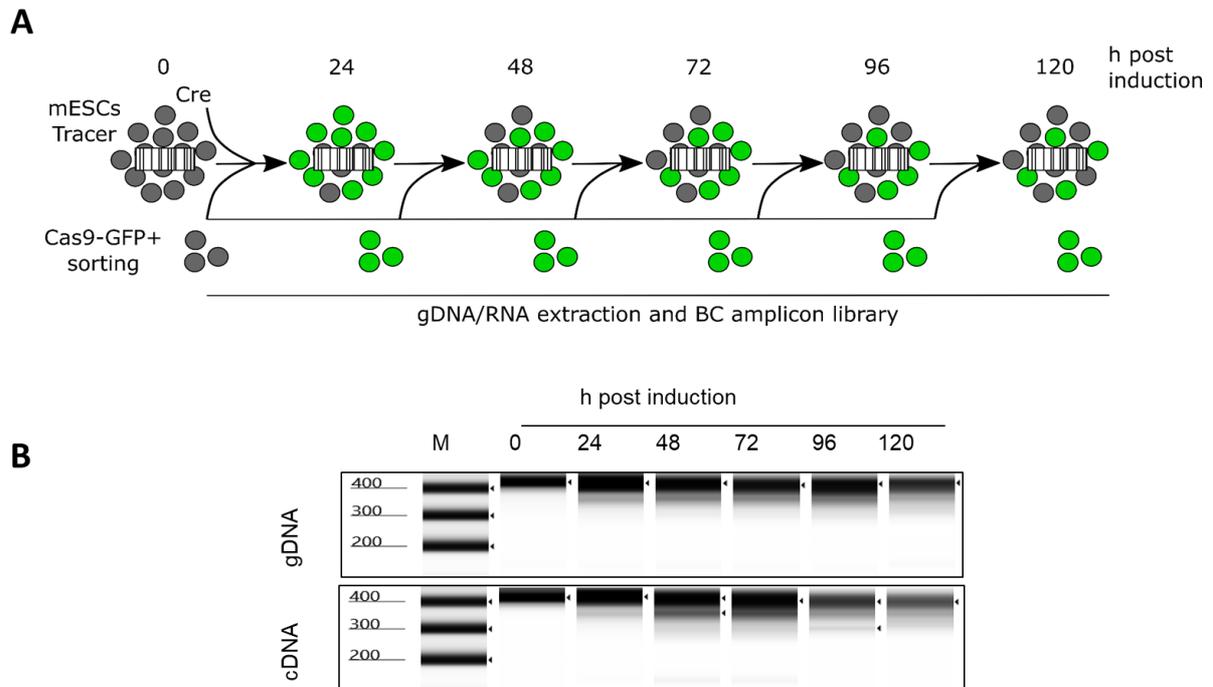


Figure 9 | A. Schematic: After nucleofection, a subset of the GFP+ cells were sorted with a 24h interval from the same starting population of cells and pooled for gDNA and RNA extraction while the remaining non-purified cells were re-plated to allow the accumulation of scars in the intBC to continue to investigate if consecutive scarring occurred. B. Screentape assay of amplified intBC from the gDNA and cDNA samples between 0h to 120h. Accumulation of deletions was observed in all samples. Schematics drawn by Adriano Bolondi using Inkscape.

5.4. 3D Scarring dynamics of Tracr TLS

It has previously been described that embedding of 96h gastruloids in 5% matrigel resulted in TLS, with segmentation of the mesodermal domain and formation of a neural tube like structure (Veenvliet, Bolondi et al., 2020, Science, in press). Since three-dimensional ES cell models capture more meaningful biological trends for cell lineage tracing, upon validation of the Tracr cell in 2D culture, we moved to 3D culture and generated TLS from Tracr ESCs.

According to Van den Brink et al., 2014, aggregate size of $\sim 300 \pm 100$ cells, develop into gastruloids upon treatment with Chiron (Van den Brink et al., 2014) and TLS were

also successfully generated from the same amount of starting cells (Veenvliet, Bolondi et al., 2020, Science, in press). Thus, we generated TLS from Tracr cells using 200 cells based on the original protocol. However, the structures either failed to elongate or did not develop proper paraxial mesoderm compartments (data not shown). To further optimize the protocol, different aggregate sizes were tested by aggregating higher cell numbers (250, 300, 350, 400, 450, 500, 550, 600). Structures generated with 550 & 600 cells displayed proper axial elongation, neuroectodermal and mesodermal domain formation (**Figure 17**).

After adjusting the Tracr TLS generation protocol, we tested the scarring dynamics in TLS. To validate the activation of scarring in single Tracr TLS, individual structures were dissociated, RNA was extracted and after reverse transcription into cDNA, intBC regions were amplified to observe the editing profile of twelve individual trunk-like-structures by screentape assay. Additionally number of GFP+ cells for individual TLS were measured by flow cytometry analysis (**Figure 10A**). In addition, Tracr TLS # 1, 2, 3, 4, 6, 8, 9 and 11 libraries were prepared for intBC-amplicon seq. Similar to Tracr cells cultured in 2D, accumulation of deletions was observed in all the structures in a range of 1- 150 nucleotides (**Figure 10B**). Flow cytometry results revealed the fraction of GFP+ cells in TLS was 20-40% with a mean of 23%, which is lower than the percentages observed in 2D at 120h (**Figure 10C**). In conclusion, flow cytometry analysis in combination with screentape assay confirmed that Tracr TLS accumulate scars in intBCs in TLS. The lower number of GFP+ cells in Tracr TLS in comparison to Tracr cells may be explained by additional stress developmental programs execute to form an embryo-like-structure.

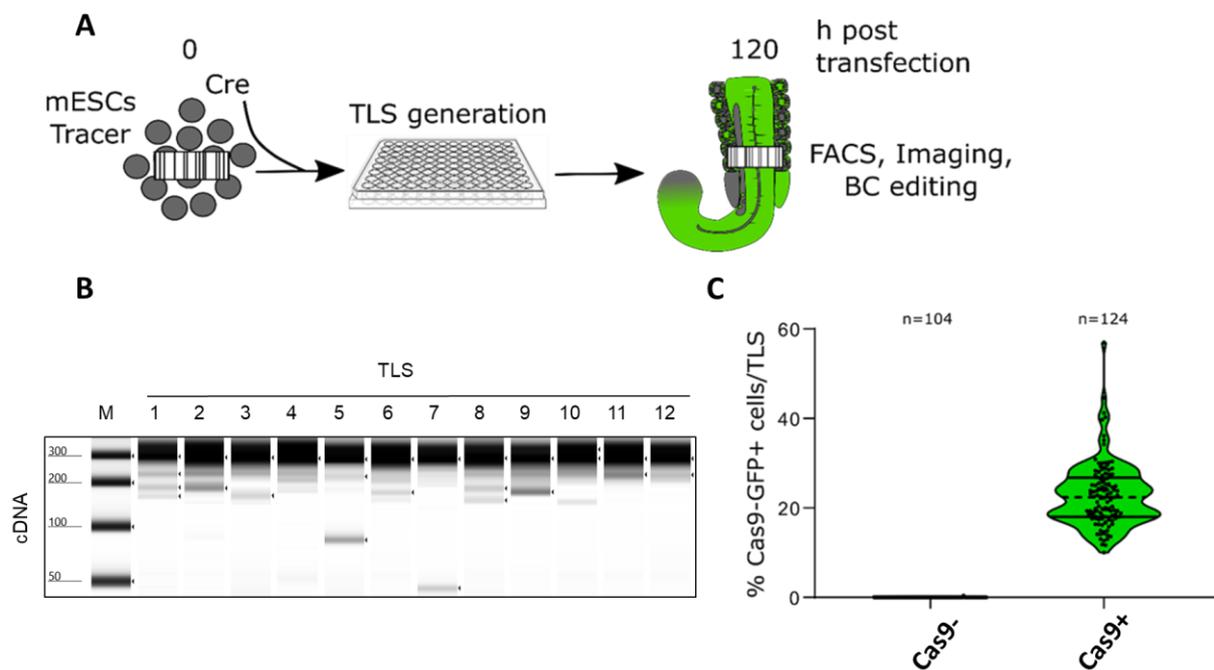


Figure 10 | Tracr TLS accumulate scars over time. A. Schematic: Tracr TLS were generated from Cas9+ and Cas9- cells and grown until 120h. B. Deletions were observed in all Tracr TLS further confirming the activation of scarring activity. C. Flow cytometry analysis show an average of 23% GFP+ cells present in each Cas9+ structure while none for Cas9- TLS. Boxes indicate interquartile range. End points of the violin plot represent minimum and maximum. Dots indicate individual Tracr TLS structures. Central dotted-line represent the mean. The two lines above and below the mean represent interquartile range.

5.6. Morphological characterization of Tracr TLS

After validating the scarring activity in Tracr TLS, we next analysed whether Cas9+ TLS display similar morphological features compared to Cas9- TLS. To this end, we scored the 120 h structures based on displayed features such as elongation, axes formation, segmentation and somite bilaterality. In both conditions, nearly all structures elongated (**Figure 13A**). In addition, about 94% of the Cas9+ structures formed a single axis, while nearly 30% of Cas9- TLS formed multiple axes. Half of the Cas9+ TLS displayed segmentation similar to results obtained by Veenvliet, Bolondi et al., 2020, Science, in press. However, only 7% of the Cas9- structures displayed segmentation (**Figure 13A**). As reported in Veenvliet, Bolondi et al., (2020), some structures failed to elongate and some formed multiples axes. Further analysis of Cas9+ segmented TLS revealed that 51% of the structures display bilateral segmentation (**Figure 13B**). Upon closer inspection, we noticed that Cas9- structures

showed a bias towards neural lineage and failed to develop paraxial mesoderm (**Appendix Figure 1**). The reason for this rather contradictory result may occur when the optimum number of cells used for aggregates was not adjusted. The protocol needs to be further optimized for Cas9- structures. In conclusion, these data show that axial elongation and segmentation efficiently occur in Tracr TLS, matching the statistics reported for wild-type TLS (Veenvliet, Bolondi et al., 2020, Science, in press).

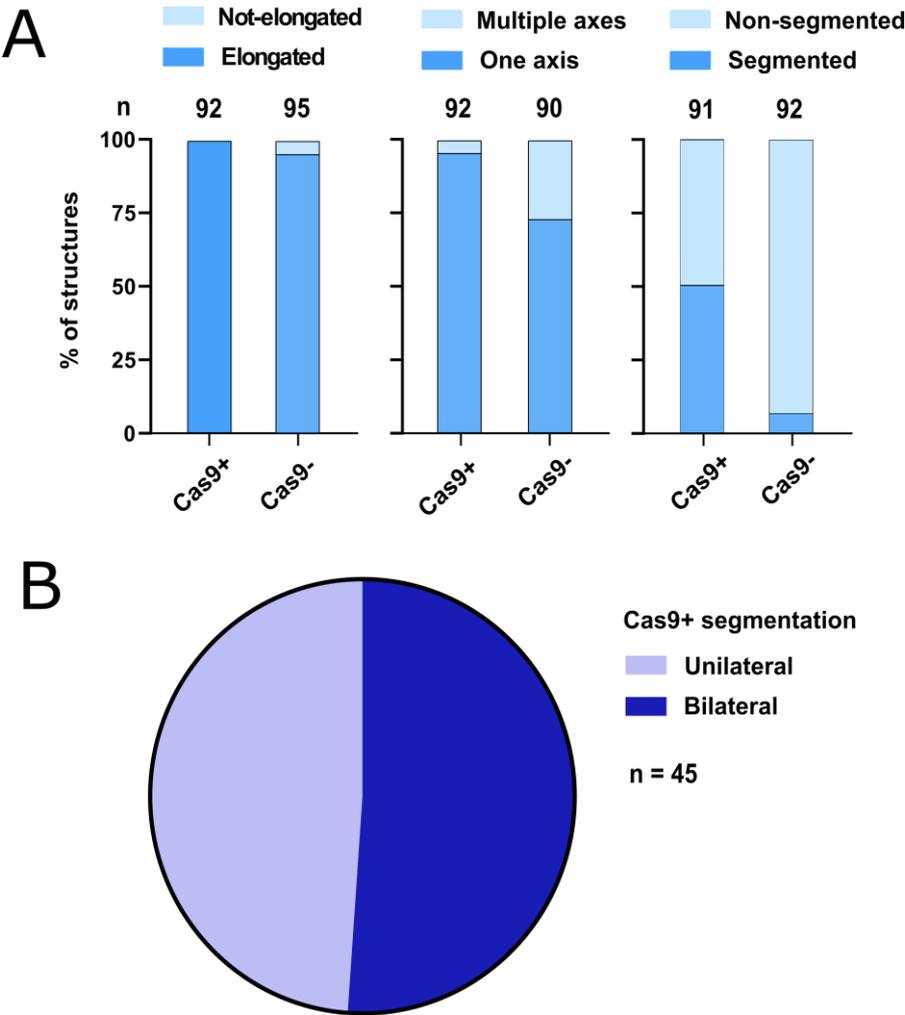


Figure 13 | Quantification of morphogenetic features in Tracr TLS A. The stacked bar plots show the percentage of Tracr TLS that display elongation, axes formation, segmentation in Cas9+ and Cas9- TLS. B. The pie chart displays the percentage of Cas9+ segmented TLS that formed unilateral- or bilateral somite-like structures. n = number of structures.

5.7. Scarring activity does not affect Tracr TLS morphology

We next performed a detailed morphometric analysis of 120h Tracr TLS to investigate whether the morphology of the structures is affected by the scarring. To this end, geometric features of 120h Tracr TLS were first measured using the ZEN blue image analysis module and plotted against the number of GFP+ cells as measured by flow cytometry. Importantly, we did not observe a correlation between the number of GFP+ cells per TLS and perimeter, area or circularity (**Figure 15**). These data show that the fraction of cells with scarring activity does not affect growth and overall morphology of the Tracr TLS.

Moreover to ensure that morphodynamic alterations were similar in Cas9- and Cas9+ Tracr TLS, we performed a time-resolved analysis of the same geometric features. Structures displayed similar morphogenetic changes over time as demonstrated by morphometric analysis (**Figure 16**).

From these data we conclude that scarring activity in the Tracr TLS did not affect the growth or morphology of the structures.

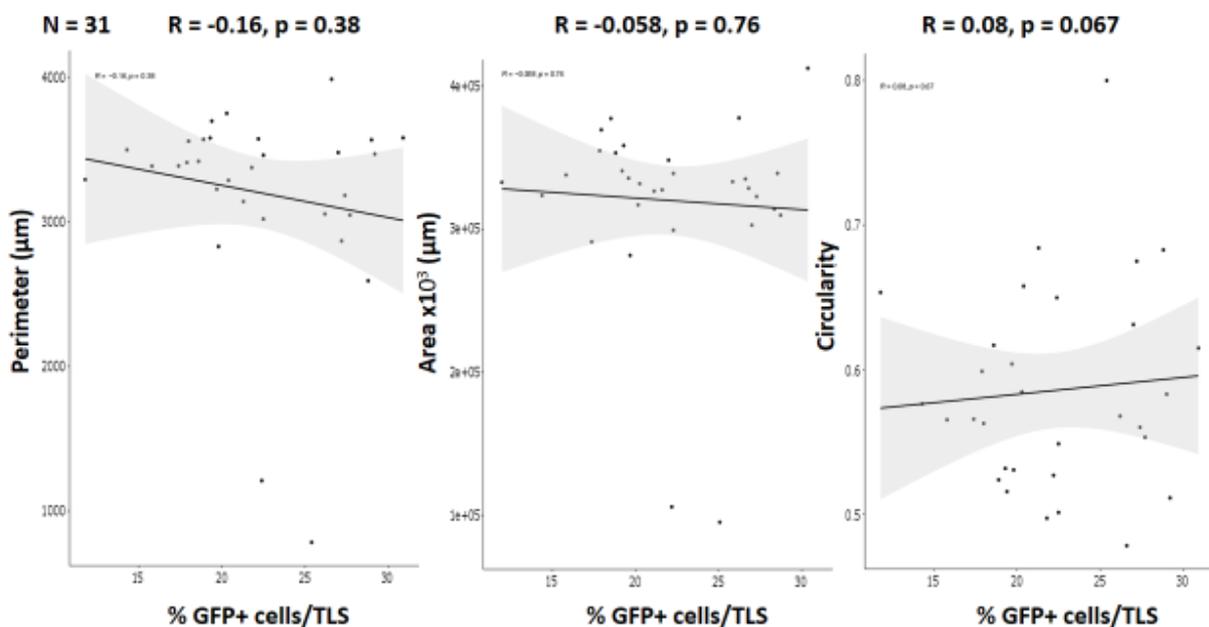


Figure 15 | No correlation between morphometric features of Cas9+ Tracr TLS and fraction of cells with scarring activity. Each dot represents an individual Tracr TLS. No correlation was

observed between the number of GFP+ cells and perimeter, area and circularity. Pearson's R = -0.16, -0.058, 0.08 P-value = 0.38, 0.76, 0.67 respectively. n = 31 The grey area represents the confidence interval.

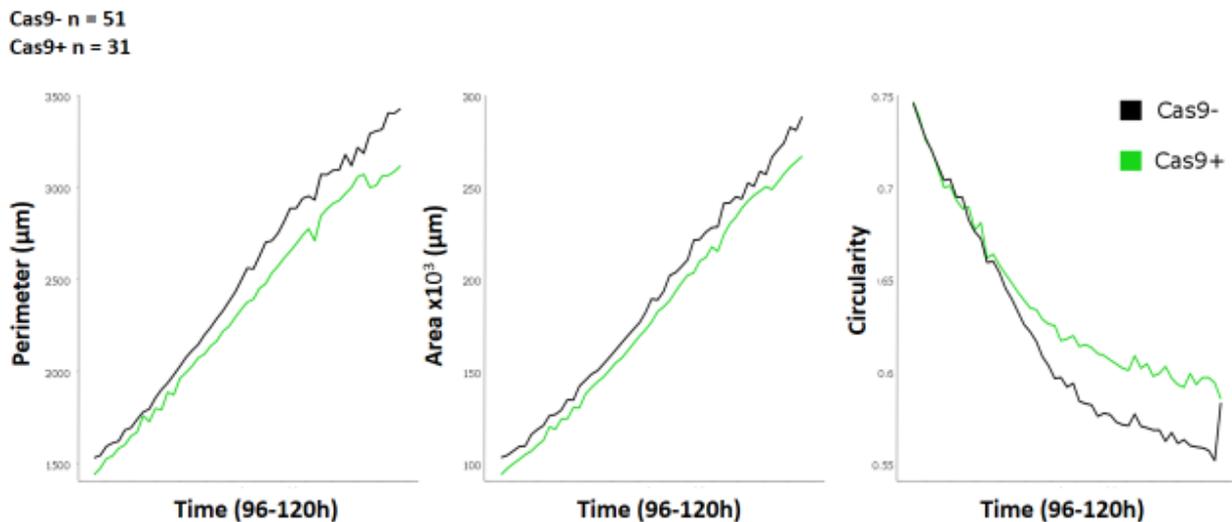


Figure 16 | Time-resolved morphometry of Cas9+ TLS versus Cas9- TLS. Tracr TLS were live imaged between 96 – 120h post aggregation. Cas9+ TLS display similar morphodynamics compared to Cas9- TLS. Cas9+ n = 31, Cas9- n = 51

5.8. Molecular characterization of Tracr TLS

To further test if the distribution of developmental marker genes resembled the organization in the embryo and in previously published TLS, we performed immunofluorescence for T, SOX2, CDX2 and SOX17 at 120h. Tracr TLS expressed T and CDX2 at the posterior pole mirroring the embryonic expression domains (**Figure 17**). We observed clear segmentation, indicative of the formation of somites. The prospective neural tube expressed SOX2, further confirming neural identity. Finally, immunofluorescent staining for SOX17, an endodermal marker, revealed gut formation (**Figure 17**).

These data show that Tracr TLS faithfully recapitulate the spatial organization of mesodermal and neural domains, and display embryo-like architecture with formation of somite-, gut- and neural tube-like structures, comparable to the TLS described by Veenvliet, Bolondi et al., 2020, Science, in press.

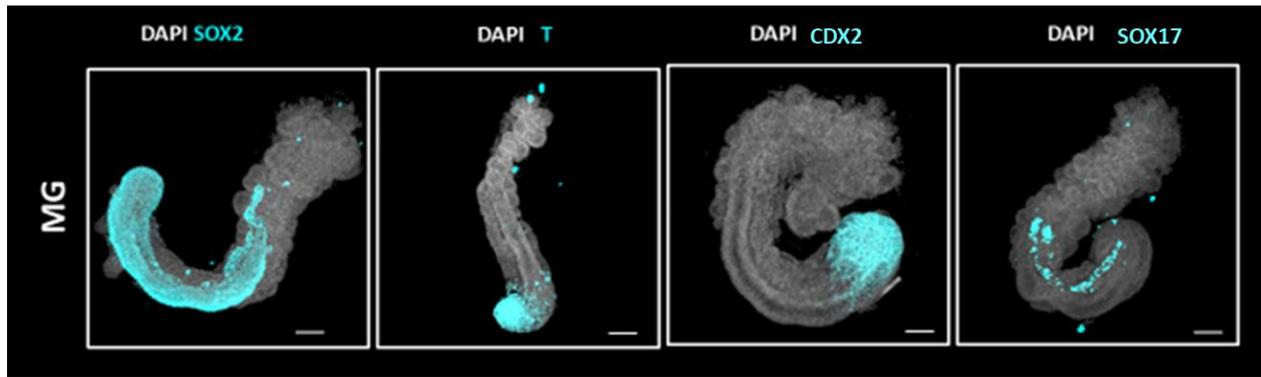


Figure 17 | Tracr TLS show mesodermal, neural and endodermal domains. Tracr TLS 120h post aggregation stained for SOX2, T, CDX2, SOX17 and counterstained with DAPI. SOX2+ cells form a neural tube like structure (first box); T-pole and Cdx2 mark the progenitor domain (second & third box); SOX17 indicating the presence of gut formation from endoderm lining (fourth box); DAPI nuclear counterstain shows tissue architecture. Scale bar indicates 100 μ m. MG - Matrigel

5.9. Combined WNT activation and BMP inhibition results in an excess of somites phenotype in Tracr TLS

CHIR99021, a constitutive activator of the WNT pathway through GSK-3 inhibition, activates T and together with WNT and FGF signaling maintain the NMPs via a positive feedback mechanism. These factors upregulate Tbx6 and together they control Msn1 expression, which is required to confer paraxial mesoderm identity (An WF et al., 2012; Koch et al., 2017). LDN-193189 blocks the activation of downstream BMP pathways such as Smad1/5/8 as well as the non-Smad pathways through MAPKs p38, ERK1/2 and the Akt pathway (Boergermann et al, 2010). *In vitro*, activation of WNT signalling and BMP inhibition in mESCs can give rise to paraxial mesoderm (Chal et al., 2015), whereas *in vivo*, WNT and BMP signalling pathways orchestrate paraxial mesoderm and lateral plate mesoderm fate (Chal et al., 2018).

To further improve the segmentation process and shift the trajectory in the paraxial mesoderm direction, we therefore treated Tracr aggregates with Chi and LDN (CL) in combination with Matrigel from 96 - 120h. We then assessed overall morphology and distribution of selected marker genes.

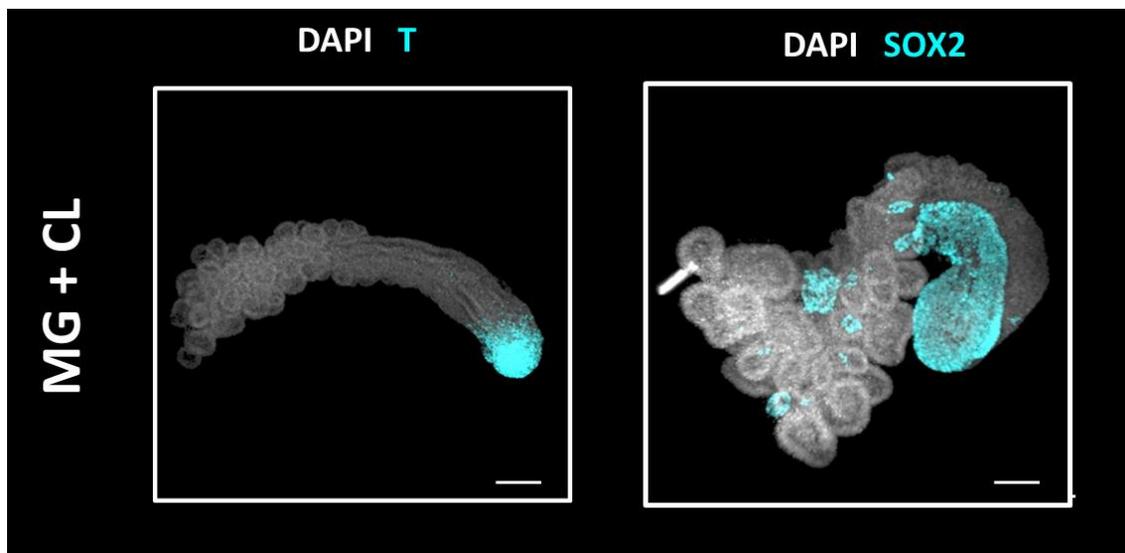


Figure 18 | Excess of somitic phenotype in Tracr TLS-CL. Tracr TLS 96h post aggregation were treated with MG + CL. At 120h, structures were stained with T, SOX2 and counterstained with DAPI. Structures display an excess of somites at the anterior pole that are organized like a “bunch of grapes”. MG + CL treated structures show a regular T pole formation (left); however, SOX2+ cells display a shortened neural domain (right). Scale bar indicates 100 μ m. MG – Matrigel; CL – Chiron + LDN

This dramatically increased the number of somites, all located at the anterior end of the TLS in a “bunch of grapes” arrangement (**Figure 18**) (Dias A.S et al., 2013; Veenvliet, Bolondi et al., 2020, Science, in press) and led to a disorganized and shortened neural domain which narrowed in length towards the anterior end. This can be clearly observed by SOX2 immunofluorescence in **Figure 18**, right. Furthermore, immunofluorescent staining revealed that T expression is unchanged compared to TLS that are not chemically modulated (**Figure 18**, left).

Thus, we showed that inhibition of BMP signaling along with the activation of WNT signaling shifts the differentiation trajectories of Tracr TLS towards the paraxial mesoderm which results in the formation of excessive somites as previously described for TLS. Thus, Tracr TLS respond to chemical modulation with CL in a similar manner as the previously published TLS (Veenvliet, Bolondi et al., 2020, Science, in press).

5.10. Sequential formation of somites in Tracr TLS

Somitogenesis in mouse starts at E8.0 and continues unto E13.0 during embryonic development (Saga, 2012; Cooke J. & Zeeman E.C., 1975). The segmentation clock controls periodicity of bilateral segment formation (Christ and Ordahl, 1995; Pourquié, 2011). The segmentation period is 2 hours in mice (Kageyama R. et al., 2012, Pourquié, 2011). To track the formation of segments as a proxy for periodicity, Tracr TLS were live imaged upon MG embedding from 96h-120h. Segments formed

consecutively on one or both sides of the neural tube-like structure in a rhythmic manner at an embryo-like pace (**Figure 19**). Around 96h, the structures begin to elongate further, giving rise to multiple segments along the anterior-posterior axis every one and half - two hours on average (**Figure 19**). Thus, Tracr TLS display sequential somite formation with embryo-like periodicity, matching the previous findings in non-Tracr TLS (Veenvliet, Bolondi et al., 2020, Science, in press).

5.11. GFP+ cells expressed in all lineages

We next asked whether GFP+ cells are expressed in specific cell types or in all the tissues of the Tracr TLS. To this end, we performed confocal imaging of the Tracr TLS and analyzed the localization of the GFP+ cells as a proxy for Cas9 activity.

Confocal imaging showed that GFP+ cells were distributed all over the structures in a salt-and-pepper manner (**Figure 20**).

Next, we performed scRNA-seq analysis and tested if Tracr TLS recapitulated all cell types detected in TLS. UMAP visualization revealed 14 distinct clusters with NMP-like cells flanked by a neural and somitic trajectory. Early & late neural-tube cells, as well as posterior and anterior PSM, somitic cells and somite derivatives like sclerotome and dermomyotome were detected. Moreover, we observed smaller clusters comprising endoderm, endothelial cells and Primordial Germ Cell Like Cells (PGCLCs) (**Figure 21**). These data show that Tracr TLS harbour all cell states previously described in canonical TLS (Veenvliet, Bolondi et al., 2020, Science, in press)

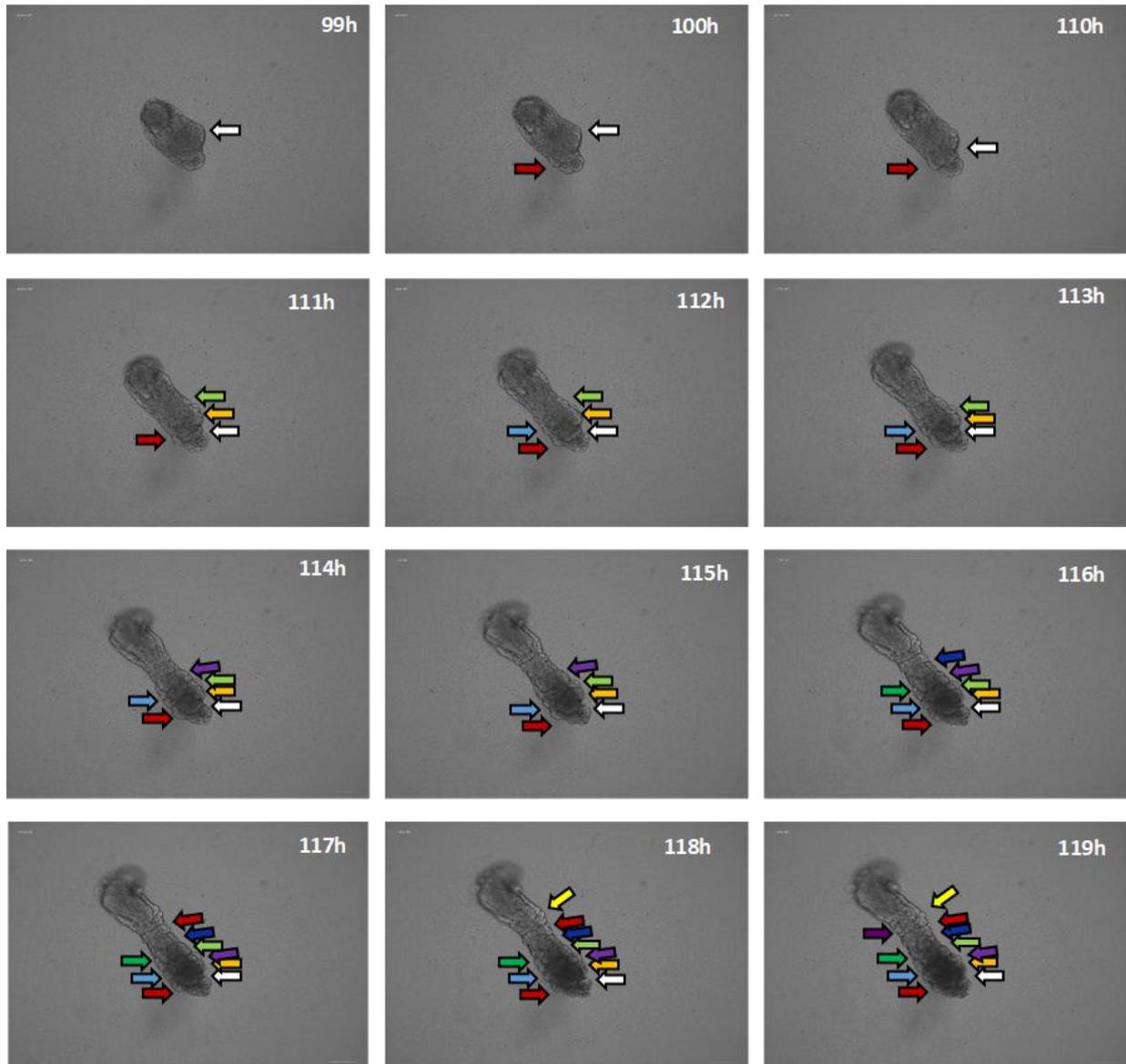


Figure 19 | Elongation of posterior T-pole and formation of neural tube and segments in Tracr TLS. After the tracr aggregates were embedded in matrigel, structures were live imaged from 99 – 119h. Arrows indicate the formation of somites sequentially from anterior to posterior axis

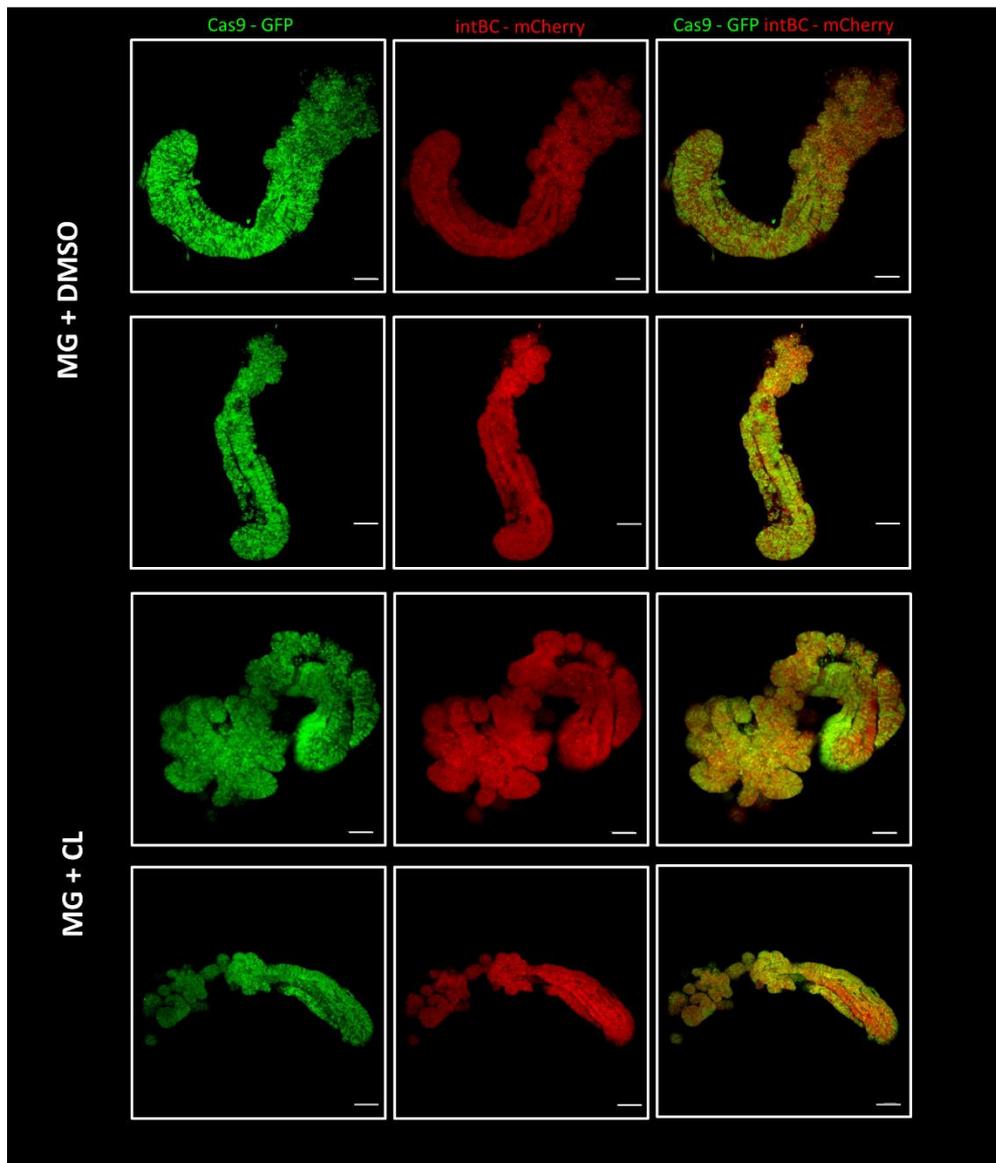


Figure 20 | CAS9-GFP+ cells are expressed in a salt-and-pepper manner in Tracr TLS. The Tracr cell line carries intBC – mCherry and Cas9 - GFP reporters. 96hGFP+ cells are detected throughout the structure in a salt-and-pepper distribution; intBCs - mCherry signal was observed in all cells; Scale bar indicates 100 μ m. MG – Matrigel; CL – Chiron + LDN

Between the two 120h Tracr TLS sequenced, inter-structure variance was detected (**Figure 21**). Lower presence of NMP-like cells, posterior pre-somitic mesoderm and higher somite sclerotome and late neural-tube cells were observed in TLS #2 (**Figure 21A**). The slight variation between the two TLS could be the result of different developmental progress of individual TLS despite sampling at the same absolute time. We also mapped Cas9-GFP and mCherry-intBCs as pseudogenes from the scRNAseq data and analyzed their expression across different cell types. UMAP visualization also showed that intBCs are expressed by every cell while Cas9-GFP showed similar salt-

and-pepper distribution as confirmed by the confocal images. Moreover, violin plots further revealed highly expressed intBCs across all cell types. Importantly, Cas9-GFP cells also express high level of the transgene. Furthermore barplots portrayed the percentage of Cas-GFP cells is expressed uniformly throughout all TLS clusters
In conclusion, the presence of Cas-GFP+ cells indicate that tracr is suitable for lineage tracing since all cell types have the potential to be traced.

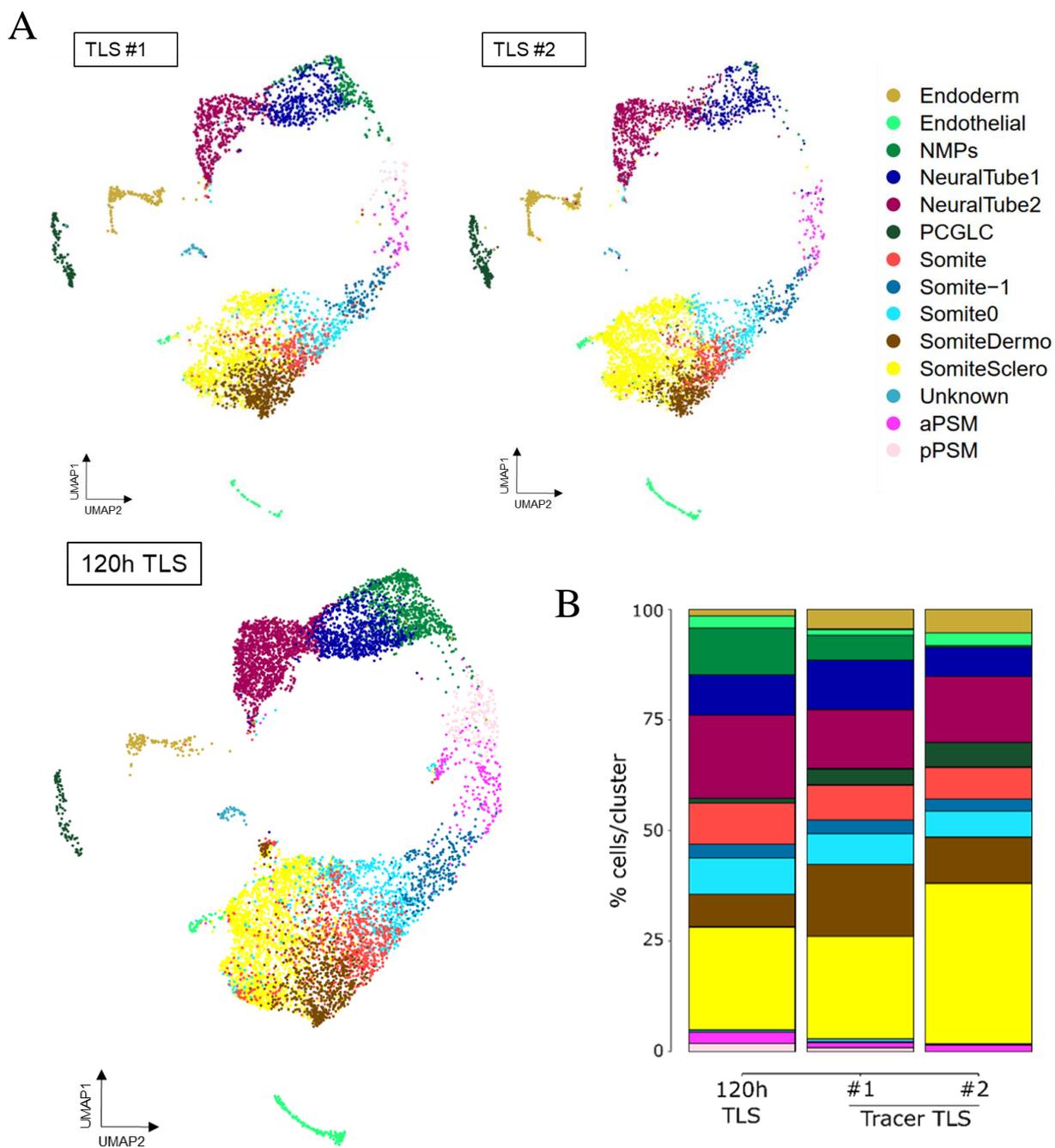


Figure 21 | Tracr TLS recapitulate cell diversity observed in TLS. 120h Tracr TLS were compared to the 120h single-cell TLS compendium previously published. A. UMAP (Uniform Manifold Approximation and Projection) coloured by the fourteen clusters identified. B. Stacked bar plot showing percentage the different cell states in 120h TLS and 120h Tracr TLS. (Veenvliet, Bolondi et al., 2020, Science, in press)

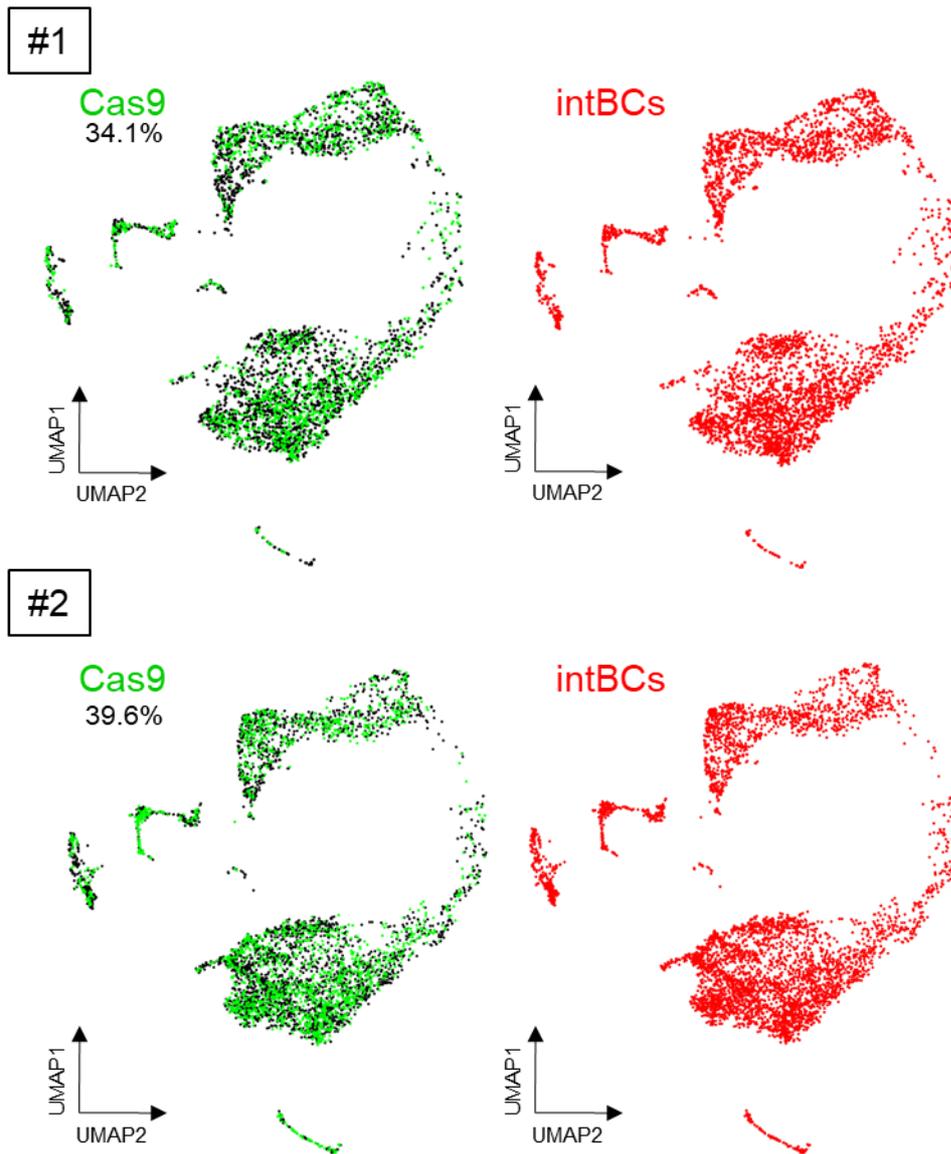
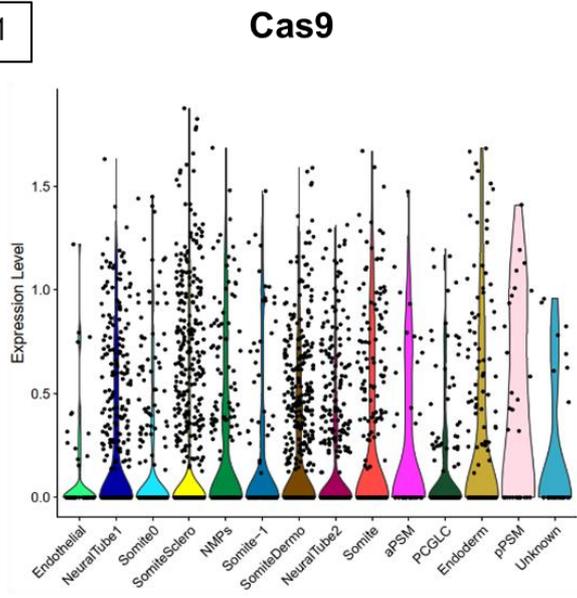
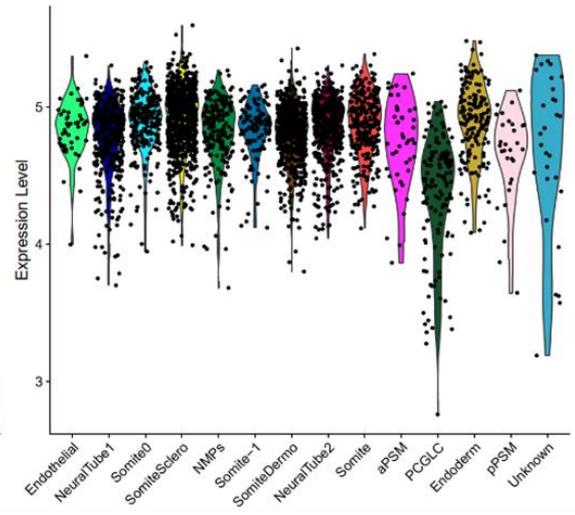


Figure 22 | Cas9-GFP and intBC expressed in all cell types UMAP (Uniform Manifold Approximation and Projection) coloured by green to identify Cas9-GFP and red to identify intBCs expressing cells. intBCs are expressed in every cell while Cas9-GFP displays a salt and pepper distribution across cells. Nevertheless, Cas9-GFP is evenly distributed across the whole UMAP.

#1



intBCs



#2

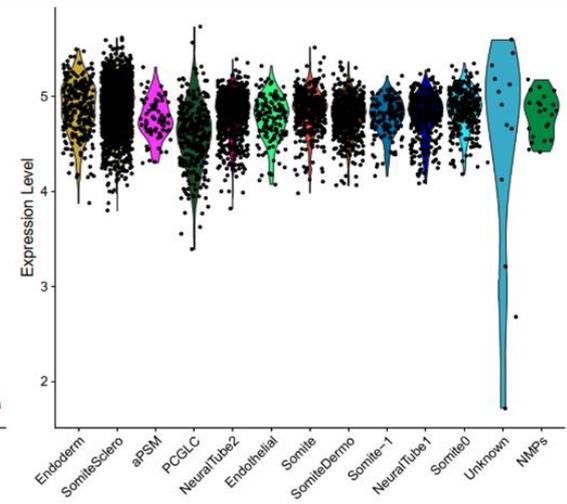
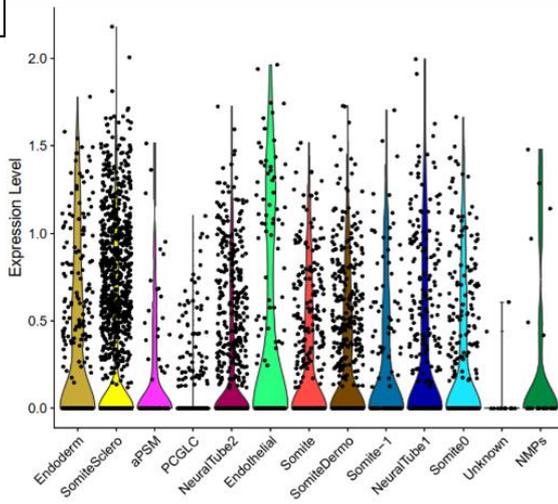


Figure 23 | Cas9-GFP and intBCs expression in single cells for the two sequenced Tracr TLS. Violin plots showing expression of Cas9-GFP or mCherry-intBCs for each cluster. End points of the violin plot represent minimum and maximum. Dots indicate individual cells. Central dotted-line represent the cluster mean.

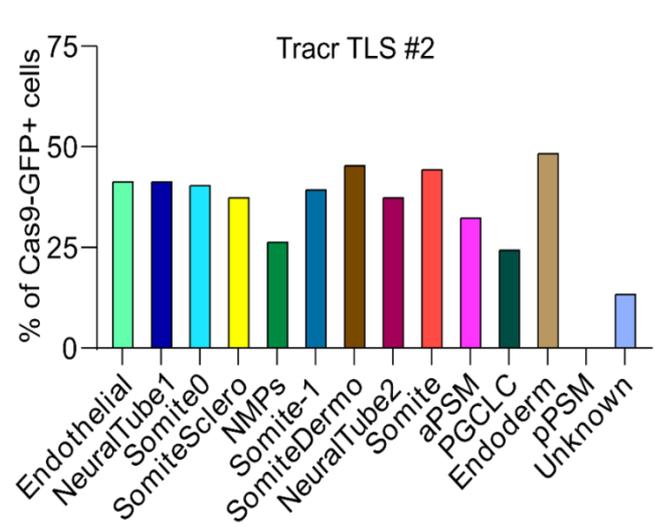
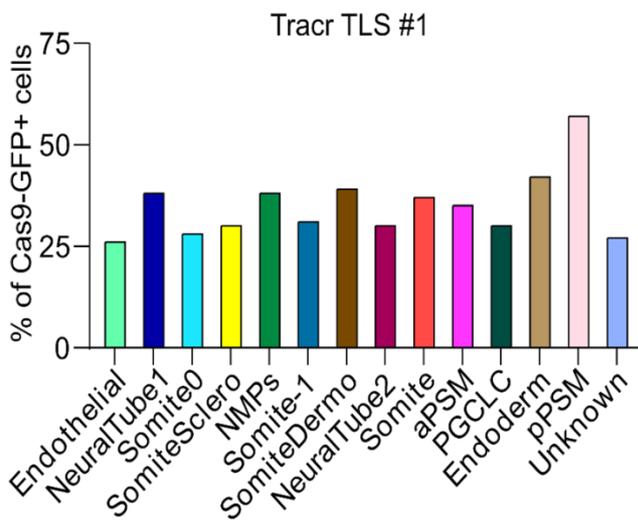


Figure 24 | Cas9-GFP is expressed uniformly throughout all TLS clusters. Bar plots showing the percentage of Cas9-GFP+ cells per cluster for Tracr-TLS #1 and #2. These data are complementary to the UMAPs and barplots from Fig. 21. Color coding represents annotated clusters from Fig.21

6. Discussion

In this project, to understand the interplay between tissue morphogenesis and lineage decisions we took advantage of two remarkable systems – a highly informative and continuous molecular recorder, Tracr cells, which was generated in-house by the Meissner Lab taking advantage of the molecular design from Chan, Smith et al. (2018), and a novel and exceptionally well-organized post-occipital embryonic model, the Trunk-like-structures (Veenvliet, Bolondi et al., 2020, Science, in press). Initially, we benchmarked the functionality of the Tracr cells by checking whether Cre induction will result in the molecular recorder activation. Importantly, intBCs, as shown by mCherry activity, were detected in almost all cells, indicating that our barcodes which may eventually give us the informative readout to trace cells were successfully recovered in all cells (**Figure 7B**). Apart from the expression of the intBCs, we need an active nuclease to create successfully generate scars. This was observed by Cas9-GFP activation in the large portion of the population at 24h (**Figure 7B**). GFP signal may be an indicator for the scarring activity on intBCs however it does not verify the generation of scars in our region of interest. Accumulation of scars, mainly deletions was confirmed by Screentape assay of amplified intBC from the gDNA and cDNA (**Figure 9B**).

Although our molecular recording approaches will ultimately rely on single cell RNA sequencing as a readout to construct lineage relationships between cells, these results led to the confirmation that Tracr cells accumulate scars on the intBCs over time and this obtained complexity could potentially be used to trace cell lineages. Nevertheless, the reason why we pooled GFP+ gDNA and RNA cells during *Luria-Delbrück-like* experiment is due to the fact that randomly integrated transgene cassettes in the genome might undergo epigenetic silencing permanently or silenced after being expressed for some time depending on their landing locations (Eissenberg J.C. 1989). This might result in the absence of specific barcodes cassette in our datasets. For this reason we will verify if the barcode recovery at the RNA level might be different than

the one observed at DNA level. Our BC amplicon seq using both gDNA and cDNA in activated Tracr cells will elucidate the extent of barcode silencing and will provide fundamental information regarding the future resolution of the system.

However, we were surprised to see the quantity of GFP+ cells gradually decrease after 24h until 120h (**Figure 7B**). Two possible scenarios may explain the observed decrease in the amount GFP+ cells over time: silencing of the CAS9-GFP transgene in the Rosa26 locus, or outgrowth competition by non-successfully transfected GFP- cells in the population. With our assay, we indirectly showed that GFP+ cells outgrow GFP- cells by isolating GFP+ cells at 24h and culturing them until they were subjected to flow cytometry analysis at 120h (**Figure 8A**). The reason for the outgrowth may be due to double strand breaks (DSBs) created by Cas9 nuclease in Cas9-GFP+ cells that will eventually enable cell tracing. Several cellular mechanisms recognize these breaks as a defect and activate several repair molecular mechanisms, including p53-mediated DNA damage response that may induce cell cycle arrest or apoptosis (Haapaniemi E., et al 2018). The type of mechanism relatively depends on the cell cycle stage; during G1 phase the error-prone NHEJ gets activated, whereas during S phase more precise homology-directed repair is active (Haapaniemi E., et al 2018; Hustedt & Durocher, 2016). Cells that undergo stress programs overcome this problem by slowing down the cell cycle which leads to slower proliferation of cells (Fragkos & Naim, 2017). One way to address the decrease in GFP+ cells is to test other transgene delivery systems. It will be interesting to check if these systems also lead to competitive outgrowth (see section: Shifting to methods that will decrease the initial stress on the cells).

Since differentiation is known to be a stressful process, we analyzed whether, in addition to DSB, differentiation is an additional cause of stress to Cas9-GFP cells. Indeed, this was the case and cells treated with differentiation media showed a decrease in GFP+ cells compared to ES + LIF media (**Figure 8D**).

This may arise as most of the differentiation pathways that regulate stem cells are also involved in stress-response mechanisms (Tower, 2012). ESC markers such as OCT4, Nanog Sox2, related to pluripotency, are downregulated in response to increasing levels of ROS, which enhances the ESC differentiation along the mesodermal and endodermal trajectory (Ae-Ri Ji, et al., 2010). This may explain the further decrease in

GFP+ cells during differentiation as cell cycle further slows or induces apoptosis, and could thus lead to GFP- fit cells outgrowing the GFP+ cells (Bowling S. et al., 2019).

6.1. Tracr TLS: molecular recording in 3D structures with an embryo-like architecture

Two-dimensional ES cell models capture meaningful biological trends however, the cell-to-cell communication present in three-dimensional environment is missing. Due to missing communication, cells could hijack their way into differentiation paths due to their unnatural microenvironment which limits their usage in modelling development. For this reason, we generated Trunk-like-structures based on the protocol used in Veenvliet, Bolondi et al. (2020) using activated Tracr cells to better understand the decisions taken by cells to exit pluripotency and to measure trajectories that will give rise to an embryo-like-structure.

To generate Tracr TLS, we found out almost 3 times as many cells are required for Tracr aggregates to mimic the architecture previously described for normal TLS (**Figure 17**). As mentioned by Van den Brink, larger aggregates tend (>600 cells) to grow in the disorganised manner. Since our protocol worked well with 550 cells, we propose that GFP- cells among the 550 cells that have not undergone DSB may be further supporting the Cas9-GFP+ cells to allow normal TLS development. Aggregate size is a major factor in the generation of stem cell models as the EB size may have an impact on the lineage bias during spontaneous differentiation (Xie A. W. et al., 2017; Ng E. S. et al., 2005). On the other hand, Cas9- cells displayed abnormal elongation (**Appendix Figure 1.**). This may also be due different aggregate size at the time of Chiron pulsing or Matrigel embedding rather than the initial aggregate size. Same sized aggregates can be imaged at 24h, 72h and 96h to check the diameter to assess whether individual aggregates undergo faster proliferation.

We next confirmed the scarring dynamics in 3D Tracr TLS by the accumulation of deletions that will enable cell lineage tracing in these models (**Figure 10B**) and were surprised to observe that Tracr TLS had a 50% reduction in GFP+ cells as compared to 2D differentiation (**Figure 10C**). Multiple reasons may lead to this difference. One factor can be the difference in the initial aggregation size per well. As shown before,

we initially observed reduction in the GFP+ cells due to DSB induced toxicity. Next, treatment with differentiation media on 2D culture further confirmed that developmental processes can be additionally stressful for the cells. In this case, since transformation from a monolayer to a multilayer environment can initiate developmental processes in Tracr TLS similar to embryos, cells are engaged in a 'survival of the fittest' battle to give rise to the best organism (Bowling S. et al., 2019) which may further explain the decrease in the GFP+ cells. Importantly, since the protocol used for these cells include N2B27 and Chi, it can be tested if these treatment also have effect on the decrease of the Cas9-GFP+ fraction.

Constructing lineage trees with 20-40% informative cells may cause some limitations as intermediate branch-point information will be lost that will lead to gaps in state trajectory. Our system may face other limitations of cell state manifolds as well, which may be hidden dynamics or delayed state divergence, by which cells become committed to distinct trajectories but continue to follow the same cell state (Wagner & Klein, 2020).

A method to obtain more informative readouts may be further addressed by multiplexing individual 120h Tracr TLS with lipid-tagged indices. Lipid- and cholesterol-modified oligonucleotide anchors are hybridized with the barcode oligonucleotide which can stably incorporate into the plasma membrane of live cells or nucleus (McGinnis C. S. et al., 2019). Once tagged, TLS can be dissociated and subjected to FACs to obtain single GFP+ cells. This way the GFP+ cells can be traced back to their TLS origin, informative Cas9-GFP+ cells can be pooled together and subjected to 10x genomic sequencing. This will lead to multiplexing of more cells in the same reaction as individual TLS contain less than 10,000 GFP+ cells. Cell lineages can be constructed by combining the information obtained from several uniquely tagged TLS.

Our Tracr model also demonstrated that the quantity of GFP+ cells, which may be indicative of the number of scarred cells, is not correlated with morphometric features (**Figure 13**), and time-resolved analysis of the geometric features displayed by Cas9+ and Cas9- cells display similar morphology (**Figure 16**). This suggests that scarring does not affect overall TLS morphology or growth.

These data may further indicate that our molecular recorder targets only the target-

site-cassette and does not interfere in other gene regulations that may affect morphology, which makes our system reliable for use.

Regardless of the scarring activity and stress-induced processes, Tracr TLS successfully recapitulated the spatial organization of mesodermal and neural domains, and displayed embryo-like architecture with formation of somite-, gut- and neural tube-like structures, comparable to the TLS described by Veenvliet, Bolondi et al., 2020 (**Figure 17**). Similar to these TLS and the post-implantation embryo, our Tracr TLS expressed T and CDX2 at the posterior end of the structure and a neural tube was observed flanked by somites that formed sequentially over time (**Figure 17, 19**). TLS displayed presence of more differentiated somite markers such as *Uncx4.1*, *Tbx18*, *Meox2* and *Pax3* (Van den Brink et al., 2020). Paraxial mesoderm leads to the specification of somites and amongst other lineages (Pei et al., 2019; Nakaya et al., 2004). Our Tracr system may further give an insight on how cell fate decisions are taken at each branching point to differentiate into somite cellular subtypes which will form skeletal muscle, bones and cartilage during normal development.

Studies have found that mutation in the Notch pathway genes such as *DLL3*, *Mesp2*, Lunatic fringe (*Lfng*) and *Hes7* lead to human pathologies that affect segmentation in humans (Turnpenny et al., 2007; Sparrow et al., 2006). Being able to record cell fate decisions during somitic trajectories in normal development may facilitate the understanding of how these pathways give rise to pathologies.

Gastruloids express heart markers such as *Gata6* and *Hand2* (Van den Brink et al., 2020). Using our tracr system in TLS, we can map the cell fates that lead to the expression of these markers and observe how intermediate mesodermal fate gives rise to the heart cells (Kumar et al., 2015). This might give an insight as to why TLS models lack the modelling of the anterior part.

Moreover, the tube structure of the gut can be clearly observed in the Tracr TLS by *Sox17* staining (**Figure 17**). *In vivo*, at a certain time point before E9.5, the lateral endoderm gives rise to the gut tube by folding ventrally like a sheet of paper (Spence, J. R. et al., 2011). Tracking gut generation is hard to achieve *in vivo*. Moreover it is still debated if the endoderm layer generates through an intermediate mesendodermal state that then resolves into mesoderm and endoderm. Our approach will allow to measure the extent of mesendoderm generation and in general will help the understanding of how the endoderm lineage specify after gastrulation.

Paraxial mesoderm and lateral plate mesoderm fate is determined by a shared common precursors in the primitive streak and absence of BMP inhibition leads to the upregulation of the lateral plate markers *in vitro* (Chal et al., 2018). In contrast, activation of WNT signalling and BMP inhibition in mESCs can give rise to paraxial mesoderm *in vitro* (Chal et al., 2015). Chemical modulation of TLS has shown that inhibition of BMP signaling along with the activation of WNT signalling results in an excess of somite formation (Veenvliet, Bolondi et al., 2020, Science, in press). This was recapitulated in Tracr TLS that displayed a dramatic increase in the number of segments at the anterior end of the structure and a disorganized and shortened neural tube (**Figure 18**). This is likely caused by dominant WNT signalling and BMP inhibition which shift differentiation trajectories in the paraxial mesoderm direction. In conclusion we observed that cell fate can actually be biased via chemical stimulation (**Figure 18**) and owing to our Tracr system, signaling molecules that orchestrate embryogenesis and affect lineage trajectories can be identified. After construction of lineage trees, Tracr cells may even portray how cells biased towards specific fates can still retain the ability to generate other cell types as indicated by Chan, Smith et al., 2019.

Although our system may have a bottle-neck regarding the fraction of GFP+ cells in each Tracr TLS, an important finding of our detailed characterization of Tracr TLS is that the Cas9-GFP+ cells are not biased towards a specific cell domain (**Figure 20, 21, 22, 23, 24**). The uniform expression of Cas-GFP cells in a salt-and-pepper manner throughout all TLS clusters show that tracr system has the potential to trace individual cell states that give rise to the embryo-like architecture of TLS.

Although Tracr TLS used for scRNAseq were sampled at 120h, we did observe inter-structure variance in terms of cell types distribution (Figure 23). This may happen due to the period at which developmental processes initiated although both the Tracr TLS were sampled at 120h. "This is relatively normal as even identical cell types within the same organism display a range of variation in terms of proliferation, genotype, function and signaling which can be easily identified by our Tracr system (Bowling S. et al., 2019). This observation highlights furthermore the power of TLS as an *in vitro* model system to study variation during exit from pluripotency.

6.2. Limitations of CRISPR-based barcode editing technology

Although CRISPR-based methods generate highly diverse evolving barcodes, it is important to consider the unfavourable effects they can have on the biological function of the cells. As mentioned earlier, several studies have revealed that the double-strand breaks created by Cas9 nuclease are toxic to the cells and induce cell death (Haapaniemi E., et al 2018; Hustedt & Durocher, 2016). To avoid the information loss by double-stranded breaks, our Cas9 nuclease could be adapted to a method designed by Hwang B. et al (2019) where CRISPR-editing uses the enzyme nickase Cas9 (nCas9) that when fused with cytidine deaminase enzyme, substitutes C:G pairs to T:A in the 4-8 nucleotide region nearby the protospacer adjacent motif (PAM), thus creating scars without inducing double-stranded breaks (Hwang B. et al, 2019; Komor A.C et al., 2016). Since, the substituted nucleotides are very limited (C:G, T:A) the question is whether this system creates enough diversity to trace lineages comparable to our system which can introduce huge amount of scars in 10 different locations in the genome.

Shortening the Cas9 half-life or mutation rates of guides may be another option to maintain gene editing and reduce toxicity (Yang S. et al, 2019). The mutation rate can be controlled by creating single or dual mismatches in sgRNAs. This is important as fast, frequent cuts can cause cellular toxicity while slower cuts may improve viability (Ihry R. J. et al., 2018; Haapaniemi E. et al., 2018). Using fast cuts enables tracing at an earlier time point in the development however; this method creates higher proportions of cells with identical scars (Chan, Smith et al., 2019). An approach used by Chan, Smith and colleagues involves the use of slow rate mismatched gRNAs to better reflect the cell cycles occurring in early mouse embryo. In our case, we used a construct with a guide combination (2,3,3) where mismatches were made in guides according to proximity to the protospacer adjacent motif (PAM) sequence..

6.3. Shifting to methods that will decrease the initial stress on the cells

Nucleofection is an electroporation technique that allows the efficient gene transfer directly into the nucleus of the cell. Various studies tested the effect of nucleofection on cells and revealed that cells undergo substantial changes until they recover an hour later. These changes included plasma membrane dilation and crease-like morphology

(Zhang M. et al 2014). Although Tracr cells give rise to correctly shaped TLS with proper morphology without causing any changes molecularly, during TLS formation, the proportion of GFP+ cells is relatively low (**Figure 10C**), which is a somewhat limiting factor during construction of TLS lineages. For this reason, our work to shift to a less stressful tracing methodology is an ongoing process. Gene expression systems that regulate the gene activity by Tet-On and Tet-Off switches are called tetracycline-controlled gene expression systems (Das A. T. et al., 2016). Plasmids designed with Cre fused to tetracycline responsive elements get activated in response to tetracycline-derivative Doxycycline (Dox) treatment. Another method is the Tamoxifen-inducible Cre-ERT2 system. Here, Cre is ligated to a mutated ERT2 - an estrogen binding domain, which gets translocated to the nucleus when Tamoxifen is administered (an estrogen antagonist) (Zhong, Z. A., et al., 2015). Both these methods were tested and the second method displayed presence of GFP+ cell which may indicate occurrence of scarring activity (data not shown). However, it needs to be further evaluated and confirmed. That said, it is important that most of the toxicity observed in Tracr TLS remains associated to the generation of DSBs more than to the strategy with which Cas9 is activated. As we previously saw our system shows 90% efficiency by the expressed GFP+ cells at 24h (**Figure 7B**), which shows nucleofection method is very efficient. Although one could test the editing efficiency by using less stressful transgene transfer systems to check the outgrowth competition. If outgrowth is also observed in these systems, this may further indicate that DSB is the causative stressor, however, if not this would imply that nucleofection is the causative stressor.

6.4. Role of ECM during development

Very little is known about the impact of Matrigel in differentiation, testing different extracellular matrices with different growth factors in Tracr TLS may reveal the different trajectories cells choose based on the different matrices. This may lead to a better understanding on how ECM components induce morphogenesis and lineage specification. The gastruloid protocol in the absence of Matrigel recapitulates symmetry breaking, germ layer formation, axial organisation and elongation (van den Brink et al., 2014). Although regulators of segmentation clock is present, gastruloids lack proper somite formation (Van den Brink et al., 2020). ECM proteins and matrix can determine the cell behaviour, polarity, migration, differentiation, proliferation and

survival by signaling and communication with the intracellular cytoskeleton (Kim S. et al., 2011). Embedding gastruloids in Matrigel induced segmentation and neural-tube formation (Veenvliet, Bolondi et al., 2020, Science, in press; Van den Brink et al., 2020). Since cells respond to the mechanical and biochemical changes in ECM through the crosstalk between integrins and the actin cytoskeleton embedding TLS in Matrigel leads to the upregulation of integrins and transmembrane receptors in TLS (Alam et al., 2007; Veenvliet, Bolondi et al., 2020, Science, in press) One could also investigate the trajectories that lead to the upregulation of these receptors mediating cell adhesion to the ECM by comparing the cell fates in gastruloids and Tracr TLS.

In addition, construction of a more endometrium-typical extracellular matrix containing, among other substances, collagens I, III, IV and VI, laminin, desmin, nidogen 1 and biglycan, can replace the use of matrigel (Boretto, M et al., 2017). Embedding Tracr TLS to an endometrium-like ECM may further unveil the relationship between endometrium ECM, cell lineage, and tissue morphogenesis.

6.5. Concluding remarks and future perspective

Trunk-like-structures hold a great promise in understanding developmental biology in depth. Combined with Tracr technology, comparing gastruloids with TLS can be used to unveil the molecular mechanisms that control embryonic architectures; since

TLS reflect the embryo at the morphological level whereas gastruloids don't, although both the models are highly similar at the molecular level (Veenvliet, Bolondi et al., 2020, Science, in press). With the available molecular recording Tracr technology, tracking how cell fate decisions and tissue morphogenesis are coordinated becomes possible.

7. Literature

- Alam, N., Goel, H. L., Zarif, M. J., Butterfield, J. E., Perkins, H. M., Sansoucy, B. G., ... & Languino, L. R. (2007). The integrin—growth factor receptor duet. *Journal of cellular physiology*, 213(3), 649-653.
- Alemaný, A., Florescu, M., Baron, C. S., Peterson-Maduro, J., & Van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. *Nature*, 556(7699), 108-112.
- Amin, S., Neijts, R., Simmini, S., van Rooijen, C., Tan, S. C., Kester, L. & Deschamps, J. (2016). Cdx and T brachyury co-activate growth signaling in the embryonic axial progenitor niche. *Cell reports*, 17(12), 3165-3177.
- Arnold, S. J., & Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nature reviews Molecular cell biology*, 10(2), 91-103.
- Aulehla, A., & Herrmann, B. G. (2004). Segmentation in vertebrates: clock and gradient finally joined. *Genes & development*, 18(17), 2060-2067.
- Beccari, L., Moris, N., Girgin, M., Turner, D., Baillie-Johnson, P., & Cossy, A. et al. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature*, 562(7726), 272-276. doi: 10.1038/s41586-018-0578-0
- Baillie-Johnson, P., Van den Brink, S. C., Balayo, T., Turner, D. A., & Arias, A. M. (2015). Generation of aggregates of mouse embryonic stem cells that show symmetry breaking, polarization and emergent collective behaviour in vitro. *JoVE (Journal of Visualized Experiments)*, (105), e53252.
- Barends, P. M. G., Stroband, H. W. J., Taverne, N., Te Kronnie, G., Leën, M. P. J. M., & Blommers, P. C. J. (1989). Integrity of the preimplantation pig blastocyst during expansion and loss of polar trophoctoderm (Raubert cells) and the morphology of the embryoblast as an indicator for developmental stage. *Reproduction*, 87(2), 715-726.
- Beccari, L., Moris, N., Girgin, M., Turner, D. A., Baillie-Johnson, P., Cossy, A. C. & Arias, A. M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature*, 562(7726), 272-276.
- Beddington, R. S. (1994). Induction of a second neural axis by the mouse node. *Development*, 120(3), 613-620.
- Beddington, R. S., & Robertson, E. J. (1999). Axis development and early asymmetry in mammals. *Cell*, 96(2), 195-209.
- Bedzhov, I., & Zernicka-Goetz, M. (2014). Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell*, 156(5), 1032-1044.
- Bergsland, M., Ramsköld, D., Zaouter, C., Klum, S., Sandberg, R., & Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes & development*, 25(23), 2453-2464.
- Boergermann, J. H., Kopf, J., Yu, P. B., & Knaus, P. (2010). Dorsomorphin and LDN-193189 inhibit

- BMP-mediated Smad, p38 and Akt signalling in C2C12 cells. *The international journal of biochemistry & cell biology*, 42(11), 1802-1807.
- Boretto, M., Cox, B., Noben, M., Hendriks, N., Fassbender, A., Roose, H., ... & Meuleman, C. (2017). Development of organoids from mouse and human endometrium showing endometrial epithelium physiology and long-term expandability. *Development*, 144(10), 1775-1786.
- Borland, R. M., Biggers, J. D., & Lechene, C. P. (1977). Fluid transport by rabbit preimplantation blastocysts in vitro. *Reproduction*, 51(1), 131-135.
- Bowling, S., Lawlor, K., & Rodríguez, T. A. (2019). Cell competition: the winners and losers of fitness selection. *Development*, 146(13), dev167486.
- Burdsal, C. A., Damsky, C. H., & Pedersen, R. A. (1993). The role of E-cadherin and integrins in mesoderm differentiation and migration at the mammalian primitive streak. *Development*, 118(3), 829-844.
- Burdsal, C. A., Damsky, C. H., & Pedersen, R. A. (1993). The role of E-cadherin and integrins in mesoderm differentiation and migration at the mammalian primitive streak. *Development*, 118(3), 829-844.
- Bush, K. T., Lynch, F. J., DeNittis, A. S., Steinberg, A. B., Lee, H. Y., & Nagele, R. G. (1990). Neural tube formation in the mouse: a morphometric and computerized three-dimensional reconstruction study of the relationship between apical constriction of neuroepithelial cells and the shape of the neuroepithelium. *Anatomy and embryology*, 181(1), 49-58.
- Bush, K. T., Lynch, F. J., DeNittis, A. S., Steinberg, A. B., Lee, H. Y., & Nagele, R. G. (1990). Neural tube formation in the mouse: a morphometric and computerized three-dimensional reconstruction study of the relationship between apical constriction of neuroepithelial cells and the shape of the neuroepithelium. *Anatomy and embryology*, 181(1), 49-58.
- Bush, K. T., Lynch, F. J., DeNittis, A. S., Steinberg, A. B., Lee, H. Y., & Nagele, R. G. (1990). Neural tube formation in the mouse: a morphometric and computerized three-dimensional reconstruction study of the relationship between apical constriction of neuroepithelial cells and the shape of the neuroepithelium. *Anatomy and embryology*, 181(1), 49-58.
- Chal, J., & Pourquié, O. (2017). Making muscle: skeletal myogenesis in vivo and in vitro. *Development*, 144(12), 2104-2122.
- Chal, J., Al Tanoury, Z., Oginuma, M., Moncuquet, P., Gobert, B., Miyanari, A., & Sumara, O. (2018). Recapitulating early development of mouse musculoskeletal precursors of the paraxial mesoderm in vitro. *Development*, 145(6).
- Chal, J., Oginuma, M., Al Tanoury, Z., Gobert, B., Sumara, O., Hick, A. & Tassy, O. (2015). Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nature biotechnology*, 33(9), 962-969.
- Chan, C., Heisenberg, C., & Hiiragi, T. (2017). Coordination of Morphogenesis and Cell-Fate Specification in Development. *Current Biology*, 27(18), R1024-R1035. doi: 10.1016/j.cub.2017.07.010
- Chan, M., Smith, Z., Grosswendt, S., Kretzmer, H., Norman, T., & Adamson, B. et al. (2019). Molecular recording of mammalian embryogenesis. *Nature*, 570(7759), 77-82. doi: 10.1038/s41586-019-

- Christ, B., & Ordahl, C. P. (1995). Early stages of chick somite development. *Anatomy and embryology*, 191(5), 381-396.
- Christ, B., Huang, R., & Scaal, M. (2007). Amniote somite derivatives. *Developmental dynamics: an official publication of the American Association of Anatomists*, 236(9), 2382-2396.
- Ciruna, B., & Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Developmental cell*, 1(1), 37-49.
- Conklin, E. G. (1905). *The organization and cell-lineage of the ascidian egg* (Vol. 13).
- Cooke, J., & Zeeman, E. C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *Journal of theoretical biology*, 58(2), 455-476.
- Davidson, K. C., Mason, E. A., & Pera, M.
- Eckert, J. J., McCallum, A., Mears, A., Rumsby, M. G., Cameron, I. T., & Fleming, T. P. (2004). Specific PKC isoforms regulate blastocoel formation during mouse preimplantation development. *Developmental biology*, 274(2), 384-401.
- Eissenberg, J. C. (1989). Position effect variegation in *Drosophila*: towards a genetics of chromatin assembly. *Bioessays*, 11(1), 14-17.
- Enders, A. C., Schlafke, S., & Hendrickx, A. G. (1986). Differentiation of the embryonic disc, amnion, and yolk sac in the rhesus monkey. *American journal of anatomy*, 177(2), 161-185.
- Fleming, T. P., Sheth, B., & Fesenko, I. (2001). Cell adhesion in the preimplantation mammalian embryo and its role in trophectoderm differentiation and blastocyst morphogenesis. *Front Biosci*, 6(1), D1000-D1007.
- Fragkos, M., & Naim, V. (2017). Rescue from replication stress during mitosis. *Cell Cycle*, 16(7), 613-633.
- Gardner, R. L. (1968). Mouse chimaeras obtained by the injection of cells into the blastocyst. *Nature*, 220(5167), 596-597.
- Gilbert, S., 2006. *Developmental Biology*. Sunderland, Mass.: Sinauer. P 382
- Gouti, M., Delile, J., Stamataki, D., Wymeersch, F. J., Huang, Y., Kleinjung, J. & Briscoe, J. (2017). A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Developmental cell*, 41(3), 243-261.
- Guo, C., Kong, W., Kamimoto, K., Rivera-Gonzalez, G. C., Yang, X., Kiritani, Y., & Morris, S. A. (2019). CellTag Indexing: genetic barcode-based sample multiplexing for single-cell genomics. *Genome biology*, 20(1), 90.
- Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature medicine*, 24(7), 927-930.
- Harrison, S., Sozen, B., Christodoulou, N., Kyprianou, C., & Zernicka-Goetz, M. (2017). Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science*, 356(6334), eaal1810. doi: 10.1126/science.aal1810
- Henrique, D., Abranches, E., Verrier, L., & Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development*, 142(17), 2864-2875.
- Hubaud, A., & Pourquié, O. (2014). Signalling dynamics in vertebrate segmentation. *Nature Reviews*

Molecular Cell Biology, 15(11), 709-721.

- Hughes, C. S., Postovit, L. M., & Lajoie, G. A. (2010). Matrigel: a complex protein mixture required for optimal growth of cell culture. *Proteomics*, 10(9), 1886-1890.
- Hustedt, N., & Durocher, D. (2017). The control of DNA repair by the cell cycle. *Nature cell biology*, 19(1), 1-9.
- Hustedt, N., & Durocher, D. (2017). The control of DNA repair by the cell cycle. *Nature cell biology*, 19(1), 1-9.
- Hwang, B., Lee, W., Yum, S. Y., Jeon, Y., Cho, N., Jang, G., & Bang, D. (2019). Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements. *Nature communications*, 10(1), 1-9.
- Hwang, B., Lee, W., Yum, S. Y., Jeon, Y., Cho, N., Jang, G., & Bang, D. (2019). Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements. *Nature communications*, 10(1), 1-9.
- Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., ... & Randhawa, R. (2018). p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature medicine*, 24(7), 939-946.
- F. (2015). The pluripotent state in mouse and human. *Development*, 142(18), 3090-3099.
- Jao, L. E., Wente, S. R., & Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences*, 110(34), 13904-13909.
- Ji, A. R., Ku, S. Y., Cho, M. S., Kim, Y. Y., Kim, Y. J., Oh, S. K., ... & Choi, Y. M. (2010). Reactive oxygen species enhance differentiation of human embryonic stem cells into mesendodermal lineage. *Experimental & molecular medicine*, 42(3), 175-186.
- Ji, A. R., Ku, S. Y., Cho, M. S., Kim, Y. Y., Kim, Y. J., Oh, S. K., ... & Choi, Y. M. (2010). Reactive oxygen species enhance differentiation of human embryonic stem cells into mesendodermal lineage. *Experimental & molecular medicine*, 42(3), 175-186.
- Kageyama, R., Niwa, Y., Isomura, A., González, A., & Harima, Y. (2012). Oscillatory gene expression and somitogenesis. *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(5), 629–641. <https://doi.org/10.1002/wdev.46>
- Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., & Church, G. M. (2018). Developmental barcoding of whole mouse via homing CRISPR. *Science*, 361(6405).
- Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., & Church, G. M. (2018). Developmental barcoding of whole mouse via homing CRISPR. *Science*, 361(6405).
- Kawagishi, R., Tahara, M., Sawada, K., Morishige, K., Sakata, M., Tasaka, K., & Murata, Y. (2004). Na⁺/H⁺ Exchanger-3 is involved in mouse blastocyst formation. *Journal of Experimental Zoology Part A: Comparative Experimental Biology*, 301(9), 767-775.
- Kester, L., & van Oudenaarden, A. (2018). Single-cell transcriptomics meets lineage tracing. *Cell stem cell*, 23(2), 166-179.
- Kester, L., & van Oudenaarden, A. (2018). Single-cell transcriptomics meets lineage tracing. *Cell stem cell*, 23(2), 166-179.
- Kim, S. H., & Kim, S. H. (2008). Antagonistic effect of EGF on FAK phosphorylation/dephosphorylation

- in a cell. *Cell Biochemistry and Function: Cellular biochemistry and its modulation by active agents or disease*, 26(5), 539-547.
- Kleinman, H. K., McGarvey, M. L., Liotta, L. A., Robey, P. G., Tryggvason, K., & Martin, G. R. (1982). Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. *Biochemistry*, 21(24), 6188-6193.
- Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., ... & Herrmann, B. G. (2017). Antagonistic activities of Sox2 and brachyury control the fate choice of neuro-mesodermal progenitors. *Developmental cell*, 42(5), 514-526.
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), 420-424.
- Kong, W., Bidy, B. A., Kamimoto, K., Amrute, J. M., Butka, E. G., & Morris, S. A. (2020). CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. *Nature protocols*, 15(3), 750-772.
- Kretschmar, K., & Watt, F. M. (2012). Lineage tracing. *Cell*, 148(1-2), 33-45.
- Kumar, N., Richter, J., Cutts, J., Bush, K. T., Trujillo, C., Nigam, S. K., ... & Willert, K. (2015). Generation of an expandable intermediate mesoderm restricted progenitor cell line from human pluripotent stem cells. *Elife*, 4, e08413.
- Kwon, G. S., Viotti, M., & Hadjantonakis, A. K. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Developmental cell*, 15(4), 509-520.
- Luria, S. E., & Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28(6), 491.
- Marikawa, Y., Tamashiro, D. A. A., Fujita, T. C., & Alarcón, V. B. (2009). 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*, 47(2), 93-106.
- McGinnis, C. S., Patterson, D. M., Winkler, J., Conrad, D. N., Hein, M. Y., Srivastava, V., ... & Chow, E. D. (2019). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nature methods*, 16(7), 619-626.
- McKenna, A., & Gagnon, J. A. (2019). Recording development with single cell dynamic lineage tracing. *Development*, 146(12), dev169730.
- McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F., & Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, 353(6298).
- Morgani, S. M., Metzger, J. J., Nichols, J., Siggia, E. D., & Hadjantonakis, A. K. (2018). Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning. *Elife*, 7, e32839.
- Muñoz-Descalzo, S., Hadjantonakis, A. K., & Arias, A. M. (2015, December). WNT/ β -catenin signalling and the dynamics of fate decisions in early mouse embryos and embryonic stem (ES) cells. In *Seminars in cell & developmental biology* (Vol. 47, pp. 101-109). Academic Press
- Naiche, L. A., Holder, N., & Lewandoski, M. (2011). FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proceedings of the National Academy of Sciences*, 108(10), 4018-

4023.

- Ng, E. S., Davis, R. P., Azzola, L., Stanley, E. G., & Elefanty, A. G. (2005). Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. *Blood*, *106*(5), 1601-1603.
- Ng, E. S., Davis, R. P., Azzola, L., Stanley, E. G., & Elefanty, A. G. (2005). Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. *Blood*, *106*(5), 1601-1603.
- Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., & Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell*, *123*(5), 917-929.
- Oginuma, M., Moncuquet, P., Xiong, F., Karoly, E., Chal, J., Guevorkian, K., & Pourquié, O. (2017). A gradient of glycolytic activity coordinates FGF and WNT signaling during elongation of the body axis in amniote embryos. *Developmental cell*, *40*(4), 342-353.
- Oginuma, M., Niwa, Y., Chapman, D. L., & Saga, Y. (2008). Mesp2 and Tbx6 cooperatively create periodic patterns coupled with the clock machinery during mouse somitogenesis. *Development*, *135*(15), 2555-2562.
- Özbudak, E. M., Tassy, O., & Pourquié, O. (2010). Spatiotemporal compartmentalization of key physiological processes during muscle precursor differentiation. *Proceedings of the National Academy of Sciences*, *107*(9), 4224-4229.
- Pei, D., Shu, X., Gassama-Diagne, A., & Thiery, J. P. (2019). Mesenchymal–epithelial transition in development and reprogramming. *Nature Cell Biology*, *21*(1), 44-53.
- Pei, W., Feyerabend, T. B., Rössler, J., Wang, X., Postrach, D., Busch, K., ... & Chen, W. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*, *548*(7668), 456-460.
- Perea-Gomez, A., Vella, F. D., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., ... & Behringer, R. R. (2002). Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Developmental cell*, *3*(5), 745-756.
- Peyriéras, N., Hyafil, F., Louvard, D., Ploegh, H. L., & Jacob, F. (1983). Uvomorulin: a nonintegral membrane protein of early mouse embryo. *Proceedings of the National Academy of Sciences*, *80*(20), 6274-6277.
- Piotrowska, K., & Zernicka-Goetz, M. (2001). Role for sperm in spatial patterning of the early mouse embryo. *Nature*, *409*(6819), 517-521.
- Pourquié, O. (2011). Segmentation of the vertebrate spine: from clock to scoliosis. *Cell*, *145*(5), 650.
- Purves, D. and Williams, S., 2001. *Neuroscience*. 2Nd Edition. Sinauer Associates. Chapter: The Initial Formation of the Nervous System: Gastrulation and Neurulation
- Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., & Schier, A. F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nature biotechnology*, *36*(5), 442-450.
- Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., ... & Schier, A. F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nature*

- biotechnology*, 36(5), 442-450.
- Ralston, A., & Rossant, J. (2005). Genetic regulation of stem cell origins in the mouse embryo. *Clinical genetics*, 68(2), 106-112.
- Richardson, B. E., & Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nature reviews Molecular cell biology*, 11(1), 37-49.
- Rivera-Pérez, J. A., & Hadjantonakis, A. K. (2015). The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harbor perspectives in biology*, 7(11), a015867.
- Rivron, N., Frias-Aldeguer, J., Vrij, E., Boisset, J., Korving, J., & Vivié, J. et al. (2018). Blastocyst-like structures generated solely from stem cells. *Nature*, 557(7703), 106-111. doi: 10.1038/s41586-018-0051-0
- Rodríguez-Hernández, C. O., Torres-García, S. E., Olvera-Sandoval, C., Ramírez-Castillo, F. Y., Muro, A. L., Avelar-Gonzalez, F. J., & Guerrero-Barrera, A. L. (2014). Cell culture: history, development and prospects. *Int J Curr Res Aca Rev*, 2(12), 188-200.
- Saga, Y. (2012). The mechanism of somite formation in mice. *Current opinion in genetics & development*, 22(4), 331-338.
- Saga, Y. (2012). The mechanism of somite formation in mice. *Current opinion in genetics & development*, 22(4), 331-338.
- Saitou, M., & Yamaji, M. (2012). Primordial germ cells in mice. *Cold Spring Harbor perspectives in biology*, 4(11), a008375.
- Saitou, M., & Yamaji, M. (2012). Primordial germ cells in mice. *Cold Spring Harbor perspectives in biology*, 4(11), a008375.
- Sakai, Y., Meno, C., Fujii, H., Nishino, J., Shiratori, H., Saijoh, Y., ... & Hamada, H. (2001). The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the antero-posterior axis within the mouse embryo. *Genes & development*, 15(2), 213-225.
- Shahbazi, M. N., & Zernicka-Goetz, M. (2018). Deconstructing and reconstructing the mouse and human early embryo. *Nature cell biology*, 20(8), 878-887.
- Shahbazi, M., Siggia, E., & Zernicka-Goetz, M. (2019). Self-organization of stem cells into embryos: A window on early mammalian development. *Science*, 364(6444), 948-951. doi: 10.1126/science.aax0164
- Siggia, E. D., & Warmflash, A. (2018). Modeling mammalian gastrulation with embryonic stem cells. In *Current topics in developmental biology* (Vol. 129, pp. 1-23). Academic Press.
- Solnicakrezel L, Sepich D S. Gastrulation: Making and Shaping Germ Layers. *Annual Review of Cell & Developmental Biology*. 2012, 28(28):687.
- Solnica-Krezel, L., & Sepich, D. S. (2012). Gastrulation: making and shaping germ layers. *Annual review of cell and developmental biology*, 28, 687-717.
- Sozen, B., Amadei, G., Cox, A., Wang, R., Na, E., & Czukiewska, S. et al. (2018). Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nature Cell Biology*, 20(8), 979-989. doi: 10.1038/s41556-018-0147-7
- Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., & Junker, J. P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic

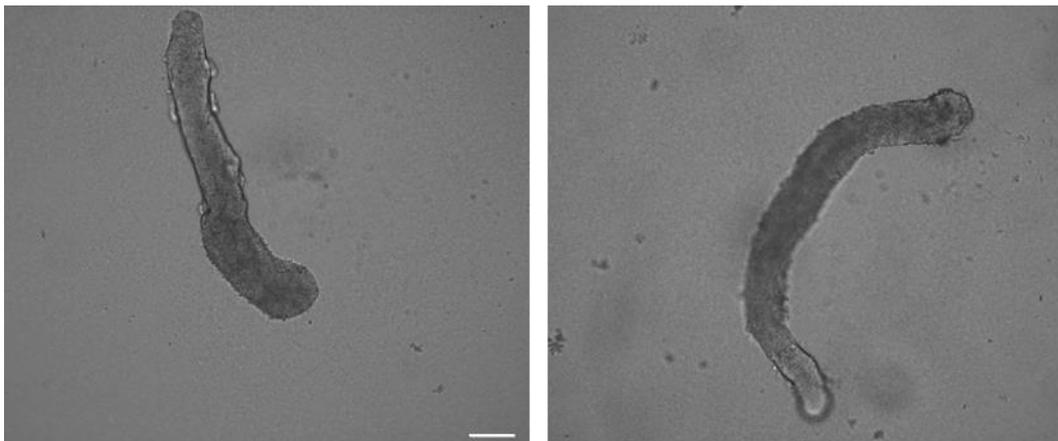
- scars. *Nature biotechnology*, 36(5), 469-473.
- Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., & Junker, J. P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nature biotechnology*, 36(5), 469-473.
- Sparrow, D. B., Chapman, G., Wouters, M. A., Whittcock, N. V., Ellard, S., Fatkin, D., ... & Dunwoodie, S. L. (2006). Mutation of the LUNATIC FRINGE gene in humans causes spondylocostal dysostosis with a severe vertebral phenotype. *The American Journal of Human Genetics*, 78(1), 28-37.
- Spence, J. R., Lauf, R., & Shroyer, N. F. (2011). Vertebrate intestinal endoderm development. *Developmental dynamics*, 240(3), 501-520.
- Spence, J. R., Mayhew, C. N., Rankin, S. A., Kuhar, M. F., Vallance, J. E., Tolle, K., ... & Shroyer, N. F. (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature*, 470(7332), 105-109.
- Spence, J. R., Mayhew, C. N., Rankin, S. A., Kuhar, M. F., Vallance, J. E., Tolle, K., ... & Shroyer, N. F. (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature*, 470(7332), 105-109.
- Srisuparp, S., Strakova, Z., & Fazleabas, A. T. (2001). The role of chorionic gonadotropin (CG) in blastocyst implantation. *Archives of medical research*, 32(6), 627-634.
- Stern, C. D., & Downs, K. M. (2012). The hypoblast (visceral endoderm): an evo-devo perspective. *Development*, 139(6), 1059-1069.
- Sulston, J. E., Schierenberg, E., White, J. G., & Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental biology*, 100(1), 64-119.
- T Das, Atze, Liliane Tenenbaum, and Ben Berkhout. "Tet-on systems for doxycycline-inducible gene expression." *Current gene therapy* 16.3 (2016): 156-167.
- Takahashi, Y., Koizumi, K. I., Takagi, A., Kitajima, S., Inoue, T., Koseki, H., & Saga, Y. (2000). Mesp2 initiates somite segmentation through the Notch signalling pathway. *Nature genetics*, 25(4), 390-396.
- Tam, P. P., & Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mechanisms of development*, 68(1-2), 3-25.
- Tarkowski, A. K. (1959). Experiments on the development of isolated blastomeres of mouse eggs. *Nature*, 184(4695), 1286-1287.
- Tower, J. (2012). Stress and stem cells. *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(6), 789-802.
- Turner, D. A., Alonso-Crisostomo, L., Girgin, M., Baillie-Johnson, P., Glodowski, C. R., Hayward, P. C., ... & Lutolf, M. (2017). Gastruloids develop the three body axes in the absence of extraembryonic tissues and spatially localised signalling. *BioRxiv*, 104539.
- Turner, D. A., Baillie-Johnson, P., & Martinez Arias, A. (2016). Organoids and the genetically encoded self-assembly of embryonic stem cells. *BioEssays*, 38(2), 181-191.
- Turner, D., Girgin, M., Alonso-Crisostomo, L., Trivedi, V., Baillie-Johnson, P., & Glodowski, C. et al. (2017). Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids. *Development*,

144(21), 3894-3906. doi: 10.1242/dev.150391

- Turnpenny, P. D., Alman, B., Cornier, A. S., Giampietro, P. F., Offiah, A., Tassy, O., ... & Dunwoodie, S. (2007). Abnormal vertebral segmentation and the notch signaling pathway in man. *Developmental dynamics: an official publication of the American Association of Anatomists*, 236(6), 1456-1474.
- Van den Brink, S. C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivié, J., ... & van Oudenaarden, A. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature*, 1-5.
- Van den Brink, S. C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A. K., Nowotschin, S., Turner, D. A., & Arias, A. M. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development*, 141(22), 4231-4242.
- Veenvliet, J. V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., ... & Wittler, L. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *bioRxiv*.
- Vukicevic, S., Kleinman, H. K., Luyten, F. P., Roberts, A. B., Roche, N. S., & Reddi, A. H. (1992). Identification of multiple active growth factors in basement membrane Matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components. *Experimental cell research*, 202(1), 1-8.
- Wagner, D. E., & Klein, A. M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. *Nature Reviews Genetics*, 1-18.
- Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., & Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, 360(6392), 981-987.
- Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D., & Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479).
- Whitman, C. O. (1878). *History of the Egg of Clepsine Previous to Cleavage*. Bartholomew Close.
- Williams, M., Burdsal, C., Periasamy, A., Lewandoski, M., & Sutherland, A. (2012). Mouse primitive streak forms in situ by initiation of epithelial to mesenchymal transition without migration of a cell population. *Developmental Dynamics*, 241(2), 270-283.
- Xie, A. W., Binder, B. Y., Khalil, A. S., Schmitt, S. K., Johnson, H. J., Zacharias, N. A., & Murphy, W. L. (2017). Controlled self-assembly of stem cell aggregates instructs pluripotency and lineage bias. *Scientific reports*, 7(1), 1-15.
- Yang, S., Li, S., & Li, X. J. (2018). Shortening the half-life of Cas9 maintains its gene editing ability and reduces neuronal toxicity. *Cell reports*, 25(10), 2653-2659.
- Yusuf, F., & Brand-Saberi, B. (2006). The eventful somite: patterning, fate determination and cell division in the somite. *Anatomy and embryology*, 211(1), 21-30.
- Zeeman, E. C. (1975). Appendix to J. Cooke. *A. Rev. Biophys. Bioeng*, 4, 210-213.
- Zhang, M., Ma, Z., Selliah, N., Weiss, G., Genin, A., Finkel, T. H., & Cron, R. Q. (2014). The impact of Nucleofection® on the activation state of primary human CD4 T cells. *Journal of immunological methods*, 408, 123-131.

- Zhang, S., Chen, T., Chen, N., Gao, D., Shi, B., & Kong, S. et al. (2019). Implantation initiation of self-assembled embryo-like structures generated using three types of mouse blastocyst-derived stem cells. *Nature Communications*, 10(1). doi: 10.1038/s41467-019-08378-9
- Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., & Yang, S. H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy-Nucleic Acids*, 4, e264.
- Zhong, Z. A., Sun, W., Chen, H., Zhang, H., Yu-an, E. L., Lane, N. E., & Yao, W. (2015). Optimizing tamoxifen-inducible Cre/loxp system to reduce tamoxifen effect on bone turnover in long bones of young mice. *Bone*, 81, 614-619.
- Zijlstra, M., Li, E., Sajjadi, F., Subramani, S., & Jaenisch, R. (1989). Germ-line transmission of a disrupted β 2 microglobulin gene produced by homologous recombination in embryonic stem cells. *Nature*, 342(6248), 435-438.

8. Appendix



Appendix Figure 1 | Cas9- display extensive elongation. Scale bar indicates 100 μ m.

9. List of abbreviations

BMP	Bone morphogenetic
Chi	CHIR99021
CL	Chiron + LDN
CLE	Caudal lateral ectoderm
DE	Definitive endoderm
DMSO	Dimethyl sulfoxide
E	mouse embryonic day
ECM	extracellular matrix
EMT	epithelial-to-mesenchymal transition
ETX	embryonic stem cell, trophoblast stem cell, extraembryonic endoderm
ETX	Gastrulating embryo like structure
FCS	fetal calf serum
GESTALT	genome editing of synthetic target arrays for lineage tracing
GFP	Green fluorescent protein
gRNA	guide RNA
gRNA	Guide RNA
IntBC	Integration barcode
LIF	leukemia inhibitory factor
MEF	Mouse embryonic fibroblast
mESC	mouse embryonic stem cell
MG	Matrigel
nCas	Nickase Cas9
NMP	neuromesodermal progenitor
NSB	Node streak border
PBS	Phosphate buffered saline
PBST-X	PBS containing Triton-X
PGCLCs	Primordial germ cell-like cells
PM	Paraxial mesoderm
pPSM	Posterior

scGESTALT	Single cell GESTALT
Sox2	Sex determining region Y-box 2
T	Brachyury
Tbx6	T-box 6
TE	Trophectoderm
TLS	Trunk-like-structure
TSC	Trophoblast stem cells
UTR	Untranslated region
VE	Visceral endoderm
XEN	Extraembryonic endoderm stem cells

Statement of independent work

Herewith I confirm that I wrote this Master’s Thesis in its entirety and that no additional assistance was provided, other than from the sources listed.

Date, Signature

Acknowledgement

I would like to thank

Prof. Dr. Bernhard G. Herrmann & Prof. Dr. Alexander Meissner for giving me the chance to be part of this incredible project and accepting to be my evaluators.

Dr. Jesse Veenliet for being my supervisor and accepting me to be a part of his team. For his patience, supervision as well as his support during the writing of this project, also for sharing his extensive knowledge with me.

Adriano Bolondi for his supervision throughout the project, for his patience, for answering all my questions at any time point and for teaching me all the molecular techniques, microscopy, inkscape and FACS.

Dr. Frederic Koch for helping me with all my questions.

Afrah Ghauri for her support throughout my thesis and master program.

I would like to thank the whole **department** for their support and kindness.