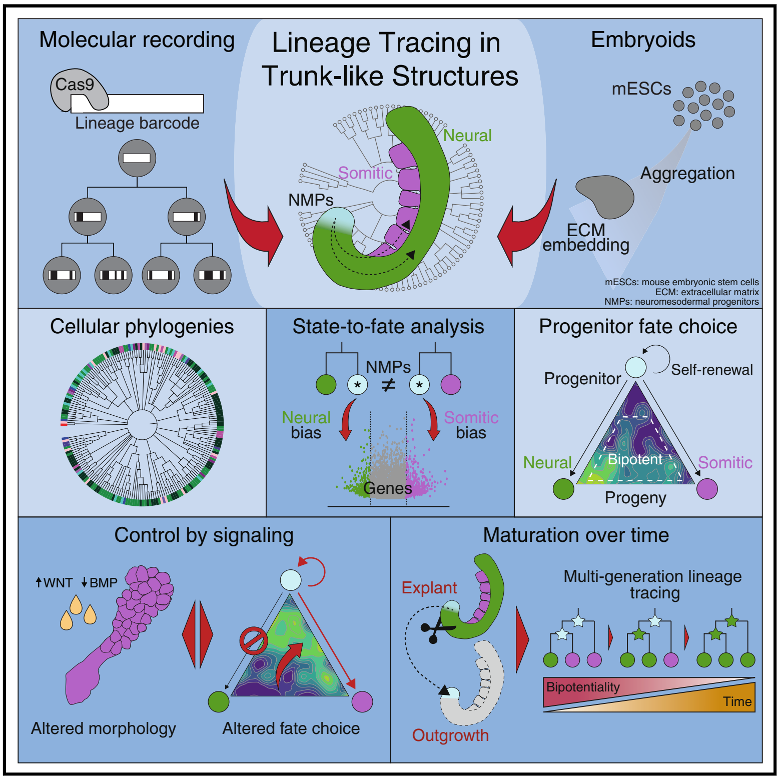
Article

Developmental Cell

Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids

Graphical abstract



Highlights

- Molecular recording in embryoids reconstructs high-resolution cellular phylogenies
- Progenitor cells share transcriptional similarities with their recent progeny
- Single axial progenitors make plastic decisions that converge to a robust form
- Axial progenitors functionally resolve toward a single fate over time

Authors

Adriano Bolondi, Benjamin K. Law,
Helene Kretzmer, ...,
Alexander Meissner, Zachary D. Smith,
Michelle M. Chan

Correspondence

meissner@molgen.mpg.de (A.M.),
z.smith@yale.edu (Z.D.S.),
mmchan@princeton.edu (M.M.C.)

In brief

Bolondi, Law, et al. engineer a single-cell recorder to reconstruct progenitor dynamics within mouse stem cell-derived embryoids. They characterize the variable fates of axial progenitors as they contribute to neural and somitic lineages, including how cellular decisions relate to transcriptional heterogeneity, the signaling microenvironment, and differentiation potential over developmental time.

Bolondi et al., 2024, Developmental Cell *59*, 1–17
June 17, 2024 © 2024 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.devcel.2024.03.024

CellPress

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Developmental Cell

CellPress

Article

Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids

Adriano Bolondi,[1,11] Benjamin K. Law,[2,3,11] Helene Kretzmer,[1] Seher Ipek Gassaloglu,[1,4] René Buschow,[5] Christina Riemenschneider,[1] Dian Yang,[6] Maria Walther,[1] Jesse V. Veenvliet,[4,7,8] Alexander Meissner,[1,9,*] Zachary D. Smith,[10,*] and Michelle M. Chan[2,3,12,*]

[1]Department of Genome Regulation, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
[3]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
[4]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[5]Microscopy Core Facility, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[6]Department of Molecular Pharmacology and Therapeutics & Systems Biology, Columbia University, New York, NY 10032, USA
[7]Cluster of Excellence Physics of Life, Technische Universität Dresden, 01307 Dresden, Germany
[8]Center for Systems Biology Dresden, 01307 Dresden, Germany
[9]Institute of Chemistry and Biochemistry, Freie Universität Berlin, 14195 Berlin, Germany
[10]Department of Genetics, Yale Stem Cell Center, Yale School of Medicine, New Haven, CT 06519, USA
[11]These authors contributed equally
[12]Lead contact
*Correspondence: meissner@molgen.mpg.de (A.M.), z.smith@yale.edu (Z.D.S.), mmchan@princeton.edu (M.M.C.)
https://doi.org/10.1016/j.devcel.2024.03.024

SUMMARY

Embryogenesis requires substantial coordination to translate genetic programs to the collective behavior of differentiating cells, but understanding how cellular decisions control tissue morphology remains conceptually and technically challenging. Here, we combine continuous Cas9-based molecular recording with a mouse embryonic stem cell-based model of the embryonic trunk to build single-cell phylogenies that describe the behavior of transient, multipotent neuro-mesodermal progenitors (NMPs) as they commit into neural and somitic cell types. We find that NMPs show subtle transcriptional signatures related to their recent differentiation and contribute to downstream lineages through a surprisingly broad distribution of individual fate outcomes. Although decision-making can be heavily influenced by environmental cues to induce morphological phenotypes, axial progenitors intrinsically mature over developmental time to favor the neural lineage. Using these data, we present an experimental and analytical framework for exploring the non-homeostatic dynamics of transient progenitor populations as they shape complex tissues during critical developmental windows.

INTRODUCTION

Although the core genetic pathways that underly early body plan formation are generally conserved, how these mechanisms control populations of differentiating cells remains a daunting scientific challenge. Unlike the deterministic and stereotypical programs found in many invertebrate embryos, such as the worm *C. elegans*[1] or the ascidian *P. mammillata*,[2] organisms such as mammals continuously self-regulate their spatiotemporal patterning. In these more regulative regimes, development precisely coordinates growth and differentiation to ensure proper allocation to downstream lineages.[3] From this perspective, progenitor fields are analogous to adult stem cell systems, which maintain long-term homeostasis through complex cellular interactions[4–6] but must encode these processes differently to robustly control morphogenetic processes within finite windows.

Understanding transient progenitor cell behavior requires holistic, system-wide quantification of their differentiation patterns across many biological replicates.[7,8] Because continuous examination of these populations is challenging, particularly in mammalian embryos, it remains unclear how developmental potential connects to ultimate fate, nor is it obvious how individual cellular decisions aggregate to ensure robust developmental programs.[8,9] The recent implementation of retrospective, high-content and single-cell genetic lineage tracing methods offers a new lens to investigate these concepts, having previously been applied to mouse embryogenesis,[10–13] tumorigenesis,[14–17] and other cell fate transitions.[10,11,18–20] However, they have not yet advanced far beyond the reconstruction of general cellular pathways or the identification of new marker genes, leaving a detailed accounting of the population-level dynamics of differentiating cells comparatively unexplored.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Stem cell-derived embryo models provide unique opportunities to study the complex behaviors of differentiating cells.[21–23] Of these, embryonic trunk-like structures (TLSs) develop from a population of multipotent neuro-mesodermal progenitors (NMPs).[24] In vertebrates, NMPs emerge in the posterior portion of the gastrulating embryo alongside fine-tuned WNT, fibroblast growth factor (FGF), retinoic acid, and other signaling gradients.[24–32] Classic labeling experiments have suggested that these progenitors functionally contribute to both the neural tube and the paraxial mesoderm tissues[29,33,34] and do so through a unique transcriptional state that utilizes factors from both programs.[26,35,36] Although the painstaking evaluation of a limited number of individual cells has established NMP ‘‘bipotentiality’’ as a proof of concept, these experiments have not addressed how these cell populations reliably produce a highly complex multi-lineage structure within a few days. As such, NMPs represent an ideal experimental model to ask general questions regarding the regulatory, cellular, and morphological dynamics of transient progenitor fields.

Here, we engineer an optimized, evolving Cas9-based molecular recorder to generate single-cell-resolved cellular phylogenies of TLS formation. Our tree-based analyses identify a notably broad distribution of individual progenitor behaviors, suggesting flexibility at the level of single cells that nonetheless converge to produce macroscopically consistent compositions and morphologies. We characterize progenitor self-renewal and differentiation dynamics as well as demonstrate the degree to which both parameters can be modulated to induce large morphological phenotypes. Finally, we generate interconnected trees produced from the same progenitor population to evaluate NMP stability, which we find resolves over time toward neural fates. Collectively, our work presents a general framework for describing progenitor cell behavior at the level of their decision-making, including their intrinsic properties and modular nature.

RESULTS

High-content lineage recording in complex stem cell-based models

To establish a reproducible, high-content molecular recording system for stem cell-based differentiation protocols, we optimized a Cas9-based lineage tracing technology previously applied to the study of mouse embryogenesis and cancer.[11,14,16] Our system builds lineage relationships from continuous, randomly generated, and heritable indels using three components: a lox-stop-lox (LSL)-Cas9-T2A-GFP cassette, an array of single guide RNAs (gRNAs), and an editable ‘‘target’’ sequence encoded within the 3′ UTR of a constitutively expressed transgene (Figure 1A). Each guide carries a single nucleotide mismatch to its respective target site (TS) to titrate the rate of indel generation (Figure S1A). To maximize information capacity, we randomly integrated multiple barcoded TS cassettes using piggyBAC transposition and selected a subclone containing 10 TSs (hereafter Tracer mouse embryonic stem cells [mESCs]) that displayed stable transgene expression, recovery of mutational information from the transcriptome, and minimal effects of recording on cell fitness (Figures 1B–1D and S1B–S1I).

The recording kinetics of our Tracer mESC line match the timescale of TLS generation, which proceeds through a gastrulation-like window over the first 4 days and rapidly elongates into the differentiated compartments of the embryonic trunk over the final day[11,14,16,24,37] (Figure 1E). To initiate recording, we delivered Cre (or an empty *mock* control plasmid) during aggregate formation (day 0, 500 cells per aggregate) and confirmed normal TLS generation and recorder activity using a combination of live imaging, immunofluorescence, and fluorescence-activated cell sorting (FACS) (Figures S1J–S1N; Video S1). We see clear spatial segregation of the expected embryonic lineages, including SOX2$^+$, T$^+$, CDX2$^+$ putative NMPs; SOX1$^+$ neural tube; and SOX17$^+$ endoderm (Figure 1F, upper panels). We also observe apically localized NCAD in somitic and neural tube cells.[24] Combined imaging and flow cytometry of 188 structures confirm that Cas9-GFP$^+$ cells consistently contribute to mature TLSs in an unbiased ‘‘salt-and-pepper’’ fashion, with minimal correlation between the fraction of actively recording cells and gross morphology (Figures 1F, S1F, S1O, and S1P). Individual Tracer TLS replicates further generate high numbers of unique ‘‘alleles’’ (n = 9,248–14,765 unique alleles per structure, Figure S1Q). Tracer mESCs can also be used more broadly to derive additional stem cell models of development using the same basic strategy, including the induction of embryoid bodies, gastruloids, and cardiac gastruloids[38] (Figures 1G and 1H; Video S2).

Reconstruction of structure-wide single-cell phylogenies

To reconstruct lineage relationships, we performed single-cell RNA sequencing (scRNA-seq) in Tracer TLSs, including targeted amplification of lineage barcodes (Figure 2A). We recovered 3,577 and 5,122 cells from two individual replicates and assigned them to distinct transcriptional states using a previously published scRNA-seq reference[24], including undifferentiated NMPs (extant NMPs), early and advanced neural tube as well as paraxial mesoderm and their more mature somitic cell types (i.e., sclerotome-like and dermomyotome-like). As a model of the posterior embryo, TLSs also produce other primitive streak derivatives, such as the endoderm, endothelium, and primordial germ cell-like cells (PGCLCs)[24] (Figure 2B).

Our initial replicates were composed of Cas9-GFP$^+$ (recording) and Cas9-GFP$^-$ (control) cells, allowing us to evaluate recording performance and lineage representation between the two subpopulations. For TLS1 (median UMIs = 27,139, median genes = 4,985), we recovered 1,221 Cas9-GFP transcript-positive cells (34.1%), which were similarly distributed across TLS cell states compared with GFP$^-$ cells (Figures S2A–S2E). Globally, 1,836 cells in TLS1 acquired at least one edit (51.3%, 4.67 mean indels per cell). For TLS1, we recovered 464 indels, 722 alleles, and 968 complete lineage barcodes, respectively, with many alleles shared across cell states as an indication of common ancestry (Figures S2A and S2F–S2L).

Next, we constructed complete single-cell-resolved phylogenies of TLS development using Cassiopeia, which builds lineage trees by maximum parsimony (Figures S2M and S2N; STAR Methods).[39] Our optimal tree harbors 665 ‘‘nodes’’ (representing bifurcating trajectory events), with a maximum and average tree depth of 9 and 3.39 nodes, respectively (Figures 2C and S2M–S2O). We further tested the statistical significance of shared ancestry by comparing our tree against 500 randomly shuffled backgrounds (Figure S2P). Notably, our second

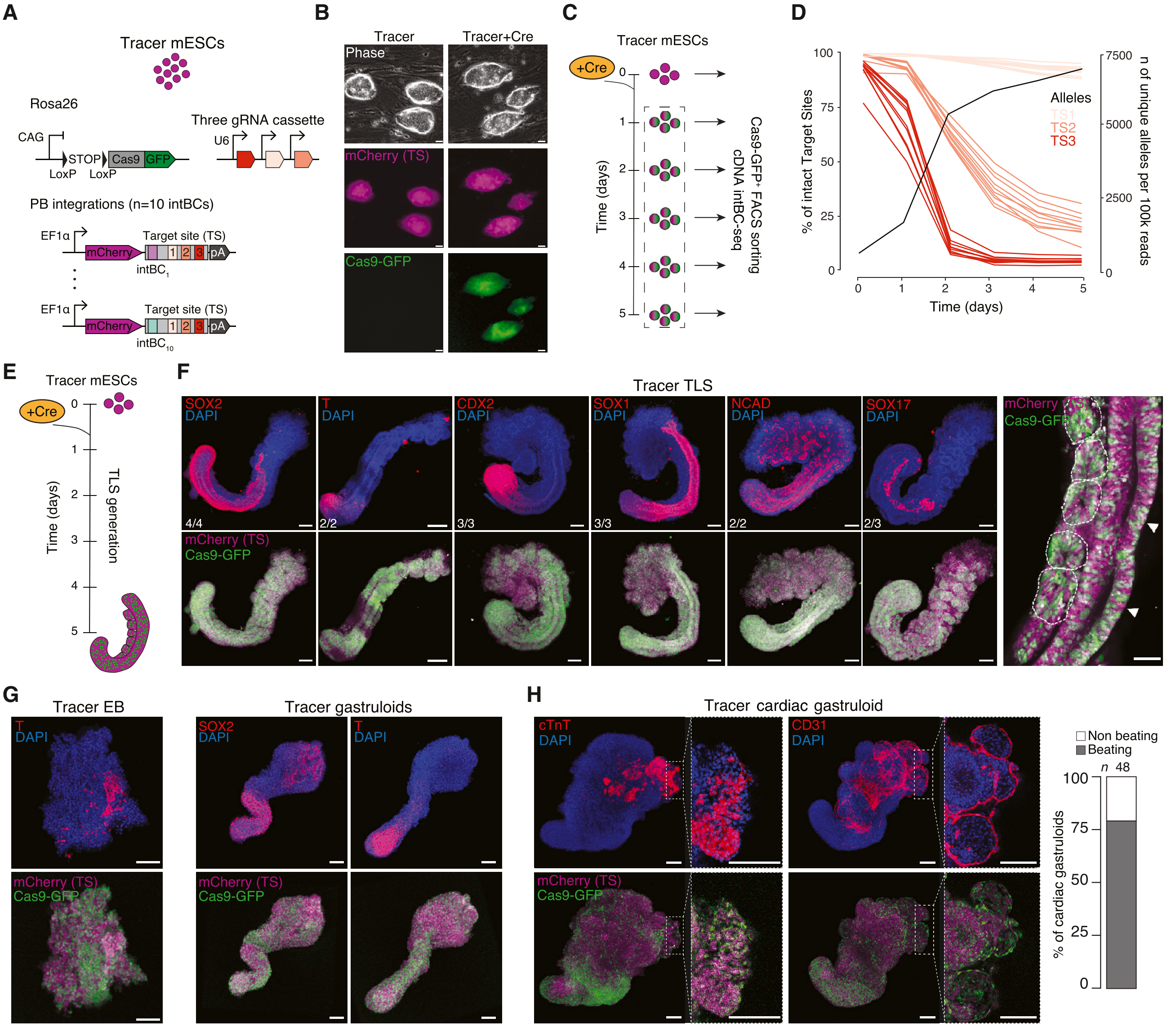


Figure 1. Molecular recording in trunk-like structures

(A) Schematic representation of the Tracer mESC line. intBC, integration barcode; pA, polyadenylation signal; PB, piggyBAC.

(B) Wide-field fluorescent and bright-field images of Tracer mESCs before and after Cre-mediated Cas9 induction. Scale bars, 20 μm.

(C) Workflow for evaluating editing efficiency across 5 days.

(D) Percentage of intact target sites (TSs) (red, right axis) and number of unique alleles (black, defined as the combined mutational profile of the three target sites within the cassette) in activated Tracer mESCs. Each line represents a target site cassette connected to a single intBC.

(E) Workflow for trunk-like structure (TLS) generation from Tracer mESCs.

(F) Confocal sections showing marker gene localization in Tracer TLS alongside TS (mCherry) and Cas9 (GFP) expression. Number of analyzed structures indicated on the bottom left. Scale bars, 100 μm and 50 μm for the magnification.

(G) Maximum intensity projections showing SOX2 and T localization in Tracer embryoid bodies (EB) and gastruloids alongside TS (mCherry) and Cas9 (GFP) expression. 43/48 (89.5%) Tracer gastruloids elongated. Scale bars, 100 μm.

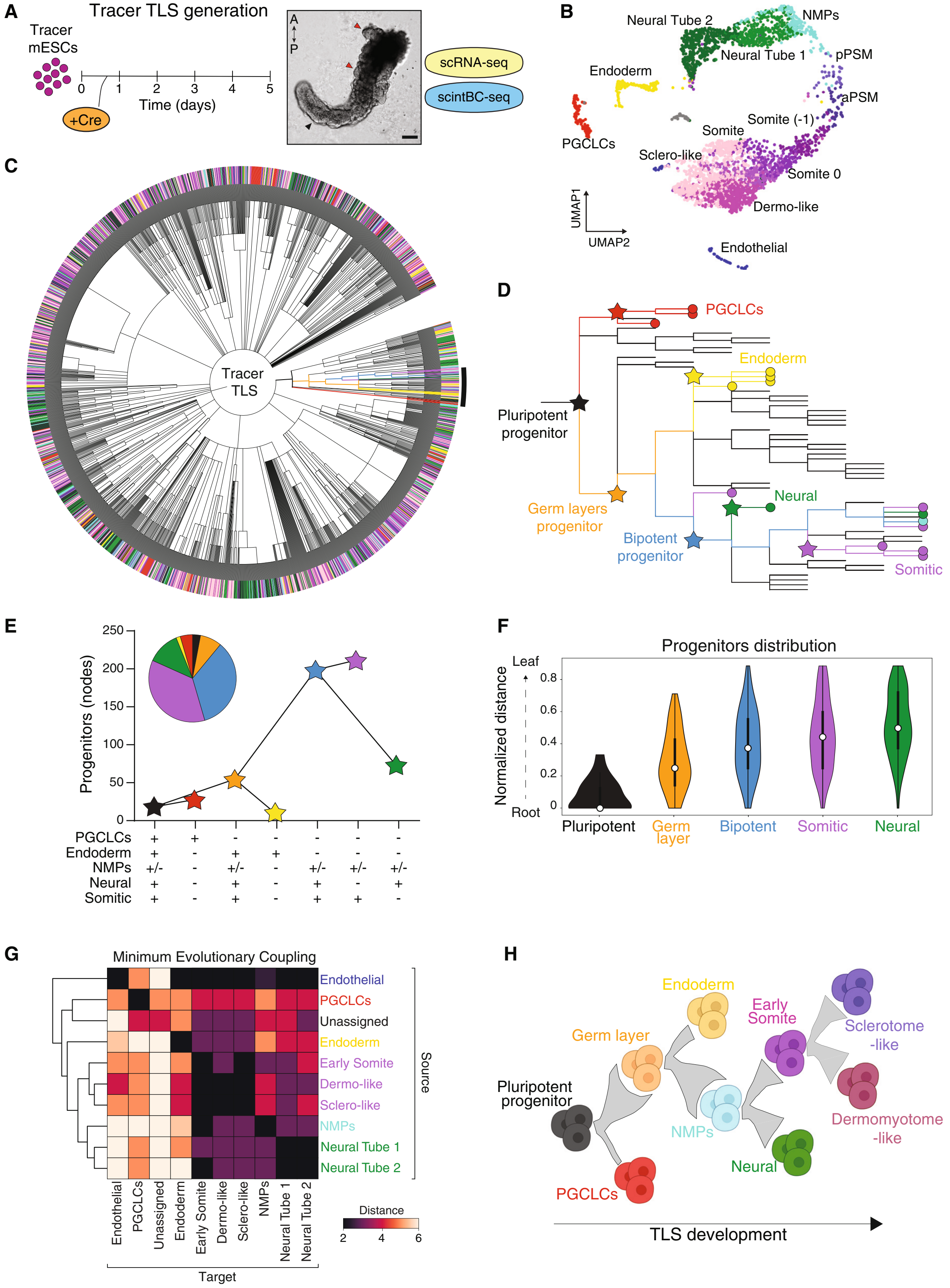
(H) Maximum intensity projections showing cardiac troponin (cTnT) and CD31 localization in Tracer cardiac gastruloids alongside TS (mCherry) and Cas9 (GFP) expression. Magnifications show the development of a cTnT$^+$ cardiac mesoderm domain and of CD31$^+$ hemogenic endothelium. 38/48 (79.1%) Tracer cardiac gastruloids displayed a beating contractile domain. Scale bars, 100 μm.

See also Figure S1.

replicate showed highly similar results (Tracer TLS2) (Figures S2Q–S2W).

TLS trees show multiple branches emanating directly from the root, which we term “clones” that likely reflect individual mESC founder cells (Figures S3A–S3C, n = 53 and 85 clones for TLS1 and 2, respectively). Individual clones are complex and variable, including in the overall number of nodes, cells within a clone and cell states generated (Figures S3A–S3C). Generally, the number

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024



(legend on next page)

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

of different cell states produced by an ancestral clone correlates with the number of total progeny, with larger clones composed of more cell states (Figures S3C and S3D). However, both TLS replicates contained several outlier clones comprised solely of PGCLCs, suggesting an early restriction event consistent with *in vivo* observations[40,41] (Figure S3D).

The neural and somitic lineages are connected through a common progenitor

To derive general properties about TLS development, we examined features of individual nodes, including their overall composition as they proceed from root to leaf (Figures 2D and S3E). Each node represents a branching event from a progenitor cell to its daughters, such that steps where cell-state composition changes reflect captured differentiation events. Using this concept, we classified our nodes according to their developmental output, including "pluripotent" (comprising all major lineages and cell states), "germ layer restricted" (producing all three germ layers, but excluding PGCLCs), "bipotent" (capturing neural and somitic lineages), or "lineage specific" (only generating states belonging to PGCLCs, endoderm, neural, or somitic) (Figures 2D and S3E). This categorization shows a clear, gradual, and lineage-specific restriction of developmental potential along the paths of the tree (Figures 2D–2F, S3E, and S3F). In particular, PGCLC-containing nodes are located near the root, suggesting early and irreversible commitment of this cell type. Similarly, germ-layer-restricted nodes tend to resolve into endoderm-specific and bipotent nodes, with the latter branching into neural and somitic restricted trajectories (Figures 2D–2F and S3E).

Notably, bipotent nodes are highly abundant in TLS (n = 203/665 and 226/1,023 for Tracer TLS1 and 2, respectively), supporting the existence of a large bipotent progenitor pool as described for NMPs *in vivo* and *in vitro*[24,25,29,32,33] (Figures 2D–2F and S3E). Neural and somitic connectivity through NMPs is also statistically supported against a null, randomly shuffled model that holds cell type composition and tree topology constant and shuffles the identity of cells around the tree (p value = 0.039, STAR Methods). This analysis also supported the gradual and hierarchical restriction of differentiation outcomes from root to leaf, with pluripotent progenitors rapidly diminishing from the roots forward, germ-layer progenitors enriched within the early portions of the trees, bipotent progenitors occupying more internal branches, and both neural and somitic progenitors existing closer to the leaves (Figure S4F). By aggregating the minimal path distance between every cell and its closest neighbor for each transcriptional state ("minimal evolutionary coupling," STAR Methods), we find that both PGCLCs and the endoderm are clear outliers. In contrast, major embryonic trunk derivatives are more closely interconnected, with short but distinct somitic and neural clusters that appear to interact through extant NMPs (Figures 2G, S3G, and S3H).

Overall, our second replicate recovered similar relationships, with a key exception being an evident switch in the connectivity between endothelial cells and somites, which deeper investigation highlighted to be specific to the sclerotome-like subpopulation (Figures S3G, S3I, and S3J). Importantly, a somitic origin for some endothelial cells has been observed *in vivo* to ensure functional diversification of vasculature in developing trunk muscles, which would be consistent with the endothelium's changing relationship to the embryonic trunk as structures mature.[42,43] Taken together, our lineage data highlight the utility of molecular recorders to generically reconstruct sequential transitions in lineage potential. We find that the relationship between cell states along the reconstructed tree provides a simple *in toto* assessment of both hierarchical and convergent developmental trajectories, such as those that shape the developing trunk (Figure 2H).

NMPs share transcriptional signatures with their recent progeny

Our investigation supports NMPs as a major progenitor population that creates both the neural and somitic lineages. We next sought to ask questions related to the dynamic relationship between progenitor populations and their differentiated products, both in terms of how individual cells relate to one another over time as well as how individual paths combine to form complex structures. These two considerations, the local decision-making process of a single progenitor cell and their cumulative ability to ensure proper morphogenesis, are key features of developmental plasticity.

To improve our ability to measure these parameters, we generated a comprehensive, replicate-powered dataset via a multiplexing approach (Multi-seq).[44] We also included an orthogonal LentiBC library (n = 2.7 × 10^5 unique barcodes [BCs]) to validate the performance of our tree-building algorithm. Overall, we recovered 11 TLS replicates after FACS purification of GFP$^+$ cells, sequencing, and quality control, with 226.9 mean cells per

Figure 2. Reconstructing developmental pathways and clonal histories

(A) Workflow for single-cell RNA sequencing (scRNA-seq) recovery of (1) transcriptional state and (2) "lineage barcode" (defined as the combination of the mutation profiles of the ten target site cassettes) in each cell. Bright-field image of Tracer TLS replicate 1. Black arrowhead, neural tube. Red arrowhead, somites. A, anterior; P, posterior. Scale bars, 100 μm.

(B) UMAP (uniform manifold approximation and projection) of Tracer TLS colored by 14 previously defined cell states from all three germ layers and the germline.

(C) Reconstructed lineage tree from 1,836 Tracer TLS cells, comprising 665 internal nodes. The colorized ring indicates the cell identity of each leaf (colors as in B). Black arc and colorized branches highlight the subclone displayed in (D).

(D) Phylogenetic subtree demonstrating the progressive restriction into the paraxial mesoderm and neural ectoderm from an initially pluripotent progenitor. Stars indicate progenitor nodes colored based on their progeny.

(E) Number of each progenitor node/class in the Tracer TLS tree. Pie chart highlights the progenitor states' distribution.

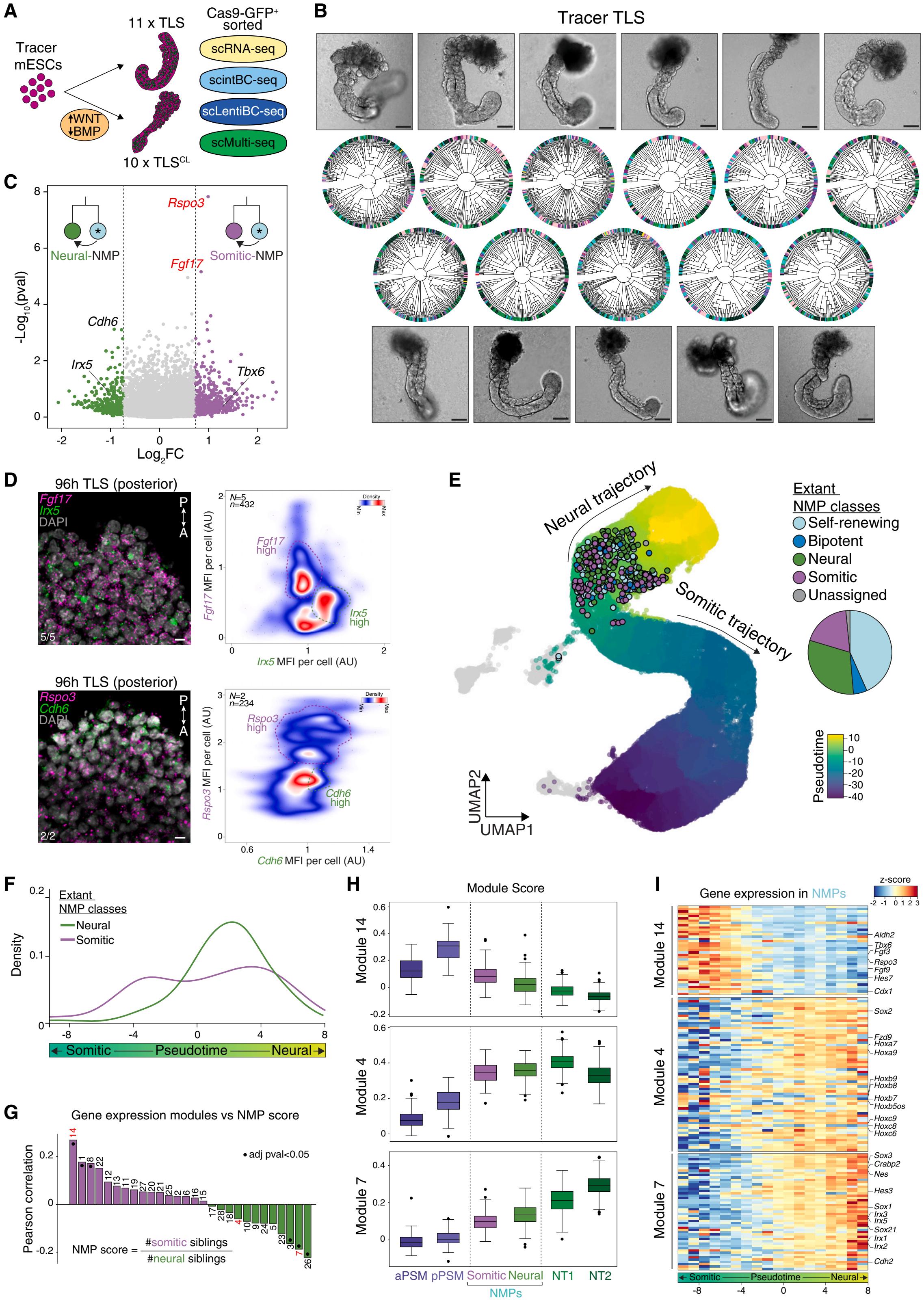
(F) Violin plot showing the clone-normalized distance for each progenitor class as they are distributed from the root to leaf. White dot represents the median, boxes the interquartile range, and edges the full range.

(G) Minimum evolutionary coupling between cell states in Tracer TLS. Heatmap shows the median minimum pairwise distance for each cell ("source") to its own and other cell states ("target", STAR Methods). Hierarchical clustering recapitulates an expected developmental trajectory for the embryonic trunk.

(H) Schematic representation of sequential differentiation events as extracted from our tree-based reconstruction.

See also Figures S2 and S3.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024



(legend on next page)

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

replicate (Figures 3A, 3B, and S4A–S4F). We generated an additional experimental cohort treated with WNT activator (CHIR99021) and BMP inhibitor (LDN193189) over the final 24 h of TLS development (hereafter "Tracer TLS^{CL}"), a perturbation that results in a strong morphological phenotype[24] that we will describe in greater detail below.

Mature TLSs still include small populations of undifferentiated, extant NMPs that show substantial connectivity to differentiated cell types within the tree. Because cell potential and molecular state cannot be measured simultaneously, we instead estimated ‘‘state-to-fate’’ relationships by evaluating individual NMP transcriptomes according to their closest phylogenetic neighbors as a proxy for their most recent differentiation event (Figure S4G). We assigned NMPs into five distinct classes: (1) self-renewing (if alone in a branch or only connected to other extant NMPs), (2) bipotent (if connected to both neural and somitic cells), (3) neural, (4) somitic, and (5) unassigned (if connected with one of the other TLS states) (Figure S4G). Overall, NMPs in different classes show a high degree of transcriptional similarity (including expression of canonical NMP-defining genes such as *Sox2*, *T*, *Cdx2*, and *Nkx1-2*).[36] We could nonetheless identify genes that were reciprocally up- or down-regulated when comparing neural against somitic NMPs, including WNT-associated ligand R-spondin-3 (*Rspo3*) and fibroblast growth factor 17 (*Fgf17*) in NMPs with somitic neighbors and genes such as Iroquois homeobox 5 (*Irx5*) and Cadherin 6 (*Cdh6*) in neural NMPs (Figure 3C). Notably, these genes are heterogeneously expressed within individual cells in the posterior region of 96-h TLSs, as measured by RNA fluorescence *in situ* hybridization (RNA-FISH), suggesting subtle molecular fluctuations within the NMP pool that may reflect their changing potential (Figure 3D).

To improve our resolution of the differentiation process, we integrated our Tracer TLS data with a transcriptional pseudotime axis of TLS formation from 96 to 120 h[24] (Figures 3E and S4H; STAR Methods). There, our distinct NMP classes also show subtle but consistent transcriptional differences, with somitic NMPs sharing higher transcriptional similarity to the somitic lineage compared with those with neural neighbors, and vice versa (Figures 3F and S4I). Somitic and neural cells that stem from bipotent nodes are also transcriptionally more similar to NMPs than those connected to other differentiated cells (Figure S4J). We then computed discrete transcriptional modules as they varied across our axis and identified six with significantly positive or negative correlations between NMP transcriptional state and local connectivity to somitic or neural cells (Figure 3G; Table S1, adjusted p value < 0.05). Of these, modules 14 and 7 were generally enriched within NMPs but nonetheless contained many regulators of somitogenesis (module 14) or neural induction (module 7) (Figures 3H, 3I, and S4I). Although these transcriptional programs appear to reflect shared transcriptional relationships between NMPs and their recently differentiated progeny, not every module displayed the same fate-based relationship. For example, module 4 is composed of posterior Hox genes and represents a shared, comparatively constant transcriptional program across all NMP groups (Figures 3F–3I and S4K).

Overall, our results highlight the value of intersecting transcriptional and lineage measurements to establish state-to-fate relationships between highly dynamic progenitors and their recent cellular progeny, including the ability to identify distinct transcriptional signatures that may explain their past differentiation.

Individual progenitor cell decisions are weakly constrained

Our extant progenitor analysis highlights the degree to which transient progenitors such as NMPs may initiate divergent transcriptional programs as part of their differentiation. Macroscopically, we expect that these individual cellular decisions aggregate to form consistent TLSs with reproducible morphological, cellular, and transcriptional properties.[24] Moreover, individual progenitor paths must be rapidly integrated given that the majority of the morphogenetic events associated with axial elongation in TLSs occur over the final 2 days, leaving little room for error correction or compensation.

Figure 3. Measuring transcriptional dynamics that reflect lineage allocation

(A) Workflow to evaluate progenitor dynamics in TLS. Multiple Tracer TLS and TLS^{CL} (treated with CHIR and LDN) replicates were generated. Cas9-GFP$^+$ cells are isolated from individual structures and used for Multi-seq-based scRNA-seq to simultaneously recover transcriptomic, lineage, clonal, and replicate information.

(B) Bright-field images of 11 Tracer TLS replicates alongside their reconstructed phylogenetic trees. Scale bars, 100 μm.

(C) Volcano plot highlighting gene expression differences between neural and somitic NMPs, defined as indicated in the schematic. Genes with a $\log_2FC < -0.75$ or > 0.75 between the two classes are highlighted in green and magenta, respectively. Significantly differentially expressed genes (q value < 0.05) are colored in red.

(D) RNA-FISH of genes that distinguish neural and somitic NMPs within the posterior portion of 96 h TLS. Images are shown alongside density plots of normalized mean fluorescent intensities (MFI) per cell. N, number of structures; n, number of cells; A, anterior; P, posterior; AU, arbitrary unit. Scale bars, 5 μm.

(E) UMAP showing the relative position of the five extant NMP classes along the pseudotime axis. Pie chart indicates the proportion of the five NMP classes across replicate TLSs. Self-renewing (n = 164), bipotent (n = 20), neural (n = 117), somitic (n = 71), and unassigned (n = 6).

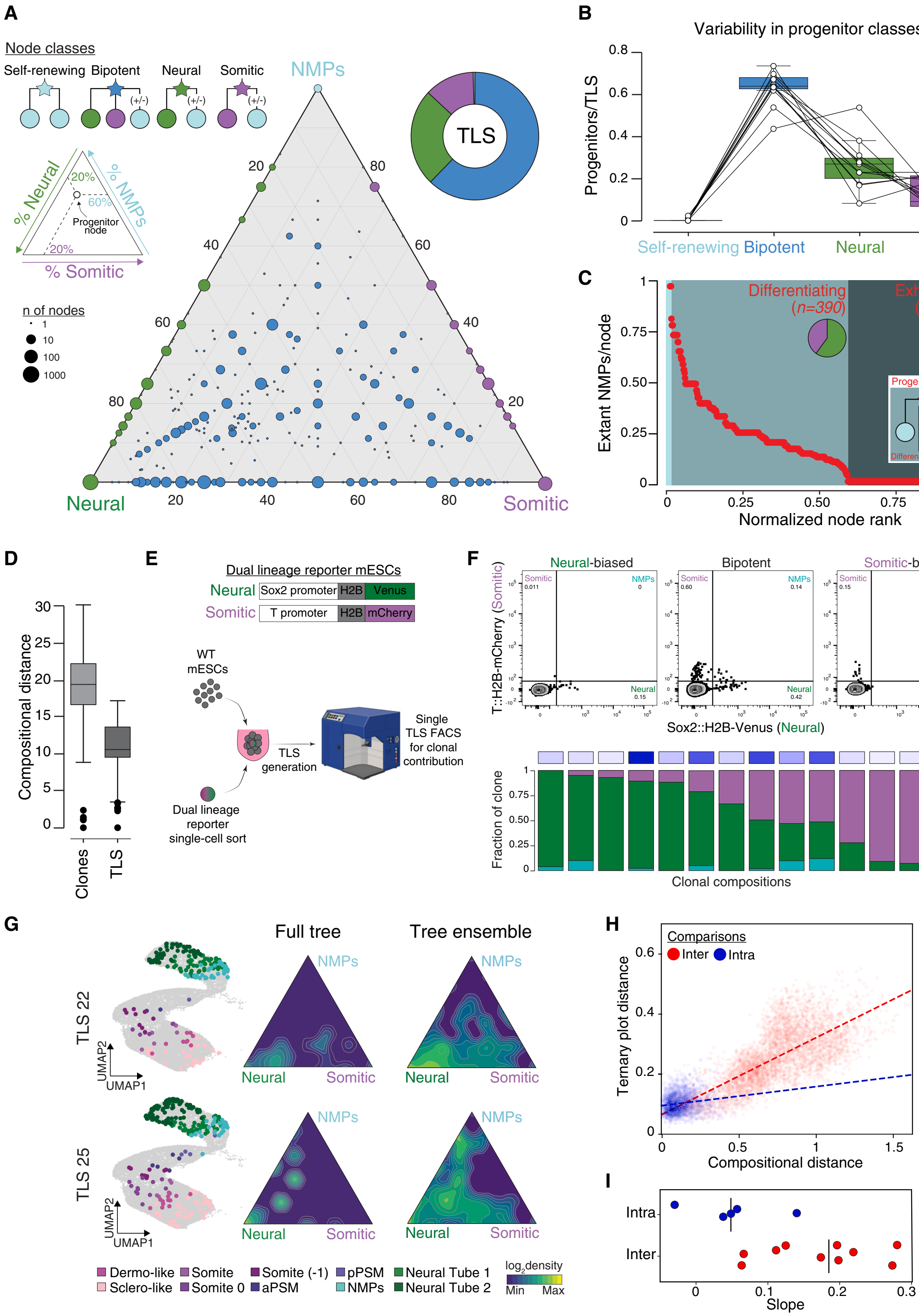
(F) Density curves showing the distribution of neural and somitic NMPs along the pseudotime axis. Axis is restricted to a window encompassing NMPs and their immediate progeny (neural tube 1 and posterior presomitic mesoderm).

(G) Pearson correlation between indicated transcriptional modules and the fraction of somitic and neural neighbors for individual NMPs (calculated as a log-ratio between somitic and neural siblings for each node, STAR Methods). Statistically significant modules (adjusted p value < 0.05) are marked with a black dot. Red, modules used for subsequent analysis. See Table S1 for full list of genes in each module.

(H) Boxplots showing module scores in neural and somitic NMPs and for closest derivative cell states. Module 14 has a higher score in somitic NMPs as compared with neural, while module 7 shows the opposite trend. Boxes indicate 25th and 75th percentiles, central line the median, and whiskers 10th and 90th percentiles. Outliers are shown. aPSM, anterior presomitic mesoderm; pPSM, posterior presomitic mesoderm; NT1, NeuralTube 1; NT2, NeuralTube 2.

(I) Row-normalized z score heatmap showing the expression of genes belonging to the three highlighted modules along the pseudotime axis (unassigned NMPs are excluded). Key genes involved in somitogenesis (module 14), axial organization (module 4), and neural induction (module 7) are indicated.

See also Figure S4.



(legend on next page)

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

To address how individual progenitor cells contribute to TLS composition, we described each of the TLS replicates profiled in our multi-structure dataset according to the decision-making of individual nodes, which represent ancestral progenitor cells. We classified each "transient progenitor node" according to the cell types it produced, including: (1) self-renewing (if composed entirely of extant NMPs), (2) bipotent (if producing both neural and somitic lineages), (3) neural, and (4) somitic (Figure 4A, upper left). To ensure that structures are described by historical progenitors and not dominated by leaves, we focused on nodes composed of ≥4 cells (which removed 861/1,526 nodes, 191/2,200 cells, and 28/143 clones in total, Figures S5A–S5D) and excluded those containing other cell types. With the exception of the self-renewing class, we did not include extant NMPs in our progenitor fate assignment to avoid making assumptions about their future potential.

The aggregate path behavior of reconstructed progenitors within a developing TLS can be described using a ternary plot, which summarizes how each historical progenitor node connects the three major trunk cell states by measuring the overall fraction of its progeny that acquire NMP, somitic, and neural fates (Figure 4A). We observe a small but distinct population of self-renewing NMPs at one origin (top corner) and a larger, more broadly distributed set of nodes with variable contributions to the TLS lineages below it (Figure 4A). The aggregate distribution of individual nodes provides further evidence for progenitor bipotentiality, as 62% of nodes produce both somitic and neural cells (bipotent, 43%–73%; neural, 8%–53%; and somitic, 0%–23% nodes per structure, Figures 4B, S5E, and S5F). Similarly, the majority of progenitor nodes (59.1%) also contain extant NMPs, suggesting that they continue to proliferate and differentiate into both lineages throughout axial elongation (Figure 4C). From this perspective, 0.5% of nodes are self-renewing, 58.6% appear to be differentiating (composed of extant NMPs connected to recently differentiated states), and 40.9% reflect "exhaustion" (composed of differentiated cell states only) (Figure 4C). Notably, the proportions of distinct node classes vary between individual TLS replicates, both in terms of their decision-making and their connection to extant NMPs (Figures 4B and S5E–S5I).

Our investigation of reconstructed progenitors confirms that each TLS generally proceeds through a bipotent intermediate but that individual progenitors contribute differently to the neural and somitic lineages (Figures 4B and S5I). In keeping with this, individual clones within a given TLS replicate are substantially more variable than individual replicates are to each other, suggesting that composition may primarily be regulated at the population level (Figure 4D). We experimentally tested clonal variability by sorting a single cell of a dual Sox2/T reporter line that stably labels the neural and somitic lineages[24] into otherwise unlabeled mESCs (Figure 4E). With this approach, all labeled cells in mature TLSs represent direct descendants of the initial seeded clone. FACS analysis on individual replicates recapitulated our tree-based findings, including a broad range of clone sizes and individual contributions to neural and somitic progeny (Figure 4F).

Progenitor plasticity is considered a feature of regulative development and appears to be represented by our path-level observations during axial elongation. For example, despite consistent cellular and morphological outcomes, ternary plots differ substantially between replicates, suggesting that the decision framework of individual progenitor cells—to self-renew or differentiate into one of two distinct outcomes—is somewhat weakly constrained. The repertoire of individual cellular decisions instead appears broad to the point that each TLS follows a unique collection of differentiation paths (Figures 4G, S5E, and S5J). However, these findings could be confounded by several forms of measurement error, including our selection of one individual tree from many potential candidates that describe our mutational data (tree building) or the fact that each recorded tree represents a subsample of the overall structure (sampling). To address the former issue, we generated 30 randomly seeded trees from each individual TLS and compared how they differ within and across structures (Figure S5K). We observe that path-level metrics are similar within structures but deviate substantially across them,

Figure 4. Estimating the boundaries of progenitor cell decision-making

(A) Ternary plot summarizing individual progenitor node contributions across replicates. Diameter scales with the number of nodes, and position along the plot highlights the fraction of NMP, somitic, or neural cells generated by that node within the tree (schematic). Colors highlight the four major progenitor classes: (1) self-renewing, (2) bipotent, (3) neural, and (4) somitic. Pie chart shows the proportion of progenitor classes across replicates.

(B) Boxplot displaying the fraction of the four progenitor classes within individual replicates. White dot represents a single replicate, and lines connect data from the same structure. Boxes indicate 25th and 75th percentiles, central line the median, and whiskers 10th and 90th percentiles.

(C) Progenitor node distribution ranked by the fraction of extant NMPs. Progenitor nodes are classified as (1) self-renewing, (2) differentiating, or (3) exhausted. Each dot represents an individual progenitor node. Pie charts show the fraction of somitic and neural progeny produced within each class.

(D) Pairwise Aitchison's distance (STAR Methods) boxplot showing cell-state composition differences within clones of the same structure or between structures. Boxes indicate 25th and 75th percentiles, central line the median, and whiskers 10th and 90th percentiles. Outliers are indicated.

(E) Schematic of somitic (T::H2B-mCherry) and neural (Sox2::H2B-Venus) dual reporter line and workflow for assessing clonal contributions. TLSs are individually analyzed by FACS to measure the clonal contribution to the neural and somitic lineages from single-seeded reporter cells. 13/86 clones (displayed in F) provided measurable progeny.

(F) Representative FACS plots of neural and somitic cells generated from three engrafted clones, as well as the clone sizes (n of cells in the clone) and contributions from all replicates.

(G) Representative UMAP and ternary plots for Tracer TLS replicates 22 and 25 shown alongside down-sampled ensembles of 30 reconstructed trees. Despite very similar cell-state compositions, these replicates followed very distinct progenitor choices.

(H) Pairwise scatter plot between cell composition (calculated using Aitchison's distance) and path (cumulative node distances in the ternary plot) for five TLS replicates down-sampled to 200 cells. Each dot represents either a within (blue, intra) or across (red, inter) replicate comparison. Lines represent the linear fit.

(I) Dot plot displaying the slopes of either intra (blue) or inter (red) TLS comparisons for each distribution plotted in (H). The larger slope for across-structure comparisons indicates a larger change in path level distance for the same change in composition. The line indicates the mean.

See also Figure S5.

suggesting that our molecular recording data build consistent trees (Figure S5K).

To quantify the effects of sampling, we down-sampled 5 individual TLSs to the same cell number and reconstructed 30 trees per structure. Down-sampling changes both cell composition and tree topology for each structure, providing the opportunity to investigate the relationship between path (tree) and composition (cell identities) against a null model of within-replicate comparisons. Effectively, this method allows us to ask how much of the observed differences in path (measured by cumulative node distances in the ternary plot) are attributable to compositional differences (calculated using Aitchison's distance on cell-state proportions, STAR Methods). As expected, path and composition are correlated, but across-structure comparisons show a larger difference in path usage for the same change in composition, supporting the general notion that developing systems can accommodate diverse ensembles of individual progenitor cell decisions (Figures 4H and 4I). Globally, these results suggest that individual TLSs proceed via highly divergent progenitor paths that nonetheless converge toward robust developmental outcomes.

Modulating progenitors' renewal and differentiation decision-making

Although individual progenitors appear to follow widely variable paths without compromising the overall phenotype, progenitor cell decision-making must still have boundaries to prevent the emergence of aberrant morphologies. To explore how these boundaries may involve the signaling microenvironment, we examined the 10 Tracer TLS^{CL} replicates recovered after quality control (Figures 3A, S6A, and S6B). The TLS^{CL} condition favors mesodermal and blocks neural differentiation through hyper-activation of WNT and inhibition of BMP signaling, in line with the known roles of these morphogen gradients during axial elongation.[24] Morphologically, this results in structures that lack distinct and organized neural tubes and form supernumerary, hyper-epithelialized somites[24] (Figures 5A, 5B, S6C, and S6D). Our scRNA-seq data confirmed these changes in cell-state composition across our replicates, including an increased fraction of extant NMPs compared to TLS controls (Figures 5B, S6D, and S6E).

Our path-dependent metrics highlight the large degree to which environmental cues can alter progenitor cell decision-making (Figures 5B, 5C, and S6F–S6J). Ternary plots show a consistent and population-wide shift in TLS^{CL} node distribution toward somitic and away from neural and bipotent states, a bias also observed at the level of extant progenitors (Figures 5C–5E, S6F, and S6G). Notably, beyond simply changing the differentiation of NMPs, TLS^{CL} branches are more connected to extant NMPs, with a corresponding increase in the proportion of clonally self-renewing nodes (Figures 5E, 5F, and S6H–S6J). These transitions are consistent across individual replicates, indicating an overall change in the progenitor path distribution that ultimately dictates their altered morphology.

We conclude that, despite having a broad distribution of possible outcomes, the transient coordination of axial progenitors' behaviors is nonetheless controlled by exogenous factors and is extremely sensitive to signaling modulation, which can ultimately distort the elongating trunk toward developmental phenotypes. Moreover, these effects appear to influence the transition between renewal and differentiation rather than change the progenitors' latent potential, a concept we explore in greater detail below.

Progenitor cell differentiation potential changes over time

Embryonic progenitor fields shape complex structures during temporally limited bursts of activity, a fundamentally different property than homeostatic adult stem cell systems. These fields likely change dynamically during morphogenesis, with co-developing cells and tissues exerting new influences on them as they grow.[45] Consistent with this idea, NMP transcriptional state transitions toward the neural portion of our pseudotime axis over the final 24 h of TLS development, a phenomenon we define as "progenitor cell maturation" (Figures 6A and S7A, data taken from Veenvliet et al.[24]).

To test whether these dynamic molecular signatures reflect changes in differentiation potential, we designed an outgrowth assay that evaluates progenitor cell decision-making as they connect to early and late differentiation events (Figure 6B). Specifically, we recorded two sequential rounds of axial elongation by isolating the posterior region of TLS and allowing it to grow for an additional 24 h. Within the integrated trees, cells collected from the first and second generations are deeply connected, both at the clonal level (74.8% ± 4.7%) and throughout the majority of the tree, with shared progenitors localized closer to the root compared with nodes comprised entirely of first- or second-generation cells (Figures 6C–6E, S7B, and S7C). Furthermore, shared progenitors are predominantly multipotent, with bipotent nodes representing the most abundant fraction (44.9% ± 8.9%, Figures 6F and 6G).

Notably, although bipotent nodes connect both generations, we observe a prominent shift in their output toward the neural lineage in the second generation (76.2% ± 11.2%, Figures 6F and 6G), consistent with the transcriptional dynamics of NMPs over time. Furthermore, inter-generational recording of TLSs initially cultured in the CL condition appear to maintain NMP bipotentiality by suppressing this process; despite the biased differentiation toward the somitic fate in the first generation, a high fraction of shared progenitors in TLS^{CL} are still bipotent (41.6% ± 3.4%), indicating that the neural lineage is generated from the same progenitor pool that contributed to the somitic lineage in the first generation (Figures 6H and S7D–S7F). Initial treatment with CL also extends somitic differentiation into the second generation, further supporting a delay in the system's maturation toward a more neural-dominant state (Figures 6H, S7E, and S7F). Experiments conducted using our Sox2/T dual reporter line further support a transition to a neural-biased fate as well as the re-establishment of bipotentiality after CL withdrawal (Figures 6I and S7G–S7K).

Our chemical modulation appears to extend the temporal window of progenitor cell bipotentiality, seemingly by blocking specific differentiation inputs on the NMP pool. To examine the stability of these conditions, we performed a long-term explant experiment where we serially isolated the posterior portion of TLS or TLS^{CL} structures and continuously maintained them in either standard (TLS) or somite-supporting (TLS^{CL}) conditions (Figure 6J). Compared with the control, re-plating explants in TLS^{CL} conditions enabled longer-term axial elongation for up to six generations, including the ability to make properly shaped

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

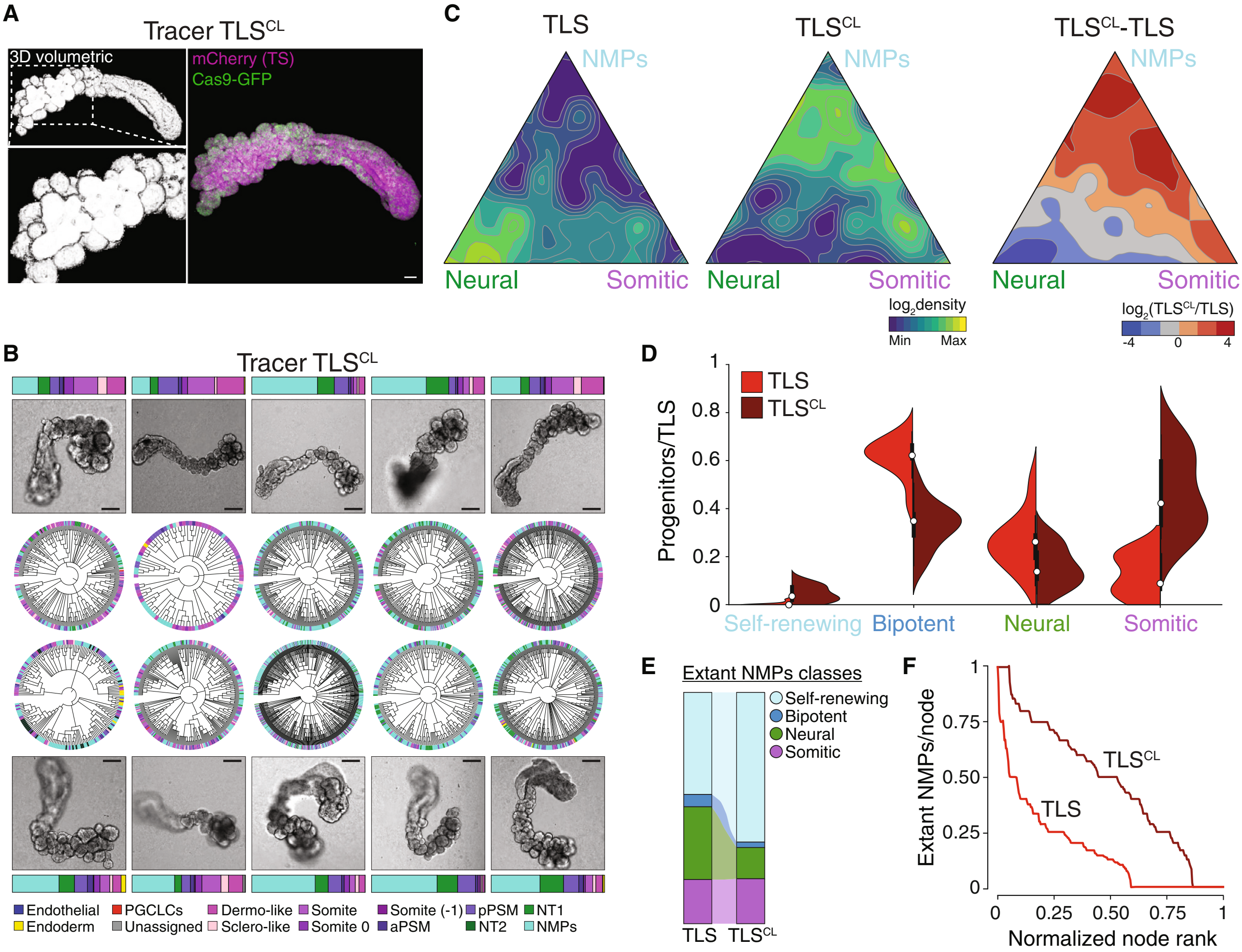


Figure 5. Chemical modulation of progenitor cell dynamics

(A) Volumetric reconstruction and confocal section showing TS-mCherry and Cas9-GFP contributions to a representative Tracer TLS^{CL}. Scale bars, 50 μm.

(B) Bright-field images of 10 Tracer TLS^{CL} replicates alongside their phylogenetic trees. Stacked bar plots summarize the percentage of cells assigned to each cell state. The TLS^{CL} condition reproducibly expands the fraction of somitic cells, from 35% ± 7% in TLSs to 50% ± 22%. Scale bars, 100 μm.

(C) Ternary gradient plots showing the progenitor node distribution in TLS^{CL} as they compare to TLS. Differential analysis identifies a continuous shift toward the somitic and self-renewing fates.

(D) Split violin plot displaying the fraction of each progenitor class within Tracer TLS and TLS^{CL} replicates. Boxes indicate 25th and 75th percentiles, white dots the median, and whiskers 10th and 90th percentiles.

(E) Alluvial plot showing the fraction of extant NMPs scored as self-renewing, bipotent, neural, and somitic in TLS and TLS^{CL} based on the classification as in Figure S4G.

(F) Cumulative distributions for TLS and TLS^{CL} nodes ranked by the fraction of extant NMPs.

See also Figure S6.

somites in the second generation (Figures 6J, red arrows, and 6K). However, all structures eventually converge to a neural-only phenotype, indicating that these conditions are only partially capable of stabilizing a bipotent NMP field (Figures 6J, 6L, and S7L). In sum, our investigation highlights the mechanisms through which transient progenitor fields mature through a period of heightened environmental sensitivity.

DISCUSSION

In this work, we combined stem cell-derived embryonic models with a high-content molecular recorder to evaluate transient progenitor cell dynamics. Specifically, we quantitatively decode macroscopic features of axial elongation and trunk tissue morphogenesis by describing them as a function of individual progenitor cell behaviors. We find that progenitor pools undertake highly distinct individual paths, the ensemble of which converge to robust cellular and morphological phenotypes. By prioritizing cell fate outcomes, we are able to describe the concept of cellular plasticity according to two related parameters—progenitor cell dynamics (self-renewal, differentiation, and exhaustion) and differentiation output (bipotent, neural, and somitic)—as well as recover their general distributions over time. We further apply this framework to determine the specific effects of signaling

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

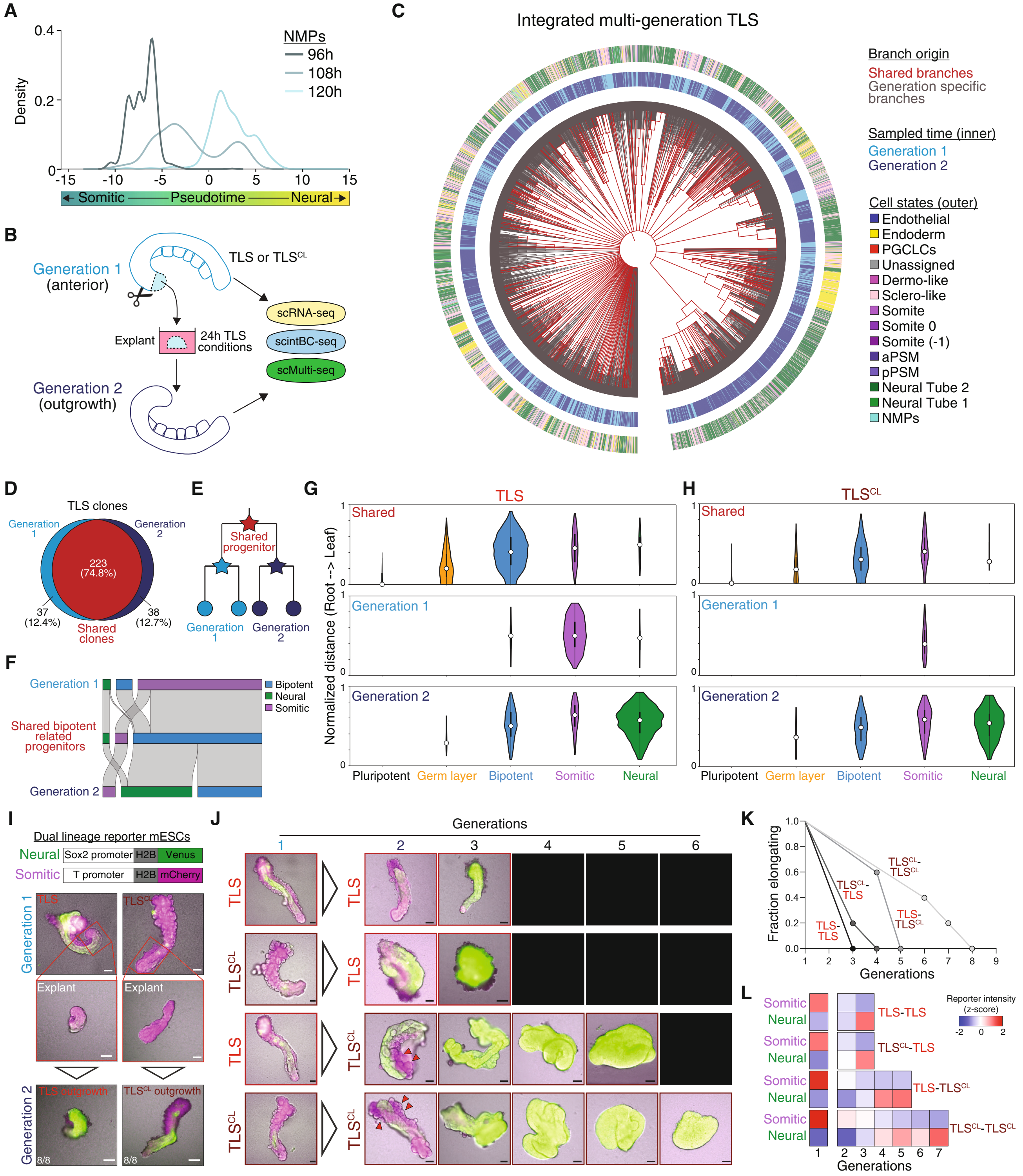


Figure 6. Progenitor cell decision-making changes as a function of time

(A) Density curves showing the distribution of NMPs recovered at 96, 108, and 120 h of TLS formation along the pseudotime axis from Figure 3D.

(B) Experiment to functionally test the differentiation potential of trunk progenitors over two consecutive generations.

(C) Reconstructed lineage tree of a single inter-generational experimental replicate. Dark red branches include cells from both generations. Outer ring highlights the cell state for each leaf, inner ring indicates the generation where the cell was sampled.

(D) Venn diagram showing the fraction of clones that are comprised of cells from both generations.

(legend continued on next page)

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

pathways, demonstrating how environmentally induced phenotypes arise by tuning the rates between self-renewal and differentiation. From a technological perspective, we envision that our Tracer mESCs will serve as a valuable platform for other complex *in vitro* differentiation systems, including the rising number of embryoid models[21,23] and those recently described for studying human development.[46–51] As with many higher order features of cellular differentiation, the ability to recover and quantify progenitor cell behaviors may open new opportunities to connect species-specific dynamics of axial progenitor populations to their underlying genetic and regulatory networks.[52]

Our biological findings have implications for understanding how developmental progression can be finely tuned to produce a complex body plan despite the apparently loose regulation of individual cells, including how developmental robustness might itself be encoded. Recent studies have found evidence of clonal transcriptional and functionally stable heterogeneity during embryonic and adult hematopoiesis, which appears instrumental to maintaining cellular diversification in blood over time.[18–20,53] In contrast to these systems, we find that transient progenitor populations proliferate and differentiate in a manner where the ultimate structure is composed of many diverse paths, suggesting that self-regulation and feedback may act on the population level instead of carefully monitoring individual cell behaviors. These findings would be consistent with detailed efforts to determine the spatial, niche-like morphology of the embryonic trunk, where NMPs appear to be spatially influenced by multiple distinct signaling gradients.[31,36,45] Similarly, our multigenerational results capture the degree to which progenitors change their differentiation output over time (Figure 7). In particular, the apparent transcriptional and functional maturation of NMPs toward the neural lineage would represent a non-homeostatic and distinct feature of this system that imposes an intrinsic and irreversible temporal window. The transcriptional instability of these cells may also explain historical challenges to indefinitely propagate this state in culture.[54] How NMP dynamics evolve to support embryogenesis *in vivo* remains an open question, particularly across species where the duration, number, and transcriptional stability of this field may change to meet the needs of distinct body plans. More broadly, the dynamic tunability of NMP pools may also be reflective of the early gastrulation window's sensitivity to genetic and environmental inputs that can otherwise lead to life-spanning congenital disorders.[55,56]

Although our work highlights the benefit of describing structures *in toto* as the cumulative product of their progenitor dynamics, a future challenge will be to connect the complex gene and cellular regulatory networks (GRNs and CRNs) that govern morphogenetic processes.[57] Compared with the stereotypical development of many invertebrates, our results show substantial macroscopic tolerance to individual cellular decisions, a property that has also been classically observed in the mammalian preimplantation embryo.[58–60] Moreover, comparative investigation of trunk development has demonstrated that a broad range of cellular and morphological phenotypes can stem from even subtle deviations in gene regulation.[61–66] In addition to encoding a more robust and scalable developmental framework, the diversity of body plans that are generated using highly similar genetic pathways may reflect evolutionary mechanisms that specifically act on population-level cellular dynamics.[67] Ultimately, understanding how the timing and coordination of progenitor pools is controlled by cell-intrinsic and -extrinsic inputs will require deeper analytical strategies that incorporate these path-based descriptions. In this manner, high-resolution fate maps may enable deeper investigation into the effects of specific genetic, epigenetic, and environmental variables as they interact to shape complex embryonic structures.

Limitations of the study

Although valuable for their reproducibility and experimental flexibility, TLSs still exhibit more variable morphologies and lack key cell types found in the embryo, including axial, lateral plate and intermediate mesodermal derivatives. Even though these lineages are not the expected progeny of NMPs, they likely function as additional sources of regulation and feedback that are not measured here. Similarly, our explant strategy relies on the enrichment of NMP progenitors in the most posterior portion of a TLS, but does not account for potential migratory events that could influence our interpretation. Our method is therefore currently suitable to study cell populations that are confined within well-defined embryonic niches, as demonstrated for NMPs, but would require additional spatial information to account for these and other aspects of morphogenesis.

Technically, our system is tuned for the rapid dynamics of early development, which requires continuous accumulation of mutational information over defined windows. We selected our Tracer mESC line based on the recorder's minimal effect on cellular toxicity or differentiation potential, but we cannot rule out subtle

---

(E) Schematic of progenitor classification by their relationship to the first and second generations. Progenitors are classified as (1) shared (comprising cells of both generations) or (2) generation-specific (comprising cells from a single generation and stemming from a shared progenitor).

(F) Alluvial plot displaying the connectivity between bipotent nodes when described using all or generation-specific cells (STAR Methods).

(G) Violin plots show the clone-normalized distance for each progenitor class as they are distributed from root to leaf and weighted to reflect their relative abundance within shared or generation-specific portions of the tree. White dot represents the median, boxes the interquartile range, and edges the full range.

(H) Violin plots as in (G) for the TLS^{CL} experiment.

(I) Fluorescent and bright-field images of representative TLS and TLS^{CL} replicates (generation 1) alongside their explants and outgrowths (generation 2). Although TLS^{CL} are predominantly somitic in generation 1, neural differentiation is restored in generation 2. N = 8 structures showed a comparable pattern. Scale bars, 100 μm.

(J) Fluorescent and bright-field images of representative TLS and TLS^{CL} replicates (generation 1) alongside their multi-generation outgrowths. Structures are cultured in the condition indicated after the arrow for the rest of the serial explants. Black windows indicate failed axial elongation. Red arrows indicate well-shaped somites. N = 5 structures showed a comparable pattern. Scale bars, 100 μm.

(K) Survival curves showing the number of structures that preserve axial elongation for each generation and condition. N = 5 structures per condition.

(L) Heatmaps showing the row *Z* score normalized mean fluorescent reporter intensities for the indicated culture conditions over consecutive generations.

See also Figure S7.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

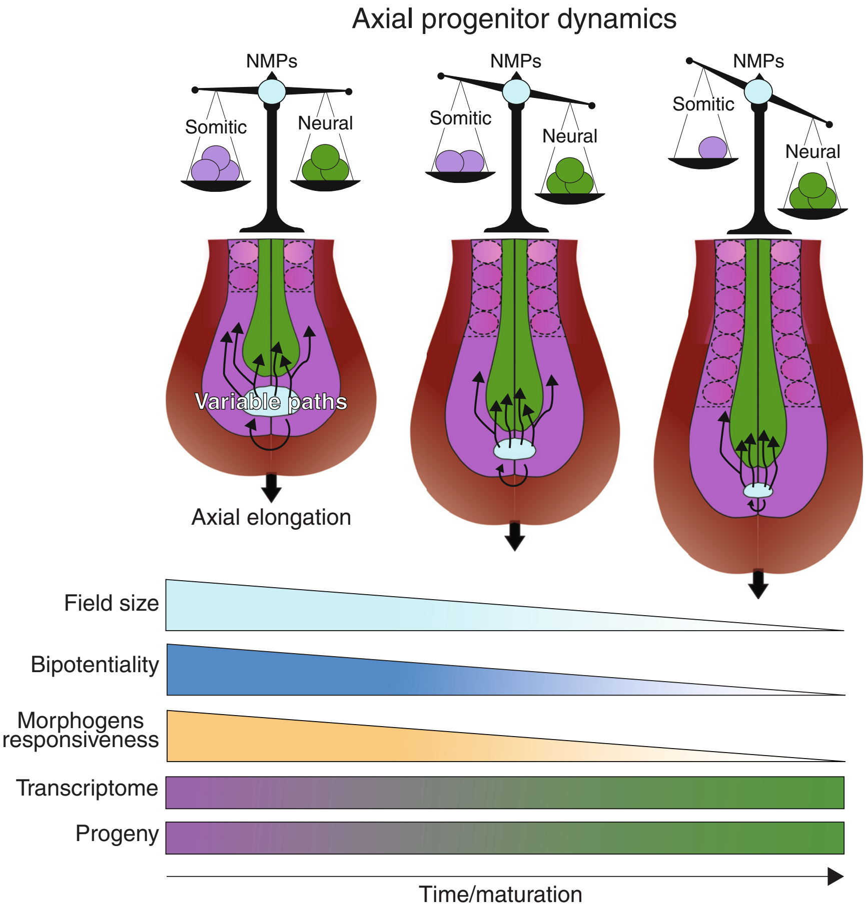


Figure 7. Progenitor field dynamics in TLSs
Schematic model for NMP differentiation dynamics during TLS development. The bipotent field generates both somitic and neural lineages via a diverse assembly of individual progenitor cell decisions (black arrows). Over time, the field shrinks, gradually loses bipotentiality, and matures toward a more neurally committed identity. Progenitor maturation is also accompanied by decreased sensitivity to signaling molecules present within the embryonic niche. Collectively, these dynamics explain the time-dependent nature of developmental windows to promote downstream differentiation and ensure proper tissue morphogenesis.

influence on cellular fitness during differentiation. This and other aspects of molecular recording—including the stochasticity in indel formation, their incomplete recovery, and the complexity in tree reconstruction—remain current limitations for the field. However, our analytical framework and future development of new methods should only improve the ability to reconstruct complex progenitor cell paths.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines
- METHOD DETAILS
  - Rosa26 targeting vector construction
  - PiggyBac library construction
  - LentiBCs lentiviral library generation
  - LentiBCs viral particles preparation
  - Generation of Tracer mESCs
  - Generation of Tracer mESCs containing the LentiBCs
  - Generation of Tracer trunk-like structures
  - TLS processing, reproducibility, and selection
  - Generation of Tracer EBs, gastruloids, and cardiac gastruloids
  - Clonal contribution to TLS
  - Multi-seq Tracer TLS cell preparation
  - Explant experimental procedure
  - Long-term explant experimental procedure
  - FACS analysis of Cas9-GFP activation in Tracer mESCs and sorting
  - FACS analysis of Cas9-GFP activation in individual Tracer TLS
  - Whole-mount immunofluorescence
  - RNA fluorescent *in situ* hybridization
  - Tissue clearing
  - Fluorescent imaging
  - Live cell imaging
  - Image analysis
  - gDNA extraction from Tracer mESCs
  - RNA extraction and cDNA synthesis
  - qPCR- and FACS-based prediction of intBC copy number
  - Amplicon-seq intBCs library preparation
  - Single-cell RNA-seq of Tracer TLS

  - ○ Single-cell RNA-seq of Tracer TLS and Multi-seq BCs recovery
  - ○ IntBCs library preparation from scRNA-seq
  - ○ LentiBCs library preparation from scRNA-seq
  - ○ Multi-seq BCs library preparation from scRNA-seq
  - ○ Explant scRNA-seq experiment
  - ○ Computational analysis
  - ○ Tracer mESCs timecourse barcode analysis
  - ○ scRNA-seq processing
  - ○ Multi-seq processing
  - ○ LentiBCs processing
  - ○ LentiBCs assignment to tree roots
  - ○ Lineage tracing data processing and tree reconstruction
  - ○ Phylogenetic distance
  - ○ Transcriptional correlation
  - ○ Allelic distance
  - ○ Minimum evolutionary coupling
  - ○ scRNA-seq integration
  - ○ Extant NMP classification
  - ○ Differential gene expression
  - ○ Pseudotime
  - ○ Gene module score
  - ○ Extant NMP score
  - ○ Compositional analysis
  - ○ Progenitor classification
  - ○ Shuffled background validation of progenitor classifications
  - ○ Ternary distance
  - ○ Down-sampling TLS replicates
  - ○ Lineage tracing processing and lineage reconstruction of explant experiments
  - ○ Explant subtree analysis
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.devcel.2024.03.024.

ACKNOWLEDGMENTS

We thank members of the Meissner, Chan, and Smith laboratories for critical discussion. We thank Christopher McGinnis and Zev Gartner for providing Multi-seq reagents. We acknowledge the service facilities at the MPIMG and LSI for the continuous support. This work was supported by the Max Planck Society (A.M. and J.V.V.), NIH New Innovator awards (DP2HD108774 [Z.D.S.] and DP2HD111537 [M.M.C.]), the Mathers Foundation (Z.D.S.), the Chen Innovation Award (Z.D.S.), an NIGMS training grant (T32GM007388 [B.K.L.]), and the Damon Runyon Postdoctoral Fellowship (DRG-2238-18 [D.Y.]).

AUTHOR CONTRIBUTIONS

A.B. and A.M. conceptualized the study; A.B. conceived, designed, and performed all experiments with the help of S.I.G. and J.V.V. (Tracer TLS generation), C.R. (LentiBC library preparation), D.Y. (lineage tracing library), and M.W. (Tracer mESC derivation); B.K.L. performed data analysis with help from A.B. and H.K.; A.B. and R.B. performed image analysis; A.B. and J.V.V. developed and optimized the outgrowth assay and co-supervised S.I.G.; J.V.V. gave critical inputs throughout the project; A.M., Z.D.S., and M.M.C. supervised the project; A.B. and Z.D.S. drafted the manuscript; A.B., Z.D.S., and M.M.C. wrote the final manuscript with critical input from all authors.

DECLARATION OF INTERESTS

Z.D.S. and A.M. are co-founders and scientific advisors of Harbinger Health. D.Y. is a co-founder of and scientific advisor for DEM Biopharma.

Received: August 15, 2023
Revised: December 13, 2023
Accepted: March 12, 2024
Published: April 4, 2024

REFERENCES

1. Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119. https://doi.org/10.1016/0012-1606(83)90201-4.
2. Guignard, L., Fiúza, U.M., Leggio, B., Laussu, J., Faure, E., Michelin, G., Biasuz, K., Hufnagel, L., Malandain, G., Godin, C., and Lemaire, P. (2020). Contact area-dependent cell communication and the morphological invariance of ascidian embryogenesis. Science *369*, eaar5663. https://doi.org/10.1126/science.aar5663.
3. Bardot, E.S., and Hadjantonakis, A.K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate. Mech. Dev. *163*, 103617. https://doi.org/10.1016/j.mod.2020.103617.
4. Beumer, J., and Clevers, H. (2021). Cell fate specification and differentiation in the adult mammalian intestine. Nat. Rev. Mol. Cell Biol. *22*, 39–53. https://doi.org/10.1038/s41580-020-0278-0.
5. Goodell, M.A., Nguyen, H., and Shroyer, N. (2015). Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. Nat. Rev. Mol. Cell Biol. *16*, 299–309. https://doi.org/10.1038/nrm3980.
6. Merrell, A.J., and Stanger, B.Z. (2016). Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style. Nat. Rev. Mol. Cell Biol. *17*, 413–425. https://doi.org/10.1038/nrm.2016.24.
7. Domcke, S.S., and Shendure, J. (2023). A reference cell tree will serve science better than a reference cell atlas. Cell *186*, 1103–1114. https://doi.org/10.1016/j.cell.2023.02.016.
8. Fang, W., Bell, C.M., Sapirstein, A., Asami, S., Leeper, K., Zack, D.J., Ji, H., and Kalhor, R. (2022). Quantitative fate mapping: A general framework for analyzing progenitor state dynamics via retrospective lineage barcoding. Cell *185*, 4604–4620.e32. https://doi.org/10.1016/j.cell.2022.10.028.
9. Zechner, C., Nerli, E., and Norden, C. (2020). Stochasticity and determinism in cell fate decisions. Development *147*, dev181495. https://doi.org/10.1242/dev.181495.
10. Bowling, S., Sritharan, D., Osorio, F.G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.C., Fujiwara, Y., Li, B.E., et al. (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. Cell *181*, 1410–1422.e27. https://doi.org/10.1016/j.cell.2020.04.048.
11. Chan, M.M., Smith, Z.D., Grosswendt, S., Kretzmer, H., Norman, T.M., Adamson, B., Jost, M., Quinn, J.J., Yang, D., Jones, M.G., et al. (2019). Molecular recording of mammalian embryogenesis. Nature *570*, 77–82. https://doi.org/10.1038/s41586-019-1184-5.
12. Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., and Church, G.M. (2018). Developmental barcoding of whole mouse via homing CRISPR. Science *361*, eaat9804. https://doi.org/10.1126/science.aat9804.
13. Wagner, D.E., and Klein, A.M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. Nat. Rev. Genet. *21*, 410–427. https://doi.org/10.1038/s41576-020-0223-2.
14. Quinn, J.J., Jones, M.G., Okimoto, R.A., Nanjo, S., Chan, M.M., Yosef, N., Bivona, T.G., and Weissman, J.S. (2021). Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. Science *371*, eabc1944. https://doi.org/10.1126/science.abc1944.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

15. Simeonov, K.P., Byrns, C.N., Clark, M.L., Norgard, R.J., Martin, B., Stanger, B.Z., Shendure, J., McKenna, A., and Lengner, C.J. (2021). Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell *39*, 1150–1162.e9. https://doi.org/10.1016/j.ccell.2021.05.005.
16. Yang, D., Jones, M.G., Naranjo, S., Rideout, W.M., 3rd, Min, K.H.J., Ho, R., Wu, W., Replogle, J.M., Page, J.L., Quinn, J.J., et al. (2022). Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. Cell *185*, 1905–1923.e25. https://doi.org/10.1016/j.cell.2022.04.015.
17. Zhang, W., Bado, I.L., Hu, J., Wan, Y.W., Wu, L., Wang, H., Gao, Y., Jeong, H.H., Xu, Z., Hao, X., et al. (2021). The bone microenvironment invigorates metastatic seeds for further dissemination. Cell *184*, 2471–2486.e20. https://doi.org/10.1016/j.cell.2021.03.011.
18. Rodriguez-Fraticelli, A.E., Weinreb, C., Wang, S.W., Migueles, R.P., Jankovic, M., Usart, M., Klein, A.M., Lowell, S., and Camargo, F.D. (2020). Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature *583*, 585–589. https://doi.org/10.1038/s41586-020-2503-6.
19. Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native haematopoiesis. Nature *553*, 212–216. https://doi.org/10.1038/nature25168.
20. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, eaaw3381. https://doi.org/10.1126/science.aaw3381.
21. Shahbazi, M.N., Siggia, E.D., and Zernicka-Goetz, M. (2019). Self-organization of stem cells into embryos: A window on early mammalian development. Science *364*, 948–951. https://doi.org/10.1126/science.aax0164.
22. Shahbazi, M.N., and Zernicka-Goetz, M. (2018). Deconstructing and reconstructing the mouse and human early embryo. Nat. Cell Biol. *20*, 878–887. https://doi.org/10.1038/s41556-018-0144-x.
23. Veenvliet, J.V., Lenne, P.F., Turner, D.A., Nachman, I., and Trivedi, V. (2021). Sculpting with stem cells: how models of embryo development take shape. Development *148*, dev192914. https://doi.org/10.1242/dev.192914.
24. Veenvliet, J.V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Guignard, L., Kumar, A.S., Pustet, M., et al. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. Science *370*, eaba4937. https://doi.org/10.1126/science.aba4937.
25. Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. Development *142*, 2864–2875. https://doi.org/10.1242/dev.119768.
26. Javali, A., Misra, A., Leonavicius, K., Acharyya, D., Vyas, B., and Sambasivan, R. (2017). Co-expression of Tbx6 and Sox2 identifies a novel transient neuromesoderm progenitor cell state. Development *144*, 4522–4529. https://doi.org/10.1242/dev.153262.
27. Kimura-Yoshida, C., Nakano, H., Okamura, D., Nakao, K., Yonemura, S., Belo, J.A., Aizawa, S., Matsui, Y., and Matsuo, I. (2005). Canonical Wnt signaling and its antagonist regulate anterior-posterior axis polarization by guiding cell migration in mouse visceral endoderm. Dev. Cell *9*, 639–650. https://doi.org/10.1016/j.devcel.2005.09.011.
28. Martin, B.L., and Kimelman, D. (2012). Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. Dev. Cell *22*, 223–232. https://doi.org/10.1016/j.devcel.2011.11.001.
29. Solovieva, T., Wilson, V., and Stern, C.D. (2022). A niche for axial stem cells - A cellular perspective in amniotes. Dev. Biol. *490*, 13–21. https://doi.org/10.1016/j.ydbio.2022.06.015.
30. Tzouanacou, E., Wegener, A., Wymeersch, F.J., Wilson, V., and Nicolas, J.F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. Dev. Cell *17*, 365–376. https://doi.org/10.1016/j.devcel.2009.08.002.
31. Wymeersch, F.J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F.C., and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. eLife *5*, e10042. https://doi.org/10.7554/eLife.10042.
32. Sambasivan, R., and Steventon, B. (2020). Neuromesodermal Progenitors: A Basis for Robust Axial Patterning in Development and Evolution. Front. Cell Dev. Biol. *8*, 607516. https://doi.org/10.3389/fcell.2020.607516.
33. Forlani, S., Lawson, K.A., and Deschamps, J. (2003). Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. Development *130*, 3807–3819. https://doi.org/10.1242/dev.00573.
34. Lawson, K.A., Meneses, J.J., and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development *113*, 891–911. https://doi.org/10.1242/dev.113.3.891.
35. Gouti, M., Delile, J., Stamataki, D., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2017). A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. Dev. Cell *41*, 243–261.e7. https://doi.org/10.1016/j.devcel.2017.04.002.
36. Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. Dev. Cell *42*, 514–526.e7. https://doi.org/10.1016/j.devcel.2017.07.021.
37. Bolondi, A., Haut, L., Gassaloglu, S.I., Burton, P., Kretzmer, H., Buschow, R., Meissner, A., Herrmann, B.G., and Veenvliet, J.V. (2021). Generation of Mouse Pluripotent Stem Cell-derived Trunk-like Structures: An in vitro Model of Post-implantation Embryogenesis. Bio Protoc. *11*, e4042. https://doi.org/10.21769/BioProtoc.4042.
38. Rossi, G., Broguiere, N., Miyamoto, M., Boni, A., Guiet, R., Girgin, M., Kelly, R.G., Kwon, C., and Lutolf, M.P. (2021). Capturing Cardiogenesis in Gastruloids. Cell Stem Cell *28*, 230–240.e6. https://doi.org/10.1016/j.stem.2020.10.013.
39. Jones, M.G., Khodaverdian, A., Quinn, J.J., Chan, M.M., Hussmann, J.A., Wang, R., Xu, C., Weissman, J.S., and Yosef, N. (2020). Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. Genome Biol. *21*, 92. https://doi.org/10.1186/s13059-020-02000-8.
40. Ginsburg, M., Snow, M.H., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. Development *110*, 521–528. https://doi.org/10.1242/dev.110.2.521.
41. Saitou, M., and Yamaji, M. (2012). Primordial germ cells in mice. Cold Spring Harb. Perspect. Biol. *4*, a008375. https://doi.org/10.1101/cshperspect.a008375.
42. Lagha, M., Brunelli, S., Messina, G., Cumano, A., Kume, T., Relaix, F., and Buckingham, M.E. (2009). Pax3:Foxc2 reciprocal repression in the somite modulates muscular versus vascular cell fate choice in multipotent progenitors. Dev. Cell *17*, 892–899. https://doi.org/10.1016/j.devcel.2009.10.021.
43. Nguyen, P.D., Hollway, G.E., Sonntag, C., Miles, L.B., Hall, T.E., Berger, S., Fernandez, K.J., Gurevich, D.B., Cole, N.J., Alaei, S., et al. (2014). Haematopoietic stem cell induction by somite-derived endothelial cells controlled by meox1. Nature *512*, 314–318. https://doi.org/10.1038/nature13678.
44. McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat. Methods *16*, 619–626. https://doi.org/10.1038/s41592-019-0433-8.
45. Buzi, G., Lander, A.D., and Khammash, M. (2015). Cell lineage branching as a strategy for proliferative control. BMC Biol. *13*, 13. https://doi.org/10.1186/s12915-015-0122-8.
46. Liu, L., Oura, S., Markham, Z., Hamilton, J.N., Skory, R.M., Li, L., Sakurai, M., Wang, L., Pinzon-Arteaga, C.A., Plachta, N., et al. (2023). Modeling post-implantation stages of human development into early organogenesis with stem-cell-derived peri-gastruloids. Cell *186*, 3776–3792.e16. https://doi.org/10.1016/j.cell.2023.07.018.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

47. Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature *582*, 410–415. https://doi.org/10.1038/s41586-020-2383-9.

48. Pedroza, M., Gassaloglu, S.I., Dias, N., Zhong, L., Hou, T.J., Kretzmer, H., Smith, Z.D., and Sozen, B. (2023). Self-patterning of human stem cells into post-implantation lineages. Nature *622*, 574–583. https://doi.org/10.1038/s41586-023-06354-4.

49. Sanaki-Matsumiya, M., Matsuda, M., Gritti, N., Nakaki, F., Sharpe, J., Trivedi, V., and Ebisuya, M. (2022). Periodic formation of epithelial somites from human pluripotent stem cells. Nat. Commun. *13*, 2325. https://doi.org/10.1038/s41467-022-29967-1.

50. Weatherbee, B.A.T., Gantner, C.W., Iwamoto-Stohl, L.K., Daza, R.M., Hamazaki, N., Shendure, J., and Zernicka-Goetz, M. (2023). Pluripotent stem cell-derived model of the post-implantation human embryo. Nature *622*, 584–593. https://doi.org/10.1038/s41586-023-06368-y.

51. Yamanaka, Y., Hamidi, S., Yoshioka-Kobayashi, K., Munira, S., Sunadome, K., Zhang, Y., Kurokawa, Y., Ericsson, R., Mieda, A., Thompson, J.L., et al. (2023). Reconstituting human somitogenesis in vitro. Nature *614*, 509–520. https://doi.org/10.1038/s41586-022-05649-2.

52. Guillot, C., Djeffal, Y., Michaut, A., Rabe, B., and Pourquié, O. (2021). Dynamics of primitive streak regression controls the fate of neuromesodermal progenitors in the chicken embryo. eLife *10*, e64819. https://doi.org/10.7554/eLife.64819.

53. Patel, S.H., Christodoulou, C., Weinreb, C., Yu, Q., da Rocha, E.L., Pepe-Mooney, B.J., Bowling, S., Li, L., Osorio, F.G., Daley, G.Q., and Camargo, F.D. (2022). Lifelong multilineage contribution by embryonic-born blood progenitors. Nature *606*, 747–753. https://doi.org/10.1038/s41586-022-04804-z.

54. Edri, S., Hayward, P., Baillie-Johnson, P., Steventon, B.J., and Martinez Arias, A. (2019). An epiblast stem cell-derived multipotent progenitor population for axial extension. Development *146*, dev168187. https://doi.org/10.1242/dev.168187.

55. Baldacci, S., Gorini, F., Santoro, M., Pierini, A., Minichilli, F., and Bianchi, F. (2018). Environmental and individual exposure and the risk of congenital anomalies: a review of recent epidemiological evidence. Epidemiol. Prev. *42*, 1–34. https://doi.org/10.19191/EP18.3-4.S1.P001.057.

56. Finnell, R.H., Waes, J.G., Eudy, J.D., and Rosenquist, T.H. (2002). Molecular basis of environmentally induced birth defects. Annu. Rev. Pharmacol. Toxicol. *42*, 181–208. https://doi.org/10.1146/annurev.pharmtox.42.083001.110955.

57. Gorfinkiel, N., and Martinez Arias, A. (2021). The cell in the age of the genomic revolution: cell Regulatory Networks. Cells Dev. *168*, 203720. https://doi.org/10.1016/j.cdev.2021.203720.

58. Kono, T., and Tsunoda, Y. (1989). Development of single blastomeres from four- and eight-cell mouse embryos fused into the enucleated half of a two-cell embryo. Gamete Res. *22*, 427–434. https://doi.org/10.1002/mrd.1120220408.

59. Maemura, M., Taketsuru, H., Nakajima, Y., Shao, R., Kakihara, A., Nogami, J., Ohkawa, Y., and Tsukada, Y.I. (2021). Totipotency of mouse zygotes extends to single blastomeres of embryos at the four-cell stage. Sci. Rep. *11*, 11167. https://doi.org/10.1038/s41598-021-90653-1.

60. Van de Velde, H., Cauffman, G., Tournaye, H., Devroey, P., and Liebaers, I. (2008). The four blastomeres of a 4-cell stage human embryo are able to develop individually into blastocysts with inner cell mass and trophectoderm. Hum. Reprod. *23*, 1742–1747. https://doi.org/10.1093/humrep/den190.

61. Aires, R., Jurberg, A.D., Leal, F., Nóvoa, A., Cohn, M.J., and Mallo, M. (2016). Oct4 Is a Key Regulator of Vertebrate Trunk Length Diversity. Dev. Cell *38*, 262–274. https://doi.org/10.1016/j.devcel.2016.06.021.

62. Chapman, D.L., and Papaioannou, V.E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. Nature *391*, 695–697. https://doi.org/10.1038/35624.

63. Concepcion, D., Washkowitz, A.J., DeSantis, A., Ogea, P., Yang, J.I., Douglas, N.C., and Papaioannou, V.E. (2017). Cell lineage of timed cohorts of Tbx6-expressing cells in wild-type and Tbx6 mutant embryos. Biol. Open *6*, 1065–1073. https://doi.org/10.1242/bio.026203.

64. Yanagisawa, K.O., Fujimoto, H., and Urushihara, H. (1981). Effects of the brachyury (T) mutation on morphogenetic movement in the mouse embryo. Dev. Biol. *87*, 242–248. https://doi.org/10.1016/0012-1606(81)90147-0.

65. Yoon, J.K., Moon, R.T., and Wold, B. (2000). The bHLH class protein pMesogenin1 can specify paraxial mesoderm phenotypes. Dev. Biol. *222*, 376–391. https://doi.org/10.1006/dbio.2000.9717.

66. Sampath Kumar, A., Tian, L., Bolondi, A., Hernández, A.A., Stickels, R., Kretzmer, H., Murray, E., Wittler, L., Walther, M., Barakat, G., et al. (2023). Spatiotemporal transcriptomic maps of whole mouse embryos at the onset of organogenesis. Nat. Genet. *55*, 1176–1185. https://doi.org/10.1038/s41588-023-01435-6.

67. Alberch, P., Gould, S.J., Oster, G.F., and Wake, D.B. (1979). Size and Shape in Ontogeny and Phylogeny. Paleobiology *5*, 296–317. https://doi.org/10.1017/S0094837300006588.

68. George, S.H., Gertsenstein, M., Vintersten, K., Korets-Smith, E., Murphy, J., Stevens, M.E., Haigh, J.J., and Nagy, A. (2007). Developmental and adult phenotyping directly from mutant embryonic stem cells. Proc. Natl. Acad. Sci. USA *104*, 4455–4460. https://doi.org/10.1073/pnas.0609277104.

69. Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1882.e21. https://doi.org/10.1016/j.cell.2016.11.048.

70. Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440–455. https://doi.org/10.1016/j.cell.2014.09.014.

71. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

72. Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis. Use R!, 2nd ed. (Springer International Publishing).

73. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

74. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

75. Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V. (2000). Logratio analysis and compositional distance. Math. Geol. *32*, 271–275. https://doi.org/10.1023/A:1007529726302.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mCherry antibody	Abcam	Cat# ab167453; RRID: AB_2571870
Anti-GFP antibody	Abcam	Cat# ab13970; RRID: AB_300798
Brachyury Rabbit mAb	Cell Signaling	Cat# D2Z3J; RRID: AB_2799983
Monoclonal anti Cdx2 antibody	Biogenex	Cat# MU392A-5UC; RRID: AB_2923402
Anti-SOX17 antibody	Abcam	Cat# ab84990; RRID: AB_1861437
Anti-SOX2 antibody	R&D System	Cat# AF2018; RRID: AB_355110
N-Cadherin (D4R1H) XP Rabbit	Cell Signaling	Cat# 13116; RRID: AB_2687616
Anti-SOX1 antibody	R&D System	Cat# AF3369; RRID: AB_2239879
Cardiac Troponin T Monoclonal Antibody	Thermo Scientific	Cat# MA5-12960; RRID: AB_11000742
Purified Rat Anti-Mouse CD31	BD	Cat# 557355; RRID: AB_396660
Chemicals, Peptides, Reagents and Recombinant Proteins		
Qubit™ dsDNA HS- Kit	Thermo Fisher Scientific	Cat# Q32851
LipoD293 DNA In Vitro Transfection Reagent	Signagen	Cat# SL100668
PEG-it virus precipitation solution	SBI	Cat# LV810A-1
Gelatin	Sigma	Cat# G1393
leukemia inhibitory factor	Chemicon	Cat# ESG1107
Tryspin-EDTA (0.05%)	Gibco	Cat# 25300054
Xfect mESC Transfection Reagent	Takara	Cat# 631320
Accutase	Sigma	Cat# A6964
Polybrene infection reagent	Merck Millipore	Cat# TR-1003-G
PBS containing $MgCl_2$ and $CaCl_2$	Sigma	Cat# D8662
NDiff227	Takara	Cat# Y40002
Trypanblue	Bio-Rad	Cat# 1450021
Amaxa 4D-Nucleofector X Kit for mESCs	Lonza	Cat# V4XP-3012
CHIR99021	Merck Millipore	Cat# SML1046
Growth-Factor-Reduced Matrigel	Corning	Cat# 356231
bFGF	Thermo Fisher Scientific	Cat# PHG0261
VEGF	R&D	Cat# 293-VE-050/CF
L-ascorbic acid phosphate	Sigma	Cat# A8960
TrypLE Express	Gibco	Cat# 12604021
BSA	Sigma	Cat# A8412
DAPI	Roche Diagnostics	Cat# 10236276001
proteinase K	Thermo Fisher Scientific	Cat# AM2546
Histodenz	Sigma	Cat# D2158
phenol:chloroform	Thermo Fisher Scientific	Cat# 15593031
Glycogen	Thermo Fisher Scientific	Cat# R0561
RNeasy Plus Micro Kit	Qiagen	Cat# 74034
RevertAid First Strand cDNA Synthesis Kit	Thermo Fisher Scientific	Cat# K1622
Q5 2X MasterMix	New England Biolabs	Cat# M0492S
AMPure XP beads	Beckman	Cat# A63881
KAPA HiFi HotStart ReadyMix	Roche Diagnostics	Cat# *KK2601*
Deposited data		
Sequencing data	GEO: GSE220949	https://github.com/ReneBuschow/AB_2022; https://github.com/BenjaminKLaw/TLS_Project; https://zenodo.org/doi/10.5281/zenodo.10791342

(*Continued on next page*)

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: cell lines		
F1G4 mESCs WT	George et al.[68]	N/A
T::H2B-mCherry/Sox2::H2B-Venus reporter mESCs	Veenvliet et al.[24]	N/A
Tracer mESCs	This study	N/A
Oligonucleotides		
See Table S1		
Recombinant DNA/Cloning		
Rosa26-LSL-Cas9-GFP-gRNAs	This study	N/A
Lineage tracing piggyBac library	This study	N/A
LentiBC library	This study	N/A
PacI	New England Biolabs	Cat# R0547S
XbaI	New England Biolabs	Cat# R0145S
NotI	New England Biolabs	Cat# R0189S
BsmbI	New England Biolabs	Cat# R0739S
ClaI	New England Biolabs	Cat# R0197S
MegaX DH10B competent cells	Thermo Fisher Scientific	Cat# C6400-03
Software and Algorithms		
Fiji	Open source	https://fiji.sc/
ZEN BLUE	Zeiss	https://www.zeiss.com/microscopy/en/products/software/zeiss-zen.html
R v.3.6.3	R	https://www.npackd.org/p/r/3.6.3
Python 3.9.11	Python	https://www.python.org/downloads/release/python-3911/
Prism 8	Graphpad	https://www.graphpad.com/features
FlowJo	BD	https://www.bdbiosciences.com/en-us/products/software/flowjo-v10-software

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Michelle M. Chan (mmchan@princeton.edu).

Materials availability

The Tracer mESC line should be directly requested to and will be fulfilled by Adriano Bolondi (bolondi@molgen.mpg.de) or Alexander Meissner (meissner@molgen.mpg.de).

Data and code availability

Raw and processed data can be downloaded from GEO under accession number GSE220949. Code used to reproduce the presented analyses is indexed on https://github.com/ReneBuschow/AB_2022 or on https://github.com/BenjaminKLaw/TLS_Project (https://zenodo.org/doi/10.5281/zenodo.10791342). Any code or additional information required to reanalyze the data reported in this work is available from the lead contact upon request. All data are available in the main text or the supplemental information.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

All cell lines used are male and derived from an F1G4 genetic background.[68] Mouse embryonic stem cells (mESCs) were cultured as previously described.[24,37] Briefly, mESCs were routinely maintained on 6cm plates (*Corning, 430166*) gelatinized with 0.1% gelatin (1:20 dilution of 2% gelatin (*Sigma, G1393*) in tissue-culture grade H_2O) and coated with mitotically inactive primary mouse embryo fibroblasts (3-4x10^4cells/cm^2) with standard mESC medium containing 15% FCS and 1000 U/ml leukemia inhibitory factor (*LIF, Chemicon ESG1107*) at 37°C and 5% or 7.5% CO_2. mESCs were split every second day with a dilution suitable to the proliferation velocity (between 1:5 and 1:9). mESC+LIF medium was refreshed daily. For splitting, media was aspirated and cells were washed

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

once with PBS and trypsinized (Tryspin-EDTA (0.05%) (*Gibco, 25300054*)) for 5-10 min at 37°C. Trypsin was neutralized by 3ml mESC+LIF and cells centrifuged for 5 min at 1000 rpm, after which the pellet was resuspended in mESC+LIF. For freezing of mESCs, cell pellets were resuspended in mESC medium with 20% FCS, and mixed in a 1:1 ratio with mESC freezing medium. Cells were frozen down o/n in the –80 °C and transferred to liquid nitrogen the next day.

METHOD DETAILS

Rosa26 targeting vector construction

The Rosa26 targeting vector includes two separate cassettes. The triple-sgRNA cassette that contains three individual sgRNAs (Ade2-1, Bri1-3 and WhtB-3) was constructed using four-way Gibson assembly (NEB), as previously described.[14,69] The 3xsgRNA cassette was then PCR amplified from the vector using the following primers:

oDY234: ctgcaactccagtctttctagccttaatCGTGACCGAGCTTGTCTGC
oDY235: gattcctgcagaattgtttaaacggttaAAAAAAAGCACCCGACTCGGGTG
and introduced by Gibson assembly into a PacI-digested Rosa26 LSL-Cas9 Targeting Vector.[70]
Individual sgRNA oligos:
Ade2-1-Top: TTGGTCCTTTGTACGCCGAAGAAGTTTAAGAGC
Ade2-2- Bottom: TTAGCTCTTAAACTTCTTCGGCGTACAAAGGACCAACAAG
Bri1-3-Top: ATGGTTGGATCATAACGATATCTCGTTTCAGAGC
Bri1-3-Bottom: TTAGCTCTGAAACGAGATATCGTTATGATCCAACCATGTTT
WhtB-3-Top: ATGGCAGAAGCTATTAATTCGCGGGTTTAAGAGC
WhtB-3-Bottom: TTAGCTCTTAAACCCGCGAATTAATAGCTTCTGCCATAAAC

PiggyBac library construction

The PiggyBac transposon vector was modified from DY001 (Yang et al.[16]). Briefly, DY001 was digested with XbaI and NotI to remove the previous triple sgRNA cassette, and then filled in with Gibson assembly using the following gene fragment.

Gene fragment 1: GACTGGATTCCTTTTTTAGGGCCCATTGGTCTAGACGTTAGGCGGCCGCGTTTAAACATGGCTTTTTCCCCG

LentiBCs lentiviral library generation

pLV_EBFP2_nuc plasmid (*Addgene #36085*) was digested using BsmBI and ClaI in order to remove the CMV promoter. The linearized backbone was then purified and a full length EF1α promoter was cloned using NEBuilder HiFi DNA Assembly Master Mix according to the manufacturer's instructions. Bacterial transformation followed by Sanger sequencing verified the successful cloning. Next, the LentiBCs library was cloned as previously described[20] with some modifications. Briefly, EF1α-modified pLV_EBFP2_nuc plasmid was first linearized using KpnI and then purified. A single strand LentiBCs library cassette (IDT) with the following sequence

LentiBC fragment:

5'CCTCCCCGCCTGGAATTCGAGCTCGGTACGACCTCCCTAGCAAACTGGGGCACAAGAT**NNNN**CT**NNNN**AC**NNNN**TC**NNNN**GT**NNNN**TG**NNNN**CA**NNNN**CGTACCTTTAAGACCAATGACTTACAAGGCA3'

was next cloned into the linearized backbone using two independent NEBuilder HiFi DNA Assembly Master Mix reactions (0.05 pmols linearized plasmid + 2 pmols single strand LentiBCs library each). The two reactions were then pooled and plasmid DNA purified using Zymo column DNA clean and concentrator kit into 10μl TE buffer (10mM TRIS pH7.5/1mM EDTA). Next, MegaX DH10B competent cells (*Thermo, C6400-03*) were used for electroporation with Gene Pulser electroporator (2.0kV, 200 Ohm, 25uF). Four independent electroporation reactions were performed using 1μl purified assembly material each followed by bacterial recovery for 1 h at 37°C in SOC medium. Half of each recovered culture was added to 400ml of ampicillin-enriched culture for a total of 4 flasks. Bacteria were allowed to grow for 6 h, after which cultures were pooled and bacteria used for Maxi Prep to extract LentiBCs plasmid library. The obtained library was quantified using Qubit™ dsDNA HS- Kit (*Thermo, Q32851*) and used for viral particles preparation.

LentiBCs viral particles preparation

For viral particles preparation, 4.5×10^6 HEK293T cells were seeded in each 10cm dish one day prior to transfection. The following day, each dish was transfected using 50μl LipoD293 DNA In Vitro Transfection Reagent (*Signagen, SL100668*) and a roughly equimolar ratio of Pax2:VSVG:LentiBC plasmid library (1:1:1 / 7μg:3.5μg:7μg). The next day, transfection reagent-containing media was discarded and replaced with 10ml mESC+LIF medium. The following two days, viral supernatant was filtered through a 45μm strainer and collected at 4°C. After the second harvesting day, viral supernatant was supplemented with PEG-it virus precipitation solution (*SBI, LV810A-1*) (5ml for each 20ml media) for 24 h at 4°C. Viral particles were finally precipitated by centrifugation at 3234g for 1 h at 4°C. Viral precipitates were resuspended in 500μl mESC+2xLIF medium and used directly for Tracer mESCs transduction (see section below). The entire lentivirus preparation and storage was carried out under S2-safety conditions and pre-cautions.

Generation of Tracer mESCs

F1G4 mESCs were thawed and cultured in Serum/LIF condition for two passages before transfection (see "Cell lines" section). At the second split, cells were transfected using Xfect mESC Transfection Reagent (*Takara, 631320*) following the manufacturer instruction. Briefly, 2×10^5 MEF-depleted mESCs were plated in a 0.2% gelatin coated well of a 12-well plate 5 h prior to transfection. After 5 h, cells were transfected with 3μg of plasmid DNA (2μg Rosa26-lox-STOP-lox-Cas9-T2A-GFP-gRNAs donor/1μg pX459-Rosa26 gRNA expressing vector) in order to knock in the construct in the *Rosa26* locus. MEFs were plated on top of the transfected mESCs 3 h post-transfection. After 48 h, culture was supplemented with Neomycin (G-418, 200μg/ml) for up to a week to select for successfully integrated clones. Monogenic subclones were manually picked, propagated and genotyped for correct integration. A single heterozygous clone was selected and expanded for further intBCs integration.

For intBCs integration, the same transfection protocol was used with the following modifications. Briefly, 2x10^5 MEF-depleted Rosa26-lox-STOP-lox-Cas9-T2A-GFP-gRNAs mESCs were plated in a 0.2% gelatin coated well of a 12-well plate 5 h prior transfection. After 5 h, cells were transfected with 3μg of plasmid DNA (2μg PiggyBac EF1α-mCherry-intBC-TS1-TS2-TS3 transposon/1μg Transposase-BFP) following the manufacturer instructions. MEFs were plated on top of the transfected mESCs 3 h post-transfection. After 48 h, cells were sorted to enrich for highly mCherry$^+$ cells and plated back as single cells at low density on MEFs to allow for colony picking. Monogenic subclones displaying homogeneous mCherry signal were manually picked and mCherry intensity verified using FACS after one week of expansion. Copy number integrations for the selected clone was then verified by qPCR (see "qPCR and FACS-based prediction of intBC copy number" section).

Generation of Tracer mESCs containing the LentiBCs

Tracer mESCs were dissociated with Accutase (*Sigma-Aldrich, A6964*) for 4 min at 37°C, following by MEF depletion on a gelatin-coated 10cm dish for 45 min at 37°C. In the meantime, a full 12-well plate was coated with 0.2% gelatin solution for 20 min at RT, and wells were then washed twice with 1xPBS. MEF-depleted Tracer mESCs were then collected, counted and seeded at a density of 1x10^5 per well in 500μl mESC+2xLIF medium. Cells were left undisturbed for 6 h in the incubator at 37°C. Next, culture medium was replaced by viral supernatant (see section LentiBCs viral particles preparation) supplemented with 8μg/ml Polybrene infection reagent (*MerckMillipore, TR-1003-G*). Transduction was carried on overnight (viral supernatant from one full 10cm dish were used for each well). The following day, 3x10^5 MEFs were seeded in each well on top of the transduced Tracer mESCs. Successfully transduced cells were FACS purified using FACS Aria Fusion (*Becton Dickinson*) enriching for BFP$^+$ cells four days after transduction. Sorted cells were next allowed to grow for one passage and frozen in stocks for subsequent experiments.

Generation of Tracer trunk-like structures

Trunk-like structures (TLS) were generated as previously described[24,37] with some modifications to activate Cas9-based molecular recording. Briefly, mESCs were trypsinized on the feeder plate as described above and washed with mESC+LIF. Cells were pelleted by centrifugation at 1000rpm for 5 min at room temperature and resuspended in 2ml mESC+LIF. On three gelatinized (0.1% gelatin) wells of a 6-well plate, cells were sequentially plated for 25 min, 20 min and 15 min during which cells were kept in the incubator at 37°C and 5% or 7.5% CO_2. With each transfer, cells were triturated to maintain a single cell suspension. mESCs were then pelleted by centrifugation for 5 min at 1000rpm and resuspended in 500μl of mESC+LIF medium. 10μl of the cell suspension was mixed with 10μl of Trypanblue (*Bio-Rad 1450021*) for automated cell counting with *Luna Automated Cell Counter*. 1x10^6 alive cells were then used for nucleofection using the Amaxa 4D-Nucleofector X Kit for mESCs (*Lonza, V4XP-3012*) in combination with 10μg of Cre-expressing plasmid DNA (*Addgene #13775*) according to the manufacturer's instruction and the CG-104 program. Nucleofected cells were then washed with 10ml pre-warmed NDiff227 (*Takara, Y40002*), resuspended in 1ml NDiff227 and counted again using the automated cell counting with *Luna Automated Cell Counter*. 500 cells were then plated in a volume of 30 to 40μl NDiff227 into each well of a 96-well round bottom, low attachment plate (*Costar, 7007 ultra-low attachment 96 well plate (7007)*). Cells were then allowed to aggregate for 48 h. After these 48h cells were pulsed with 3μM CHIR99021 (CHIR, *Merck Millipore*) in 150μl NDiff227. After 72h, medium was refreshed by removing 150μl of the old media and adding the same volume of new, pre-incubated (37°C and 5% or 7.5% CO_2) NDiff227. At 96h, structures were embedded in 5% Phenol-Red-Free-Growth-Factor-Reduced Matrigel (MG) (*Corning, 356231*). To this end, fresh NDiff227 medium was pre-incubated for at least 20 min at 37°C and 5% or 7.5% $CO_{2.}$ Pre-incubated medium was then put on ice for 5 min, after which MG was added to achieve a final concentration of 5% in the culture wells. Medium was then put at room temperature for 5 min, during which 150μl of old medium was removed from the aggregates. New medium with MG (150μl) was then added, and the cultures were returned to the incubator and further cultured at 37°C and 5% or 7.5% CO_2. Tracer TLS cultures were allowed to settle for at least 30 min before proceeding to further experimentation. For Tracer TLS treated with CHIR+LDN (Tracer TLS^{CL}), 5μM CHIR together with 600nM LDN was added together with MG embedding and left from 96h to 120h.

TLS processing, reproducibility, and selection

The TLS protocol was developed to reproducibly generate embryoids with a clearly defined neural tube and somites.[24] Under our experimental conditions, TLS elongate and display a T positive posterior domain with 95% efficiency, develop a neural tube domain with an 85% efficiency and show somite segmentation with an efficiency of ∼60%.[24] Importantly, Tracer TLS exhibited similar performances, as detailed in Figure S1. Tracer TLS used in this study were pre-selected before further processing based on their gross morphologies to ensure molecular profiling and fate mapping of correctly developed structures. In particular, Tracer TLS had to be elongated (indicating developmental symmetry breaking along a clear anterior-posterior axis), display a clear neural tube domain, and include at least four segmented structures (somites) at one or both sides of the neural tube domain.

After 120h, a single TLS is expected to comprise on the order of 10,000 cells (based on TLS1 and TLS2 cell count measurements). We recovered 3,577 and 5,122 high quality cells after Cell Ranger and quality control for Tracer TLS1 and 2, respectively. Of these, 1,836 (51.3%) and 2,453 (47.8%) cells had at least one edit and were sufficient for phylogenetic tree reconstruction for TLS1 and

TLS2, respectively. Based on the estimated size of a TLS, we therefore expect our lineage data to cover roughly 20-25% of each structure, with major sources of cell loss occurring at the experimental processing, quality control, and filtering stages.

Generation of Tracer EBs, gastruloids, and cardiac gastruloids

Tracer EBs, gastruloids and cardiac gastruloids were generated as previously described[24,38] with some modifications.

Tracer EBs

500 Cre-activated Tracer mESCs were plated in a volume of 30 to 40μl NDiff227 into each well of a 96-well round bottom, low attachment plate. Cells were then allowed to aggregate for 48h. After these 48h, 150μl NDiff227 were added to each well. From 72 to 120h, medium was refreshed by removing 150μl of the old media and adding the same volume of new, pre-incubated (37°C and 5% or 7.5% CO_2) NDiff227. Tracer EBs were analyzed at 120h after aggregation.

Tracer gastruloids

500 Cre-activated Tracer mESCs were plated in a volume of 30 to 40μl NDiff227 into each well of a 96-well round bottom, low attachment plate. Cells were then allowed to aggregate for 48h. After these 48h, cells were pulsed with 3μM CHIR99021 in 150μl NDiff227. From 72 to 120h, medium was refreshed by removing 150μl of the old media and adding the same volume of new, pre-incubated (37°C and 5% or 7.5% CO_2) NDiff227. Tracer gastruloids were analyzed at 120h after aggregation.

Tracer cardiac gastruloids

800 Cre-activated Tracer mESCs were plated in a volume of 30 to 40μl NDiff227 into each well of a 96-well round bottom, low attachment plate. Cells were then allowed to aggregate for 48h. After these 48h, cells were pulsed with 3μM CHIR99021 in 150μl NDiff227. At 72h, medium was refreshed by removing 150μl of the old media and adding the same volume of new, pre-incubated (37°C and 5% or 7.5% CO_2) NDiff227. At 96h, the old media was removed and structures were pulsed with 150μl NDiff227+++ (NDiff227, 30ng/ml bFGF (*Thermo, PHG0261*), 5ng/ml VEGF (*R&D, 293-VE-050/CF*), 0.5mM L-ascorbic acid phosphate (*Sigma, A8960*)). At 120h, medium was refreshed by removing 100μl of the old media and adding the same volume of new, pre-incubated (37°C and 5% or 7.5% CO_2) NDiff227+++, while at 144h the old media was refreshed with normal NDiff227. Tracer cardiac gastruloids were analyzed at 144 and 168h after aggregation.

Clonal contribution to TLS

For the clonal contribution experiment, 300 F1G4 feeder-free WT mESCs were seeded in NDiff227 in each well of a 96-well plate, as done for conventional TLS generation. Next, a single dual reporter cell, T::H2B-mCherry/Sox2::H2B-Venus, was sorted in each well and the plate was centrifuged for 1 min at 300g. Next, TLS were generated as described above and individual structures analyzed for clonal contributions using BD FACSCelesta Flow Cytometer as described below.

Multi-seq Tracer TLS cell preparation

Tracer TLS and Tracer TLS^{CL} were generated as described above but using Tracer mESCs containing LentiBCs. For the Multi-seq experiment, 12 Tracer TLS and Tracer TLS^{CL} were picked with a p200 cut tip and transferred in a new well of a 96-well plate. Next, structures were washed twice with ice cold PBS and trypsinized in 20μl TrypLE Express (Gibco) for 25 min in the incubator at 37°C, pipetting every 5 min to help dissociation. Multi-seq labeling was then performed as previously described.[44] Briefly, single-cell suspensions of each sample were incubated with a unique BC-Lipid modified oligonucleotide "anchor" mix (200nM each final) 5 min on ice. Next, a 200nM "co-anchor" mix was added to each sample and cells were incubated for additional 5 min on ice. The reaction was then quenched by addition of 200μl 1xPBS/1%BSA, and cell suspensions were then washed twice with 1xPBS/1%BSA in the plate. Next, the 24 samples were pooled in a 1.5 ml DNA low Bind tube in 1xPBS/1%BSA and FACS purified to enrich for the GFP^+ cell population in 500μl 1xPBS/0.4%BSA using a FACS Aria Fusion (Becton Dickinson). The recovered cell suspension was then subjected to 10x single-cell RNA-seq using the 10x Genomics Chromium™ Single Cell 3' v3.1 (see section below).

Explant experimental procedure

For the explant experiments, TLS and TLS^{CL} were derived as previously described using our previously published dual reporter cell line, T::H2B-mCherry/Sox2::H2B-Venus, which marks the somitic and neural tube compartments in the developing trunk both *in vitro* and *in vivo*.[24] 350 cells per well were seeded for these experiments. At day 5 TLS and TLS^{CL} were imaged using the Zeiss Celldiscoverer 7 with incubator chamber temperature set at 37°C and CO_2 content at 5%. Next, the most posterior ends from eight TLS and eight TLS^{CL} (~10% of the structure overall) were manually dissected using microdissection forceps under a stereomicroscope, using the T::H2B-mCherry signal to confirm the uniform activity of the reporter within the posterior end. Then, each explant was transferred in a well of an Ibidi 8-well glass-bottom plates (*Ibidi 80827*) using a p10 pipette with minimal transferring volume. The dissection was carried out to include most of the posterior end, paying attention not to include any somitic structure. Pre-incubated NDiff227 medium (see "generation of Tracer trunk-like structures" section) was then put on ice for 5 min, after which MG was added to a concentration of 5%. Medium was then put at room temperature for 5 min. MG-supplemented medium (200μl) was then added in each well containing an explant, and each of the explant was repositioned in the center of each well and imaged using the Zeiss Celldiscoverer 7 with incubator chamber temperature set at 37°C and CO_2 content at 5%. Explant cultures were then returned to the incubator and further cultured at 37°C and 5% or 7.5% CO_2 for 24h. The following day, outgrowths from the explants were imaged and fixed for immunofluorescent staining (see "whole-mount immunofluorescence" section).

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Long-term explant experimental procedure

For the long-term explant experiments, the same procedure described above was used for up to seven consecutive rounds of outgrowth. From the third generation onward, posterior explants were dissected from the structure as described above. We identified the posterior portion by its rounded shape, as the anterior tissue is generally more disorganized and has multiple protrusions. Explants from either TLS or TLS^{CL} were transferred either in TLS or in TLS^{CL} conditions and outgrowths imaged every 24h for expression of the reporters. Structures that did not show axial elongation at each explant step were not transferred further.

FACS analysis of Cas9-GFP activation in Tracer mESCs and sorting

To activate the lineage tracing system, Tracer mESCs were nucleofected as described above (see "generation of Tracer trunk-like structures" section). Next, activated Tracer mESCs were seeded in individual wells of a gelatin-coated 12-well plate and each well was analyzed over the 5-day period to quantify the percentage of Cas9-GFP^+ cells. BD FACSCelesta Flow Cytometer was used for the analysis. For the mESCs timecourse experiment (Figure 1), Cas9-GFP^+ cells were sorted each day along the five days to enrich for active tracing cells and cell pellets were snap frozen at -80°C before proceeding with gDNA/RNA isolation. For the sorting and replating experiment (Figure S1), $3x10^5$ Cas9-GFP^+ activated Tracer mESCs were sorted 24 h post-nucleofection and seeded in three independent wells of a gelatin-coated 12-well plate. Each well was then analyzed to quantify the percentage of Cas9-GFP^+ cells using a BD FACSCelesta Flow Cytometer at day 5 post-nucleofection. Flow cytometry data was analyzed using FlowJoV10.

FACS analysis of Cas9-GFP activation in individual Tracer TLS

Tracer TLS were generated as described above (see "generation of Tracer trunk-like structures" section). At day 5 post aggregation, a full plate of Tracer TLS was washed twice with 200μl ice cold 1xPBS and then each structure was dissociated by incubation with 50μl TrypLE Express (Gibco) for 25 min at 37°C. Next, each trypsinization reaction was quenched by addition of 200μl FACS buffer (PBS, 2%FBS, 0.5mM EDTA), followed by plate centrifugation at 300g for 5 min at 4°C. Each pellet was resuspended in 50μl FACS buffer and the full plate was subjected to FACS analysis using the HTS plate reader in a BD FACSCelesta Flow Cytometer. Percentages of Cas9-GFP^+ cells per Tracer TLS were later analyzed using FlowJoV10.

Whole-mount immunofluorescence

Whole mount immunofluorescence was performed as previously described.[24] Individual Tracer TLS were picked using a p200 pipette with the tip cut-off at the 50μl mark and transferred to either 96-well plates (*Costar, 3596*) or Ibidi 8-well glass-bottom plates (*Ibidi, 80827*). Tracer TLS were washed twice with PBS + $MgCl_2$ and $CaCl_2$ + 0.5% BSA (*Sigma, A8412*), once with PBS, and then fixed in 4% PFA for 1h at 4°C on a rocking platform. Subsequently, Tracer TLS were washed twice in PBS for 5 min, permeabilized by 3 x 20 min incubation in 0.5% Triton-X/PBS (PBST) and blocked in 5% fetal calf serum/PBST (blocking solution) overnight at 4°C. For antibody staining, Tracer TLS were transferred to Ibidi 8-well glass bottom plates. Primary antibody incubation was performed in blocking solution for 48-72h at 4°C, after which Tracer TLS were washed three times with blocking solution and three times with PBST. After the last washing step, Tracer TLS were incubated in blocking solution overnight at 4°C. The next day, secondary antibodies diluted in blocking solution were added, and structures were incubated for 24h at 4°C. Afterwards, Tracer TLS were washed three times with blocking solution and three times with PBST. The last PBST washing step after secondary antibody incubation included DAPI (0.02%, *Roche Diagnostics, 10236276001*). DAPI was incubated overnight and washed once with PBST. All primary and secondary antibodies are listed in Table S2.

RNA fluorescent *in situ* hybridization

RNA–FISH in 96h TLS was performed according to the protocol from Molecular Instruments with some modifications. Briefly, 96h TLS were fixed with 4% paraformaldehyde (PFA) at 4°C for 1 h and washed three times for 10 min each with 1× PBS with 0.1% Tween 20 (PBST) at 4°C. Structures were dehydrated in an increasing concentration series of methanol + PBST washes, for 5 min each wash at 4°C (25% methanol; 50% methanol; 75% methanol; 100% methanol). Structures were stored at −20°C overnight or longer. Next, structures were rehydrated in a decreasing concentration series of methanol + PBST washes, for 5 min each wash at 4°C (100% methanol; 75% methanol; 50% methanol; 25% methanol; 100% PBST). After two washes for 10 min at 4°C in PBST, structures were treated with 10μg/ml proteinase K (*Thermo, AM2546*) for 10 min at room temperature. After two washes with PBST for 15 min each, structures were post-fixed in 4% PFA for 15 min at room temperature and washed three times in PBST for 15 min each step. Structures were then prepared for hybridization by incubating in hybridization buffer at 37°C for 1 h. Probes were resuspended in hybridization buffer at a concentration of 1pM and incubated with structures overnight at 37°C. Structures were washed four times with probe wash buffer for 15 min each wash at 37°C, followed by three washes in 5× SSCT buffer + 0.1% Tween 20. Fluorescent hairpins were prepared as described by the manufacturer at a concentration of 0.06μM each hairpin in amplification buffer. Structures were then incubated in amplification buffer before incubation with hairpin probes overnight at room temperature in the dark. Excess probes were removed by five washes of 15 min each step in 5×SSCT at room temperature in the dark. Nuclei were counterstained by incubation with 2μg/ml DAPI. The buffers and probe sequences used in this study are available at Molecular Instruments and their unique ID can be found in Table S2.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Tissue clearing

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

Fluorescent imaging

Tracer TLS stained with antibodies, RNA probes or carrying fluorescent reporters were imaged with the Zeiss LSM710 (laser-scanning microscope), Zeiss LSM880 (laser-scanning microscope with Airyscan) or Celldiscoverer 7 with appropriate filters for mCherry, Venus, DAPI, Alexa Fluor 488, Alexa Fluor 568, Alexa Fluor 647, and combinations thereof. Post-acquisition image processing was performed using Zen or Fiji.[71]

Live cell imaging

Live cell imaging experiments were performed using the Zeiss imaging platform Celldiscoverer 7, running under ZEN blue v3.1 (*Carl Zeiss, Germany*). Screening experiments in 96-well plates were performed using semi-automated strategies. Structures were imaged using a 5x/NA 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 μm/pixel. Prior to the experiment, all 96 imaged positions were manually controlled in x, y, and z. Typical field of view in xy-sizes were 2.52x2.02mm. To ensure a proper representation of the entire structure, z-stacks recording was performed. Typically, 13 slices with a z distance of 19.56μm were captured, resulting in a full z-depth of 235μm. All images were acquired under full environmental control including 100% humidity, 5% (v/v) CO_2 at 37°C.

Image analysis

The processing and analysis for TLS morphometries were carried out in ZEN blue v3.1 (*Carl Zeiss, Germany*) or ImageJ/Fiji.[71] All used models, macros and snippets are publicly available at https://github.com/ReneBuschow/AB_2022.

TLS morphometrics using ZEN

Before images were analyzed, single z-planes were projected (maximum intensity) from individual scenes using custom macros in ZEN blue v3.1. Next, projected images were classified by either median, rolling-background subtraction plus fixed intensity thresholds or previously trained pixel classifier module (*ZEN Intellesis module*). The image analysis module used these thresholded images or probability maps to assign regions of interest (TLS). For both strategies, the resulting structures were filtered by an area ranging from 25,000-750,000μm^2. Morphometric and gene expression parameters of interest (area, perimeter, max axis, circularity, T::H2B-mCherry mean intensity and Sox2::H2B-Venus mean intensity) were then calculated and stored by an automated routine. The model and the related images analysis scripts can be found at https://github.com/ReneBuschow/AB_2022.

TLS morphometrics using Fiji

With this semi-automated routine, anteroposterior axis (max axis) for each TLS, explant or outgrowth sample are manually drawn, while all other morphometric features are subsequently calculated by the macro. Outlines for TLS structures are generated automatically based on the bright field images. This outline was then used to compute morphometric and gene expression parameters of interest (area, perimeter, max axis, circularity, T::H2B-mCherry mean intensity and Sox2::H2B-Venus mean intensity). Finally, the generated masks were applied to all fluorescent channels and pixels intensity values calculated for each structure. The obtained values were subsequently used for gene expression analysis using custom R scripts. All custom Fiji and R codes can be found at https://github.com/ReneBuschow/AB_2022.

RNA-FISH analysis in TLS

Quantification of single cell RNA levels from confocal images was performed within ZEN blue v3.5 (*Carl Zeiss, Germany*). We analyzed a total of 760 cells by a robust intensity-based method. Herein, we segment in a first step the nuclear counter staining (DAPI) by a gaussian smoothing function, followed by background subtraction and fixed intensity thresholds and water-shedding. Subsequently nuclear regions were dilated by 30 pixels (~4.2μm) as a proxy for cytoplasm. The combination of the resulting nuclear and cytoplasmatic regions was used to measure the mean fluorescent intensity per cell. For the statistical analysis we used custom R scripts, and plotted mean fluorescence intensity values for each transcript. The image analysis scripts can be found at https://github.com/ReneBuschow/AB_2022.

gDNA extraction from Tracer mESCs

gDNA was extracted from isolated Cas9-GFP$^+$ Tracer mESCs at different days after Cas9-GFP activation for downstream intBCs library preparation. Briefly, $1x10^5$-$1x10^6$ cell pellets from the indicated days were collected and lysed in 200μl Lysis buffer (10mM Tris-HCl pH=8.0; 10mM NaCl; 10mM EDTA; 0.5% SDS; 300μg/mL Proteinase K) overnight at 55°C. The following day, gDNA was purified using phenol:chloroform (*Thermo, 15593031*) extraction followed by precipitation overnight at -80°C (2.5V EtOH, 1/10V 5M NaCl and 1/100V molecular biology grade Glycogen (*Thermo, R0561*)), centrifugation and resuspension. gDNA concentrations were measured using Qubit™ dsDNA HS- Kit (*Thermo, Q32851*) and same amount of input material was used for each sample for intBCs library preparation.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

RNA extraction and cDNA synthesis

RNA was extracted from isolated Cas9-GFP$^+$ Tracer mESCs at different days after Cas9-GFP activation or from individual Tracer TLS at day 5 post aggregation for downstream intBCs library preparation. Briefly, RNA was extracted from 1x10^5 isolated Cas9-GFP$^+$ Tracer mESCs or individual Tracer TLS using RNeasy Plus Micro Kit (Qiagen) and RNA concentration and quality was measured using Qubit™ RNA HS (Thermo, *Q32852*). cDNA synthesis was performed using 1μg total RNA (for Tracer mESCs) or all extracted RNA (for individual Tracer TLS) using the RevertAid First Strand cDNA Synthesis Kit (*Thermo, K1622*) following the manufacturer's instructions. OligoT were used as primers first strand cDNA synthesis.

qPCR- and FACS-based prediction of intBC copy number

qPCR on gDNA extracted from a set of fifteen selected Tracer mESCs clones was performed to evaluate the number of EF1α-mCherry integrations. Briefly, gDNA was isolated as described above (see "gDNA extraction from Tracer mESCs" section). Next, qPCR was performed using 25ng gDNA/well and amplifying using the following primer pairs:

mCherry_qPCR_F: CCGACATCCCCGACTACTTG
mCherry_qPCR_R: ATGAACTCGCCGTCCTGC
Tfrc_qPCR_F: CTCAACCAAATGGTTCGTACAG
Tfrc_qPCR_R: ACATCTCATAGTCCAGGTTCAAT

IntBC copy number was then calculated as:

$$\text{n intBCs} = \left[2^{\left(\text{ct}_{\text{mCherry}} - \text{ct}_{\text{Tfrc}}\right)}\right] * 2$$

The number of measured intBCs for each clone was then compared to the mCherry mean fluorescent intensity (MFI) measured by FACS for each clone and this information used to fit a linear function:

*n intBCs = 0.0003*mCherryMFI+2.411*

The obtained function was subsequently used to estimate the number of intBCs in a set of 171 clones analyzed by FACS. A Tracer mESC clone harboring 10 intBCs was finally selected and used in this work.

Amplicon-seq intBCs library preparation

Isolated gDNA and prepared cDNA from Tracer mESCs and individual Tracer TLS were used for Amplicon-seq intBCs library preparation. 200ng gDNA and 100ng cDNA were used as input material. Two subsequent PCR reactions were carried on to prepare the NGS library. In the first step the input material was used for 4x25μl PCR reactions for each sample. Amplification was carried out using Q5 2X MasterMix (*NEB, M0492S*) with the following steps: 98°C 30 s/ 98°C 10 s – 65°C 30 s – 72°C 45 s (12 cycles)/ 72°C 5 min/ 4°C hold. Next, the four reactions for each sample were pooled together, followed by a double-sided AMPure XP beads (*Beckman, A63881*) cleanup (0.55x-0.8x). Purified intBCs amplicons were eluted in 30μl EB buffer.

In the second step, the whole eluate from the previous PCR reaction was used for 4x25μl PCR reactions for each sample. Amplification was carried out using Q5 2X MasterMix (*NEB, M0492S*) with the following steps: 98°C 30 s/ 98°C 10 s – 65°C 30 s – 72°C 45 s (12 cycles)/ 72°C 5 min/ 4°C hold. Next, the four reactions for each sample were pooled together and purified using AMPure XP beads (*Beckman, A63881*) cleanup (0.8x). Quality and concentration of the obtained libraries were measured using Agilent High Sensitivity D5000 ScreenTape on an Agilent 4150 TapeStation. Libraries were then sequenced using paired end sequencing (500 cycles kit; 250/250) on a Miseq platform at a minimum of 50 million fragments per sample. Oligonucleotides sequences used for library preparations are listed in Table S3.

Single-cell RNA-seq of Tracer TLS

Single-cell RNA-sequencing (scRNA-seq) experiments have been performed as previously described.[24] Briefly, Tracer TLS were generated as described above. Two individual 120h TLS (Tracer TLS1 and 2) were selected based on the presence of one axis of elongation, clear formation of a neural tube domain and somite segmentation. Individual 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 deplete excess Matrigel. Each structure was then dissociated in 200μl TrypLE Express (Gibco) for 25 min at 37°C, with pipetting every 5 min intervals. The cell suspension was filtered using Scienceware Flowmi Cell Strainers, 40μm. Cells were washed twice with 1ml 1xPBS/0.4%BSA with centrifugation steps performed for 5 min at 1200rpm in 1.5 ml DNA-lobind Eppendorf tubes. The cell concentration was determined using a hemocytometer and cells were subjected to scRNA-seq (10x Genomics, Chromium™ Single Cell 3' v3.1; one structure per reaction). Single-cell libraries were generated according to the manual, with one modification: fewer PCR cycles (n=8) were run than recommended during cDNA amplification or library generation/sample indexing to increase library complexity. Quality and concentration of the obtained libraries were measured using Agilent High Sensitivity D5000 ScreenTape on an Agilent 4150 TapeStation. Libraries were sequenced with a minimum of 400 million paired end fragments according to parameters described in the manual.

Single-cell RNA-seq of Tracer TLS and Multi-seq BCs recovery

The steps described in the previous section were applied to generate the scRNA-seq library of the Multi-seq sample, with two modifications: (i) during the cDNA amplification step, 1μl of an oligonucleotide to enrich for the Multi-seq BCs was added to the reaction (see Table S3 for oligo sequence) and (ii) after cDNA amplification and incubation with SPRIselect beads, the Multi-seq BCs containing supernatant was collected and subjected to further SPRIselect beads incubation in order to recover the Multi-seq BCs as previously described.[44] Multi-seq BCs recovery and integrity were measured using Agilent High Sensitivity D5000 ScreenTape on an Agilent 4150 TapeStation. The obtained material was then used as input for Multi-seq BCs library preparation (see section below).

IntBCs library preparation from scRNA-seq

Lineage tracing barcodes (intBCs) libraries were prepared for all samples via PCR amplification from the double stranded 10x cDNA. Briefly, 100ng of 10x cDNA was used to prepare 4x25μl PCR reactions for each sample. Amplification was carried out using Q5 2X MasterMix (*NEB, M0492S*) with the following steps: 98°C 30 s/ 98°C 10 s – 68°C 30 s – 72°C 40 s (10 cycles)/ 72°C 2 min/ 4°C hold. Next, the four per-sample reactions were pooled together and purified using double-sided AMPure XP beads (*Beckman, A63881*) cleanup (0.45X/0.7X). Quality and concentration of the obtained intBC libraries were measured using Agilent High Sensitivity D5000 ScreenTape on an Agilent 4150 TapeStation. Libraries were then sequenced using asymmetric end sequencing (150 cycles kit; 28/91 FC-410-1002) on a Novaseq platform at a minimum of 100 million fragments per sample. Oligonucleotides sequences used for library preparations are listed in Table S3.

LentiBCs library preparation from scRNA-seq

LentiBCs libraries were prepared via PCR amplification from the double stranded 10x cDNA. Briefly, 100ng of 10x cDNA was used to prepare 4x25μl PCR reactions for each sample. Amplification was carried out using Q5 2X MasterMix (*NEB, M0492S*) with the following steps: 98°C 30 s/ 98°C 10 s – 68°C 30 s – 72°C 40 s (10 cycles)/ 72°C 2 min/ 4°C hold. Next, the four reactions for each sample were pooled together, followed by a double-sided AMPure XP beads (*Beckman, A63881*) cleanup (0.45X/0.7X) to purify the intBCs libraries. Quality and concentration of the obtained libraries were measured using Agilent High Sensitivity D5000 ScreenTape on an Agilent 4150 TapeStation. Libraries were then sequenced using asymmetric end sequencing (150 cycles kit; 28/91 FC-410-1002) on a Novaseq platform to obtain 100 million fragments. Oligonucleotides sequences used for library preparations are listed in Table S3.

Multi-seq BCs library preparation from scRNA-seq

Multi-seq BCs libraries were prepared as previously described.[44] Briefly, 10ng input material obtained from the 10x cDNA purification (see section above) was used to perform library PCR using KAPA HiFi HotStart ReadyMix (*Roche, KK2601*) in 50μl reaction with the following steps: 95°C 5 min/ 98°C 15 s – 60°C 30 s – 72°C 30 s (13 cycles)/ 72°C 1 min/ 4°C hold. Next, AMPure XP beads (*Beckman, A63881*) cleanup (1.6X) was performed to purify the Multi-seq BC libraries. Quality and concentration of the obtained libraries were measured using Agilent High Sensitivity D5000 ScreenTape on an Agilent 4150 TapeStation. Libraries were then sequenced using asymmetric end sequencing (150 cycles kit; 28/91 FC-410-1002) on a Novaseq platform at a minimum of 50 million fragments per sample. Oligonucleotides sequences used for library preparations are listed in Table S3.

Explant scRNA-seq experiment

For the explant scRNA-seq experiment, the same procedure described in the previous sections was applied to microdissected TLS and TLS^{CL} replicates with the only difference that 500 cells per well were used as starting material (as for Tracer TLS in general). Three TLS and three TLS^{CL} structures were selected at 120 h and the anterior portion of each structure (generation 1) was used for Multi-seq based scRNA-seq as described in the corresponding section (one 10x reaction). Outgrowths of all six structures (generation 2) were subjected to the same exact procedure 24 h later (two 10x reactions). Single-cell libraries for cDNA, lineage tracing and Multi-seq barcodes were generated and sequenced as described above.

Computational analysis

All analyses were carried out using R v.3.6.3 and Python (3.9.11) unless stated otherwise.

Tracer mESCs timecourse barcode analysis

intBC amplicon data (derived from either gDNA or cDNA) were processed using the reference and algorithms from the Cassiopeia tool.[39] First, the reads were aligned to the reference (“PCT48”) using the Smith-Waterman algorithm with parameters -gapopen 20 and -gapextend 1. Next, the intBC cut site information for each individual barcode was determined using the callAlleles-PCT48.pl script. intBC patterns were classified based on the code from the indel table into Deletion & Insertion (D and I), Deletion (D), Insertion (I) or uncut. Barcodes that did appear in the known barcode list were assigned to the closest barcode if the edit distance was less than 4 and otherwise removed. The number and type of edit and position was summarized by barcode and time-point and visualized using the ggplot package.[72]

scRNA-seq processing

Raw reads (fastq) were generated using Cell Ranger (https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest) (v.4) from 10x Genomics Inc. with the command “cellranger mkfastq.” Reads from all timepoints were aligned against the mouse genome (mm10 with additional pseudogenes input for the GFP and mCherry transcripts), and barcodes and unique molecular identifiers were counted using “cellranger count”.

Gene-barcode matrices were loaded into R and converted into a Seurat object.[73] Cells were filtered for quality (nFeature_RNA > 3000 & nCount_RNA > 10000 & nCount_RNA < 80000 & percent.mt < 5 for TLS1, TLS2, and the TLS Multiseq; nFeature_RNA > 1500 & nCount_RNA > 5000 & nCount_RNA < 80000 & percent.mt < 5 for the Tracer Explant experiment). Next, the data were normalized (NormalizeData) and cell types were predicted by comparison to our own previously published TLS time-course[24] (FindVariableFeatures with selection.method = "vst", nfeatures = 2000; FindTransferAnchors with dims = 1:30; TransferData with dims = 1:30).

For visualization, a UMAP was calculated for an integrated dataset of the respective lineage tracer objects and the published TLS 96h, 108h and 120h data sets by (1) selection of integration features on the list of objects (SelectIntegrationFeatures), (2) scaling of all objects and regression of cell cycle as well as mito count (CellCycleScoring and ScaleData with vars.to.regress = c(“S.Score”, “G2M.Score”, “percent.mt”)), (3) running a PCA (RunPCA), (4) finding integration anchors (FindIntegrationAnchors with dims = 1:30 and union of these with the cell cycle genes), (5) integration of the objects (IntegrateData with dims = 1:30) and finally (6) scaling of the integrated object, PCA and UMAP calculation (RunUMAP with dims = 1:30, n.neighbors=10).

Multi-seq processing

Multi-Seq data were processed in R using the “deMULTIplex" package (https://github.com/chris-mcginnis-ucsf/MULTI-seq).[44] First, the barcode list from scRNAseq cells matching our quality criteria (see scRNA-seq processing) were loaded to R as cell.id.vec. Next, the Multi-seq sample barcode FASTQs were preprocessed using MULTIseq.preProcess with parameters cellIDs = cell.id.vec, cell=c(1,16), umi=c(17,28), tag=c(1,8). Then, the Multi-seq sample barcode alignment was done using MULTI-seq.align and three rounds of quantile sweeps for sample classification were done following the tutorial.

LentiBCs processing

Raw reads (fastq) were generated using Cell Ranger (v.4) from 10x Genomics Inc. with the command “cellranger mkfastq.” Reads from all timepoints were aligned against the mouse genome (mm10 with additional genes for GFP and mCherry), and barcodes and unique molecular identifiers were counted using “cellranger count.” The unaligned reads were extracted from the BAM file using samtools (-bh -f 4) and tranformed into a fastq file (bedtools bamtofastq). The known pre- and suffix (ACCTCCCTAGCAAACTGGGGCACAAGAT and GTACCTTTAAGACCAATGACTT) were trimmed away using cutadapt and only resulting reads with the correct pattern (....CT....AC....TC....GT....TG....CA....C) and length (n=41) were kept for further processing (LentiBCs). Based on the read name, the LentiBCs and cell BC joined, resulting in a table containing read name, cell BC and LentiBCs.

LentiBCs assignment to tree roots

To calculate the proportion of the dominant lentiviral barcode for each clone in a tree, all leaves, i.e. cells, with a missing lentiviral barcode was excluded from the data set. Of the remaining cells, the proportion of the top lentiviral barcode was calculated as the highest cell count over the sum. For each tree which is made of several clones, the median top proportion is calculated to represent the sample, and then the median of the sample medians is reported.

Lineage tracing data processing and tree reconstruction

The lineage tracing libraries for the TLS1 and the Multi-seq experiments were sequenced on the same flow cell. However, indices could not be fully determined, and reads from the two libraries were merged. A stringent filtering strategy based on the cell-barcode information was therefore applied to recover the data. First, cell barcodes discovered in the associated scRNA-seq dataset were extracted. Next, unique cell barcodes for each of the two datasets were determined and the lineage tracing reads were filtered down to include only those matching cells with experiment-specific cell barcodes. These extremely stringent criteria did not only require lineage tracing reads to match their respective scRNA-seq dataset, but also to be unique for that experiment. Trees were then reconstructed with Cassiopeia[39] using the “Cassiopeia-ILP” or “Cassiopeia-Hybrid” algorithms. Briefly, Cassiopeia-ILP infers phylogenies using a near-optimal Steiner-Tree maximum parsimony approach. For larger trees, Cassiopeia-Hybrid splits cells into clades using a “Greedy” heuristic based on the frequency and probability of Cas9 mutations that are likely to have occurred at the beginning of the experiment. Then, each clade is reconstructed using the Cassiopeia-ILP approach.

Given the size and complexity of our datasets, we found that the Cassiopeia-ILP and -Hybrid approaches were unable to converge within a reasonable timeframe. To maintain consistency, we applied a convergence time limit of 1 h to the reconstruction process, as we observed minimal decreases in tree likelihood from extended convergence times. In principle, the ILP algorithm creates an exact tree. However, missing values in the dataset, and the inability to generate a complete Steiner Tree adds variability to the algorithm, even for small trees. In this case, Cassiopeia utilizes a “Random Seed” to set the solution path of the tree optimizer. To investigate the effect of the Random Seed, we generated 30 trees for each TLS dataset using a range of potential seeds. We established the robustness of our approach by comparing the likelihood of 35 independent trees, including the consistency of specific nodes when the Hybrid and ILP methods were run with distinct (n = 30) random seeds. The optimal random seed for each experiment was selected and used in each reconstruction pipeline. We then defined clones to be the first branches off the root of the fate map.

To graph the resulting TLS trees, we utilized the Cassiopeia “upload-and-export-itol” function to generate iTOL tree representations. We then plotted each tree separately using a built-in circular plotting and metadata functions available in iTOL.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Phylogenetic distance

The phylogenetic distance between two objects on a tree is the sum of branch lengths between them. To compute this distance between cells A and B, we utilized the Ete Toolkit python package to compute the sum of branch distances where each branch was defined by the mutations that separate each node from another (see previous section).

Transcriptional correlation

Pairwise expression correlation was performed using a Pearson correlation and was estimated using the top 10 differentially expressed genes for each cell state cluster.

Allelic distance

To assess the validity of tree reconstruction, we used a modified allelic distance score that calculates the degree of shared indels between cells.[11,14,16] Given cells A and B, the allelic distance compares the distance of the i^{th} character with the following:

$$aD(a_i, b_i) = \begin{cases} 2, \mathit{if}\ a_i \neq b_i\ \mathit{and}\ a_i \neq 0\ \mathit{and}\ b_i \neq 0 \\ 1\ \mathit{if}\ (a_i == 0\ \mathit{or}\ b_i == 0)\ \mathit{and}\ a_i \neq b_i \\ o.w. \end{cases}$$

Then, the distance between cells A and B is calculated as the sum of all $aD(a_i, b_i)$ for all indexes i in A and B. The allelic distance is compared against phylogenetic and transcriptional distances as an orthogonal measure of lineage similarity that does not require tree reconstruction.

Minimum evolutionary coupling

Given cell states X and Y, we first calculate the distances between each pair of cells.

$$\begin{pmatrix} d(X_1, Y_1) & \cdots & d(X_1, Y_y) \\ \vdots & \ddots & \vdots \\ d(X_x, Y_1) & \cdots & d(X_x, Y_y) \end{pmatrix}$$

The goal of minimum evolutionary coupling is to identify the closest ancestral relationship between cells. To account for parallel differentiation trajectories within each tree, we calculate the lineage relationship from X (source) to Y (target) as the minimum distance from each cell in X to the closest cell in Y. Then, the median of this vector is taken to calculate the Minimum Evolutionary Coupling from X to Y:

$$mD(X, Y) = \text{median} \begin{pmatrix} \min\left(\left\{d(X_1, Y_1)\ldots d(X_1, Y_y)\right\}\right) \\ \min\left(\left\{d(X_x, Y_1)\ldots d(X_x, Y_y)\right\}\right) \end{pmatrix}$$

The resulting minimum evolutionary coupling values were then clustered and visualized using the Seaborn “clustermap” function. Note that minimum evolutionary coupling is asymmetric.

scRNA-seq integration

Pre-processed TLS1, TLS2, and TLS Multi-seq replicates were integrated together with the previously published TLS time-course[24] to create a large dataset comprising TLS and TLS^{CL} replicates. Briefly, the integration was carried out by (1) selection of integration features on the list of objects (SelectIntegrationFeatures), (2) scaling of all objects and regression of cell cycle as well as mito count (CellCycleScoring and ScaleData with vars.to.regress = c(“S.Score”, “G2M.Score”, “percent.mt”)), (3) running a PCA (RunPCA), (4) finding integration anchors (FindIntegrationAnchors, reduction = rpca, with dims = 1:30 and union of these with the cell cycle genes), (5) integration of the objects (IntegrateData with dims = 1:30) and finally (6) scaling of the integrated object, PCA and UMAP calculation (RunUMAP with dims = 1:30, n.neighbors=10). Once completed, the integrated object was visualized via UMAP.

Extant NMP classification

To assess the impact of NMP differentiation on transcriptomic signatures, we categorized extant NMPs according to the identity of their sibling cells on the tree, which represents their most recently connected cell. Narrowing our investigation to NMP trajectories, we focused on NMPs that are closely related to Neural, Somitic, or other extant NMP cells, and excluded any cell that neighbors an Endoderm, Endothelial, PGCLCs, or unassigned cells. To account for low sampling of rare cell types, we combined the Neural and Somitic cell types into singular categories. Specifically, NeuralTube1 and NeuralTube2 cells were labeled as “Neural”, while aPSM, pPSM, Somite (-1), Somite 0, Somite, Sclerotome-like, and Dermomyotome-like cells were labeled as “Somitic”.

Next, extant NMPs were classified based on the relative proportions of NMP, Neural, and Somitic sibling cells. First, NMPs that are siblings with exclusively NMPs were categorized as “Self-renewing” to capture the self-renewing progenitors within an experiment. Second, NMPs that neighbor both Neural and Somitic cell states were classified as “Bipotent”. Finally, the remaining NMPs were classified as “Neural” or “Somitic” if they were found to be siblings with only Neural or Somitic cells, respectively.

Differences within the extant NMP categories were subsequently investigated using differential gene expression and pseudotime analysis.

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Differential gene expression

Using the integrated TLS dataset and NMP categories described above, differential gene expression between NMP categories was tested for all genes expressed in the categories of interest. Differential gene expression was calculated using the Wilcoxon rank-sum test offered in the scanpy rank_gene_groups function and the log fold change was calculated as the $\log_2(\text{Group1} + \text{pseudocount}) - \log_2(\text{Group2} + \text{pseudocount})$ with a pseudocount of 0.01.

Pseudotime

Trajectory inference and pseudotime analysis were performed for the integrated TLS dataset using the Monocle 3 package.[74] Since the trajectory identified spanned the differentiation of NMPs into both the Neural and Somitic lineages, we centered the 0 pseudotime value on the median pseudotime value of all NMPs. The remaining pseudotime values were shifted accordingly with the Somitic trajectory extending towards -45 and the Neural trajectory towards +15. The previously calculated UMAP and cell cluster assignment were used for visualizing pseudotime trajectories.

Gene module score

To characterize gene sets involved in NMP differentiation, the Monocle 3 package[74] was used to identify modules that co-vary along the pseudotime axis. First, genes that vary significantly across the full pseudotime axis were identified (graph_test, neighbor_graph = 4) using the Morans I spatial autocorrelation analysis. Given the complexity of the full pseudotime axis, a higher number of the genes were found to be significant (22,292, $q\_value < 0.05$), so genes were further filtered using a threshold of 0.1 Moran_I statistic, resulting in 1,153 remaining genes. Genes were then grouped into modules using Louvain community clustering based on the cells that express them (find_gene_modules).

The resulting gene modules were investigated by scoring the average gene expression within different NMP categories and individual cells using the scanpy score_genes function.

Extant NMP score

To quantify the level of commitment that an extant NMP has towards the Neural or Somitic fates, we reasoned that the number of sibling cells correlates to the degree of commitment that an NMP has towards the neural and somitic lineages. Therefore, we scored NMPs by the relative ratio of Neural and Somitic sibling cells. Specifically, the NMP score is defined as the following:

$$NMP\ score = \frac{\#\ of\ Somitic\ siblings + 1}{\#\ of\ Neural\ siblings + 1}$$

Compositional analysis

Compositional relationships between TLS replicates were calculated using Aitchison Distance.[75] Briefly, the Aitchison Distance measures the distance between two D-dimensional compositional vectors M and N as follows:

$$d(M,N) = \sqrt{\sum_{i=1}^{D}\left[\ln\left(\frac{M_i}{g(M)}\right) - \ln\left(\frac{N_i}{g(N)}\right)\right]^2}$$

where g(M) and g(N) are the geometric means of the M and N vectors respectively. By applying a centered-log ratio (clr) transformation, the Aitchison's distance metric surpasses Euclidean distance for compositional comparisons by providing scale, permutation, and perturbation invariance and accounting for the sub-compositional dominance present in compositional comparisons.[14]

To assess the compositional distances between structures and clones, we calculated the Aitchison's Distance using the complete 14 cell state vector for a comparison of two structures or clones. As some clones were poorly sampled and Aitchison's Distance requires values greater than 0, a pseudo count of 0.0001 was added to each value prior to calculating the Aitchison Distance. Then, the Aitchison's Distance was calculated between all pairwise comparisons of TLS replicates or clones within the same structure.

Progenitor classification

To assess ancestral progenitors within TLS replicates, we categorized internal tree nodes by the extant cells that they produce. Narrowing our investigation to NMP trajectories, we focused specifically on NMP, Neural, and Somitic cell producing nodes, excluding any node that produces Endoderm, Endothelial, PGCLCs, or unassigned cells. To account for low sampling of rare cell types, we combined the Neural and Somitic cell types into singular categories. Specifically, NeuralTube1 and NeuralTube2 cells were labeled as "Neural", while aPSM, pPSM, Somite (-1), Somite 0, Somite, Sclerotome-like, and Dermomyotome-like cells were labeled as "Somitic". Finally, to focus analysis on progenitors, we filtered out nodes that produce fewer than 4 cells. All together, these filtering steps removed 2,198 of the 3,794 nodes across all replicates.

Next, the remaining nodes were classified based on the relative proportions of NMP, Neural, and Somitic extant cells. First, nodes that produce exclusively NMPs were categorized as "Self-renewing" to capture the self-renewing NMPs within an experiment. Second, nodes that produce both Neural and Somitic cell states were classified as "Bipotent". Finally, the remaining nodes were

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

classified as “Neural Committed” or “Somitic Committed” as they were found to produce Neural or Somitic cells, respectively. As the population of NMPs is consumed over time, the “Bipotent”, “Neural”, and “Somitic” nodes were classified regardless of the absence or presence of extent NMPs.

We next classified nodes by the fraction of extant NMPs that they contain. With this, we divided nodes into 3 categories; first, nodes that produce exclusively NMPs were classified as “Self-renewing”. Second, nodes that contain both NMPs as well as some Neural or Somitic were classified as “Differentiating”. Finally, we reasoned that nodes that produce no NMPs had consumed their renewing potential, and therefore we classified them as “Exhausted”.

To visualize the progenitor nodes, we utilized the “Python-Ternary” package to plot the nodes as a ternary plot with each point defined by the relative ratios of NMP, Neural, and Somitic extant cells. To create ternary contour plot heatmaps, we used the “Plotly” package.

Shuffled background validation of progenitor classifications

Internal node dynamics and classifications are dependent upon the composition of the structure and the topology of the tree. To account for this, a shuffled background validation was performed to test the significance of progenitor node counts. Briefly, the shuffled background test randomly shuffles the state assignments of leaves in the tree while maintaining the tree topology to measure the number of progenitor node classes in a random tree. This process was then repeated 500 times to generate a background distribution of progenitor node counts and depths. Then, the significance of the selected tree counts was determined by comparing the measured value to the background distribution.

Due to the continuous editing of Cas9, our lineage tracing approach continues recording nodes even if a cell does not change its developmental potential. Within a tree, this results in nested branches of unipotent nodes that produce exclusively one cell type. Since uni- and multipotent nodes require the same cell states, excess unipotent nodes can hamper our quantification of multipotent nodes. To correct this effect and test the significance of progenitor bipotentiality, we therefore removed nested unipotent nodes and performed the shuffled background test on the pruned tree as described above.

Ternary distance

To quantify the difference in progenitor dynamics between two trees, we measured the distance between the set of nodes of each tree. The ternary distance is defined as the sum of the minimum distances, without replacement, between pairs of nodes in the tree. Specifically, to measure the distance between tree A and tree B, we first identified the pairs of nodes that are closest together. All pairwise node distances between tree A and B are calculated as the Euclidean distance using the fraction of Neural, Somitic, and NMP extant cells for each node.

$$\mathit{node\ distance}(X,Y) = \sqrt{(X_{NMPs} - Y_{NMPs})^2 + (X_N - Y_N)^2 + (X_S - Y_S)^2}$$

$$\begin{pmatrix} A_1B_1 & \cdots & A_1B_n \\ \vdots & \ddots & \vdots \\ A_nB_1 & \cdots & A_nB_n \end{pmatrix}$$

The node pair with the smallest distance is then selected and removed. This process is repeated until there are no more nodes in tree A or tree B, resulting in a set of n node pairs that have the minimum distances between them where n is the number of nodes in the smaller tree. Finally, the ternary distance is calculated as the sum of the distances between all pairs of nodes, normalized by the size of the smaller tree.

$$\mathit{ternary\ distance} = \sum (A_1B_1)\ldots(A_nB_n)$$

Down-sampling TLS replicates

To account for sampling variability as a confounding factor when comparing ternary plots, we down-sampled multiple TLS replicates to a consistent size of 200 cells per tree. Multi-seq Bar5, Bar10, Bar16, Bar19, and Bar22 were selected as each was a larger TLS replicate with greater than 200 cells. Prior to reconstruction, 200 cells were randomly sampled without replacement from each dataset, followed by lineage reconstruction using the Cassiopeia-ILP algorithm and pipeline as described above. This process was repeated 30 times for each of the 5 TLS replicates, resulting in 150 down-sampled trees of 200 cells each. Each of these trees was then analyzed according to the NMP-focused progenitor analysis to collect progenitor node dynamics and ternary plots.

Using these datasets, ternary and compositional distances were calculated between each pairwise tree from the same original dataset (intra) as well as randomly selected pairs of down-sampled trees from different datasets (inter). Finally, the correlation and slope relating the composition and ternary distances in intra- and inter- comparisons was calculated using a linear regression (scipy, linregress, alternative = “two-sided”).

Please cite this article in press as: Bolondi et al., Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids, Developmental Cell (2024), https://doi.org/10.1016/j.devcel.2024.03.024

Lineage tracing processing and lineage reconstruction of explant experiments

To reconstruct trees across multiple timepoints, lineage datasets from each of the recorded timepoints (generation 1 and 2) were merged and reconstructed using the Cassiopeia-Hybrid algorithm as described above. Due to the size of the explant datasets, we ran tree reconstruction once using a single seed (using the optimal seed identified from the TLS1 dataset in Gurobi). To validate the reconstruction of merged datasets, we reasoned that lineage relationships between cells of the same timepoint should be consistent when compared against a tree generated from only cells of that timepoint. To estimate the quality of the combined tree, the combined tree was pruned to contain cells of only a single timepoint and compared against a tree of a single timepoint using the Cassiopeia Triplet Correct Score. This function calculates the proportion of triplet cells in the tree that are the same across two separate trees, providing a quantitative measure of lineage relationship consistency when generation 1 and generation 2 timepoints are merged.

Explant subtree analysis

To investigate the dynamics of progenitors as they restrict to generation 1 or generation 2, nodes on the tree were categorized by the timepoint of the extant cells that they produced. Nodes that produce extant cells of both generation 1 and generation 2 were classified as “Shared” while nodes that produced exclusively generation 1 or exclusively generation 2 cells were classified as “Generation 1” or “Generation 2”, respectively. To investigate the dynamics of progenitors that share a common ancestor, we focused on clones that start as a shared node. Then, subtrees of exclusively Generation 1 or Generation 2 nodes that connect to a Shared ancestor were identified for further study. The dynamics of the Shared ancestor nodes were investigated by classifying each node using only cells from generation 1, generation 2, or all cells. Narrowing our investigation to only NMP trajectories, we removed all nodes that were not classified as NMP-related (Bipotent, Neural, and Somitic) in all three timepoint classifications. The resulting dynamics and conversions between timepoints were visualized via alluvial plots using Plotly.

The dynamics of subtrees and shared nodes were classified using the progenitor analysis described above and visualized via violin plots. The violin plots across each timepoint were scaled by the number of nodes to match across the three subtree categories (seaborn, violinplot, scale = “count”).

QUANTIFICATION AND STATISTICAL ANALYSIS

All plots and statistical analyses were performed with GraphPad PRISM 8, R v.3.6.3 and Python (3.9.11). Information regarding these analyses can be found in the STAR Methods (see computational analysis section) or in the figure legends. No methods were used to determine whether the data met the assumptions of the statistical approach.