

Zebrafish Neuromesodermal Progenitors Undergo a Critical State Transition *in vivo*

Kane Toh¹ Dillan Saunders¹ Berta Verd² and Benjamin Steventon¹

1) Department of Genetics, University of Cambridge, Cambridge, UK, CB2 3EH

2) Department of Zoology, University of Oxford, Oxford, UK, OX1 3SZ

Summary

The transition state model of cell differentiation proposes that a transient window of gene expression stochasticity precedes entry into a differentiated state. As this has been assessed primarily *in vitro*, we sought to explore whether it can also be observed *in vivo*. Zebrafish neuromesodermal progenitors (NMPs) differentiate into spinal cord and paraxial mesoderm at the late somitogenesis stages. We observed an increase in gene expression variability at the 24 somite stage (24ss) prior to their differentiation. From our analysis of a published 18ss scRNA-seq dataset, we showed that the NMP population possesses a signature consistent with a population undergoing a critical transition. By building *in silico* composite gene expression maps from our image data, we were able to assign an 'NM index' to each *in silico* NMP based on the cumulative expression of its neural and mesodermal markers. With the NM index distributions, we demonstrated that cell population heterogeneity of the NMPs peaked at 24ss. We then incorporated stochasticity and non-autonomy into a genetic toggle switch model and uncovered the existence of rebellious cells, which we then confirmed by reexamining the composite maps. Taken together, our work supports the transition state model within an endogenous cell fate decision making event.

Introduction

Neuromesodermal progenitors (NMPs) are axial progenitors that co-express the lineage-specific transcription factors *Brachyury/T/Tbxta* and *Sox2* and are competent to generate both neural (e.g. spinal cord) and mesodermal (e.g. somite) fates at the single-cell level (Henrique et al., 2015; Wymeersch et al., 2021). Bipotent NM cells have been identified in amniotes such as mouse (Cambray & Wilson, 2007; Cambray & Wilson, 2002) and chick (Brown & Storey, 2000; Guillot et al., 2021; Wood et al., 2019) as well as anamniotes such as *Xenopus* (Davis & Kirschner, 2000; Gont et al., 1993), axolotl (Taniguchi et al., 2017) and zebrafish (Martin & Kimelman, 2012). Therefore, they are an evolutionary conserved cell population whose decision to generate spinal cord and paraxial mesoderm provides an ideal system to explore the mechanisms of cell fate decision making *in vivo*.

The degree to which NMps divide to produce daughter cells of both neural and mesodermal fates depends on species-specific growth dynamics (Steventon & Martinez-Arias, 2017). In the zebrafish embryo, there is little volumetric growth associated with posterior body elongation (Steventon et al., 2016) and proliferation stops abruptly within the embryo around the 10 somite stage (ss) (Bouldin et al., 2014; Zhang et al., 2008). Correspondingly, zebrafish tailbud NMps are a largely quiescent pool of monofated progenitors that give rise to a limited portion of the posterior body axis (Attardi et al., 2018; Bouldin et al., 2014). In the mouse embryo, using retrospective clonal analysis, long clones originating from a single cell have been observed in both neural and mesodermal tissues (Tzouanacou et al., 2009), which is consistent with a proliferative phase in the mouse NMps at around E9.5 (Wymeersch et al., 2016). Despite this difference in developmental dynamics, two independent lines of evidence support the notion that zebrafish NMps are, like all other vertebrate NMps, competent towards both neural and mesodermal fates. Firstly, single cell transplantation experiments demonstrate that zebrafish NMps can be steered towards either neural or mesodermal fates upon manipulation of the canonical Wnt pathway (Martin & Kimelman, 2012). Secondly, a single cell transcriptomic signature that contains conserved markers of both spinal cord and paraxial mesoderm states have been discovered for the zebrafish NMps at late gastrulation/early tailbud stages of development (Lukoseviciute et al., 2021). Thus, a conceptual clarification between NM competent cells and NM progenitors (NMps) has been proposed, of which a differing proportion of NM competent cells act as NMps in a stage and specific-specific manner dependent on the rate of proliferation (Binagui-Casas et al., 2021; Sambasivan & Steventon, 2021). In this paper, we refer to these cells as zebrafish tailbud ‘NMps’ to remain consistent with previous literature, although they are better understood as NM competent cells at post 10ss of development.

How do these zebrafish tailbud NMps differentiate into their NM derivatives? Differentiation has been widely characterized as an ordered and largely deterministic succession of cellular states, specifically transcriptomic states, that emerge from the activation of a set of master transcription factors in a gene regulatory network (Davis et al., 1987; Whyte et al., 2013). If transcriptomic states strongly correlate to developmental lineage, then we can sort single cells along a pseudotemporal axis of developmental progression using their transcriptomic states as the similarity measure and infer the gene expression trajectories within these differentiating cells. Elucidating the pseudotemporal axis has uncovered numerous insights into development (Wagner et al., 2018; Wolf et al., 2019) and disease (Mukherjee et al., 2020; Petti et al., 2022). Despite their utility, pseudotemporal ordering algorithms make a critical simplifying assumption: cells with similar transcriptomic profiles are assigned to be closer together in their developmental maturity along a lineage (Schier, 2020; Tritschler et al., 2019). This biological assumption has been challenged by several observations. First, *in vitro* studies revealed the prevalence of non-genetic heterogeneities within clonal stem cell populations, where cells

stochastically transition between distinct metastable states despite being functionally homogeneous (Canham et al., 2010; Hayashi et al., 2008; Trott et al., 2012). In addition, global transcriptomic trajectories may be driven by complex dynamics such as slow fluctuations that persist across cell division cycles (Chang et al., 2008) and oscillatory dynamics in key regulators (Verd et al., 2018). Furthermore, distinct trajectories may converge to the same terminal fate (Packer et al., 2019). These observations suggest that the relationship between cell fate and transcriptomic state can be complex (Casey et al., 2020) and additional information is required before constraining the possible dynamics that arise from snapshot data with the maximum parsimony assumption (Tanay & Regev, 2018; Weinreb et al., 2018).

An alternative class of models proposes that differentiation is a two-stage process: an initial regime of increased heterogeneity within the differentiating population due to elevated stochasticity in gene expression, followed by the stabilisation of the gene expression pattern during cell-fate determination. During the initial stochastic phase, transcriptomic states and cell fates are less correlated as gene expression heterogeneity increases. Models that belong in this class include the Darwinian model of cellular differentiation (Kupiec, 1997; Minelli et al., 2014; Paldi, 2020), the ‘exploratory’ model of stem cell decision-making (Halley et al., 2009) and the ‘transition state’ model (Antolović et al., 2019; Arias & Hayward, 2006; Brackston et al., 2018; Moris et al., 2016; Muñoz-Descalzo et al., 2012; Rué & Martinez Arias, 2015). Taking a statistical mechanical perspective, this phenomenon of ‘regulated stochasticity’ is consistent with cellular differentiation being a critical phase transition (Teschendorff & Feinberg, 2021). Experimental observations of a surge in gene expression variability that precedes a commitment phase are found predominantly in *in vitro* models such as hematopoietic stem cell differentiation models (Hu et al., 1997; Mojtahedi et al., 2016; Moussy et al., 2017; Pina et al., 2012; Richard et al., 2016), induced pluripotent stem cells (iPSCs) (Bargaje et al., 2017; Buganim et al., 2012) and mouse embryonic stem cells (mESCs) (Moris et al., 2018; Semrau et al., 2017; Stumpf et al., 2017). In contrast, *in vivo* observations of this phenomenon has been comparatively rare (Antolović et al., 2019; Peláez et al., 2015). *In vivo* evidence are vital to ensure that the preceding *in vitro* observations are not due to artifacts of cell culture conditions (MacArthur & Lemischka, 2013; Smith, 2013) or reporter dynamics (Smith et al., 2017).

In this paper, we assessed the transition state hypothesis *in vivo* during the zebrafish tailbud NMP differentiation event. Our results can be grouped according to two features of the hypothesis:

1. Transient increase in transcriptional heterogeneity during NMP differentiation.

As photolabels of the NMP region at the 12ss revealed that cells only contribute to somites and spinal cord from the 24 somite level onwards (Attardi et al., 2018), we focused on a time-window between the 18ss and 30ss to capture the commitment

event. By quantifying the single-cell levels of nuclear *sox2* and *tbxta* expression in NMps from 18ss to 30ss *in situ*, we demonstrate that the heterogeneity in expression of both genes as well as the variability in Nmp number peak at 24ss. In addition, by examining a publicly available 18ss scRNA-seq dataset of the zebrafish embryo (Wagner et al., 2018), we found that NMps have a higher critical index and transcriptional noise relative to their derivatives, supporting the view that the Nmp population is undergoing a critical transition. Furthermore, by combining the expression of multiple Nmp marker genes across multiple samples with an image alignment pipeline (ZebReg) and computing the ‘NM index’, we found that the Shannon entropy, a measure of the population heterogeneity, also peaks at 24ss.

2. Loosening of the relationship between cell state and cell fate: existence of ‘Rebellious’ cells.

Given the importance of stochasticity in the initial stage of the transition state model, we modelled the behaviour of a stochastic non-autonomous toggle switch to mimic the gene expression heterogeneity observed during the Nmp differentiation event. The incorporation of non-autonomy in the stochastic model reflects the fact that embryonic development *in vivo* is a dynamic process that unfolds over time, where intrinsic changes in a cell’s molecular state exists in a causal feedback loop with extrinsic changes that occur in its surrounding environment (Busby & Steventon, 2021). Interestingly, we identified a subset of trajectories that transiently reside in the competing attractor before eventually switching into the primary attractor at the end of the simulation. Following the work of Mojtahedi and colleagues, we labelled these trajectories as ‘Rebellious’ (Mojtahedi et al., 2016). We exploited the relative biological simplicity of the zebrafish Nmp system to relate cellular states to cellular fates by examining the spatial locations of the NMps within our ZebReg composite maps. Reexamining our composite maps, we identified an increase in the number of Rebellious cells at the 24ss within the mesoderm-fated domain.

Taken together, our work supports the existence of a transition state and the presence of ‘rebellious’ cells *in vivo* during zebrafish Nmp differentiation.

Results

Heterogeneity in *sox2* and *tbxta* expression and variability in the number and locations of NMps peak at 24ss

To assess the number and location of zebrafish tailbud NMps over time, we performed HCR stains for *sox2* and *tbxta* to quantify the mRNA expression of single cells *in situ* within the zebrafish tailbud (Figure S2A-E). First, we compared the expression of nuclear *sox2* and *tbxta* in the NMps against the posterior notochord and posterior neural tube populations (Figure S2F-H’). We find that the posterior notochord population has a tight distribution of nuclear *sox2* with a mean normalised

intensity close to 0 and a broader nuclear *tbxta* distribution (Figure 1D-D'). Conversely, the posterior neural tube population has a tight distribution of nuclear *tbxta* with a mean normalised intensity close to 0 and a broader nuclear *sox2* distribution (Figure 1E-E'). On the other hand, the NMps are distinct from both populations as they have broad marginal distributions of both nuclear *sox2* and *tbxta* (Figure 1F-F").

Next, we quantified the nuclear *sox2* and *tbxta* levels within the NMP population across the different somitogenesis stages. In the gene expression scatterplots (Figure 1J), we find that most NMps are *sox2*_{low}*tbxta*_{low}. However, at 24ss, we also find a greater number of *sox2*_{int}*tbxta*_{int} NMps, reflecting a transient increase in the transcriptional heterogeneity of the NM gene expression states. We then quantified the number and position of NMps at each stage across multiple individual tailbud samples. We found significant variation in the position (Figure 1G-I) and number (Figure 1K) of NMps across samples at all stages under study. Notably, peak variability in NMP number occurred at the 24ss (Figure 1K). Taken together, our analysis demonstrates that a transient phase of increased heterogeneity in *sox2* and *tbxta* expression states occurs around 24ss. This closely matches the developmental stage at which labelled NMps contribute to both spinal cord and paraxial mesoderm (Attardi et al., 2018) and therefore suggests that the increased heterogeneity precedes the commitment to either NM fate.

Analysis of 18ss scRNA-seq data reveals a peak in the critical index and transcriptional noise index in the NMP population relative to its derivatives.

A second prediction of the transition state model is that cells should explore a larger region of gene expression space prior to cell fate commitment as the progenitor basin flattens, resulting in a more dispersed 'cloud' of points in state space (Huang, 2009). Consequently, cell population heterogeneity increases, whereas between-gene variation decreases as cells up-regulate groups of either neural or mesodermal genes in coordinated fashion (Mojtahedi et al., 2016). To assess this in the context of zebrafish NMps *in vivo*, we made use of a recently published single-cell RNAseq dataset at 18ss (Wagner et al., 2018).

First, we reanalysed the scRNA-seq data using an independent dimensional reduction and clustering approach to obtain the 8 tailbud subclusters that include the NMps and their derivatives (See STAR Methods). We include the expression of selected differentially expressed genes for the neural, mesodermal and NMP clusters as dot plots (Figure 2C-C") and provide information on marker gene expression for all 8 clusters in Table S3. In support of our manual annotation of the NMP cluster, we find that most of the *sox2*₊*tbxta*₊ co-expressing cells are found within the NMP cluster, and the NMP cluster is sandwiched between two neural clusters and five mesodermal clusters (Figure 2A-B). This is consistent with *sox2* and *tbxta* emerging as differentially expressed genes in this cluster (Table S3). In addition, we validated

a subset of the identified NMP marker genes experimentally via HCR and find that they are all expressed within the NMPs within the tailbud (Figure S6), supporting the robustness of our *in silico* analysis.

We find that the NMP cluster is enriched for the posterior Hox genes *hoxc13b*, *hoxc13a* and *hoxa13b*, with avg_log2FC of 1.02, 0.94 and 0.78 respectively (Table S3). Avg_log2FC measures the log fold-change of the average expression of these genes in the NMP cluster versus the other clusters (Stuart et al., 2019). This is consistent with reports of their roles in mediating the termination of axial elongation (Aires et al., 2019; Young et al., 2009) and directing mesoderm formation in the NMP niche (Ye & Kimelman, 2020). In terms of signalling pathways, *wnt8a* and *fgf8a* appear as the top two genes enriched in the NMP cluster, both of which are actively involved in NMP maintenance and differentiation (Goto et al., 2017; Row et al., 2016). Notably, *fgf8a* is expressed in >80% of cells in the NMP cluster (PCT1 = 0.816) and <8% of cells in all other clusters (PCT2 = 0.075). Finally, our analysis also identified *cyp26a1*, a retinoic-acid degrading enzyme that safeguards the Wnt/*tbxta* positive feedback loop, with an avg_log2FC of 1.02 (Martin & Kimelman, 2010).

Besides identifying the known molecular players in NMP differentiation, we also uncovered numerous other genes involved in a diverse range of processes. Genes with annotations that implicate their roles in signalling pathways feature prominently and include *wls*, *wnt8-2*, and *depdc7* for Wnt, *angptl2b* and *her12* for Notch, *nog2* and *id3* for BMP and *fgf4* for FGF signalling. Also, three genes were annotated with cytoskeleton-associated processes (*tagln3b* and *enc1* have actin-binding activity and *kif26ab* regulates microtubule motor activity), two possess histone deacetylase binding activity (*znf703* and *kdm6a*) and another two are associated with ubiquitination (*traf4a* and *ubl3a*). Interestingly, *foxd3*, a neural crest marker, emerged as a candidate that is enriched in the NMP cluster, with an adjusted p-value of 1.92×10^{-6} . It is expressed in >20% of cells in the NMP cluster and <4% of cells across all the other 7 clusters (Table S3). This observation confirms the results of recent study that revealed a common transcriptomic signature of the neural crest and NM populations (Lukoseviciute et al., 2021).

Cell fate decision making has been proposed to be a critical transition event, with both these features captured in a single 'Critical Transition Index' that has previously been shown to predict a cell fate decision making event within blood progenitors as they commit to either myeloid or erythroid lineages (Mojtahedi et al., 2016). In similar vein, we computed the critical indices for the NMP, neural (*pou5f3*+ posterior NT and posterior NT) and mesodermal (posterior PSM and tail somites) clusters (Figure 2D-D'). Along both the neural and mesodermal differentiation trajectories, the NMP cluster cells have the highest critical index, which is consistent with a cell population undergoing a dynamical bifurcation. Next, we assessed the level of transcriptional noise in the population. We observed that the NMP cluster cells have a higher

transcriptional noise relative to the other cell populations (Figure 2E-E'). Therefore, both quantitative indices support the hypothesis that NMPs are approaching a critical transition at 18ss.

Gene expression imputation and the construction of a composite map via ZebReg demonstrates a peak in the NM index entropy at 24ss.

Our observation that the number and position of NMPs vary extensively between stage-matched embryos, especially at the 24ss, suggests that there is significant variability in *sox2* and *tbxta* expression within the NMPs. Consequently, fixed measurements of gene expression from a single sample alone would be inaccurate as it can only give an instantaneous snapshot capturing one out of many different gene expression states that the NMP population can potentially explore. To leverage the gene expression information across multiple tailbud samples, we developed a tool called ZebReg that takes images of stage-matched zebrafish tailbud samples as inputs, converts them into point clouds and registers the point clouds together to construct composite gene expression target maps (Figure 3A).

We first used ZebReg to impute the expression of 8 genes into three composite maps, one for each stage at 18ss, 24ss and 28ss. Each composite map was constructed by combining the gene expression of 6 different images across different HCR experiments (Figure S8; Table S2), and each map displays the expression of *sox2* (Figure 3B), *tbxta* (Figure 3B') and 6 other neural or mesodermal marker genes selected from our scRNA-seq analysis within a target point cloud. As these maps contain spatial information of the tailbud cells, we could identify the *in silico* NMPs by virtue of their *sox2* and *tbxta* co-expression as well as their locations on the composite map (Figure S12). These *in silico* NMPs in the composite maps were also found to be within the NMP regions in our probability map which were identified via a different approach (Figure S14).

To assess the fidelity of our gene expression imputation procedure, we performed a gene-by-gene qualitative inspection of the spatial expression patterns in the composite maps to the corresponding patterns observed in the original HCR images at 18ss. A visual comparison between the imputed and original images demonstrates a strong resemblance in their expression patterns (Figure 3C). For instance, *sox2* and *cdh6* are expressed strongly in the posterior neural tube and hypochord but not in the notochord. *tbxta* is expressed strongly in the notochord progenitor zone and the dorsal PW. Thus, the overall visual correspondence between the original and imputed gene expression images is evidence that ZebReg has aligned these images appropriately, at least when assessed on a qualitative level. We also performed additional *in silico* validation experiments and demonstrated that ZebReg also preserves the quantitative relationships between genes (See STAR Methods).

Using the composite maps, we first constructed an ‘NM index’ which combines the information across the 8 genes and plotted the NM index distributions for the NMps in the three composite maps (Figure 3D). These distributions reflect the neural/mesodermal biases of the *in silico* NMps in these three stages, and cells can be classified as being either neural-biased, mesoderm-biased or indecisive based on their NM index value (Figure S13A). We find that there is a consistent neural bias in the NMps across all three stages which is reflected by the median NM index value.

To quantify the NM heterogeneity of the *in silico* NMps between these stages, we computed a series of Shannon entropy estimators. Examining the empirical maximum-likelihood entropy estimator (H) with the natural unit of information (nat), we observed a surge in value at 24ss with $H = 6.23$ nat, compared to the neighboring values of $H = 4.17$ nat at 18ss and $H = 3.18$ nat at 28ss (Figure 3E). This increase in entropy followed by a decline was also observed in the other entropy estimates (Figure 3E). Thus, our data suggest that the Nmp population heterogeneity, measured by the entropy, peaks at 24ss.

Stochastic modelling of a genetic toggle switch with a time-varying input uncovers ‘rebellious’ cells.

Our results indicate widespread cellular variability in the expression of NM genes during the process of Nmp differentiation. Given the low-level expression of *sox2* and *tbxta* for the majority of NMps (Figure 1F), stochastic effects are likely to dominate and drive the cellular heterogeneity in the population (Kærn et al., 2005). To understand how noise-induced heterogeneities can arise in the context of the zebrafish tailbud, we first focused our analysis on constructing a single-cell resolution quantitative spatial landscape of the canonical Wnt signaling environment as it is one of the main signaling factors that bias the NM fate balance (Martin & Kimelman, 2012). We examined the nuclear expression of *tcf* mRNA (a readout of canonical Wnt) in the dorsal PW, intermediate zone and posterior PSM at the 18ss, 24ss and 30ss (Figure 4A-C). *tcf* expression was found to increase consistently along the dorsal PW to intermediate zone to PSM trajectory (Figure 4D), which traces the approximate path of the mesoderm-fated NMps.

To incorporate the changes in Wnt signalling levels as NMps differentiate into their mesodermal fates, we constructed a stochastic non-autonomous genetic toggle switch model of *sox2* and *tbxta*, where both genes mutually repress each other (Figure 4E). Whilst a direct repression between *tbxta* and *sox2* has not been shown, the interaction between *sox2* and *tbxta* in other Nmp systems, whether direct or indirect, is an antagonistic one that drives the mutually exclusive choice between the neural or mesodermal fate (Bouldin et al., 2015; Gouti et al., 2017).

When stochastic trajectories were first equilibrated within a tristable regime in the first interval of our simulation (Figure 4F, 01 Constant), they spread out to occupy all

three attractors with considerable frequency (Figure 4G). To reflect the increase in Wnt signalling within the differentiating NMPs along the mesodermal trajectory, we then increased the external activation parameter α_1 linearly over the next simulation interval (Figure 4F, 02 Changing). At the end of the second interval, we observed that most stochastic trajectories reside within the primary ‘mesodermal’ attractor (attractor 0) (Figure 4G’). However, some trajectories deviate from this behaviour - around 9.5% of trajectories exhibit transient behaviour and are situated between both neural and mesodermal attractors. Unexpectedly, a smaller percentage of trajectories (5.1%) still reside within the alternate ‘neural’ attractor (attractor 1). In the final simulation interval (03 Constant) during which the external activation parameter α_1 is kept constant at a value of 0.8 (Figure 4G’), these trajectories eventually switch into the primary attractor and adopt the appropriate cell fate.

Thus, the introduction of non-autonomy into a stochastic toggle switch leads to a novel phenomenon that was not present in the autonomous case (Figure S16B) - the emergence of ‘rebellious’ cells with gene expression states (high *sox2*, low *tbxta*) inconsistent with their signalling environment (high Wnt) and eventual mesodermal fate.

ZebReg’s composite maps reveal that the number of Rebellious cells peak at 24ss.

We returned to ZebReg’s composite maps to assess whether these Rebellious cells can be found within the *in silico* NMP population. To identify these cells, we need to compare their canonical Wnt signalling activities (*tcf* levels) and eventual cell fates (neural or mesodermal) against their NM gene expression states (NM index levels). Given that our data consist of fixed snapshot images of the NMPs, we do not have direct information of their prospective fates. Despite this, as NMPs at the 21ss onwards are spatially segregated and have low levels of proliferation (Attardi et al., 2018, Figure S11A-B, D-E), we reason by an argument of elimination that since NMPs also display low levels of apoptosis throughout 18ss to 30ss (Figure S11A,C), we can reliably predict the cell fate of NMPs based on their spatial locations in the composite maps at 18ss, 24ss and 30ss.

Having established that we can infer an *in silico* NMP’s fate prospectively from its location in the composite map, we defined approximate neural-fated and mesodermal-fated domains following the fate map of Attardi and colleagues and assessed the NM index levels of the NMPs within these domains (Figure 5A). We found cells with NM gene expression profiles that are inconsistent with their prospective fates and labelled these cells as ‘Incongruent’. Conversely, cells with compatible state-fate relationships are labelled ‘Congruent’. We observed Incongruent cells with low/high NM index levels residing in the neural/mesoderm-fated domains (Figure 5A blue arrows). As *sox2* and *tbxta* are the primary orchestrators of the neural and mesodermal gene expression programmes

respectively, these ‘incongruent’ cells are also found in our original HCR images as *sox2 (tbxta)*-high NMPs within the neural (mesoderm)-fated domains (Figure 5B).

Incongruent cells can be further classified as ‘Compliant’ or ‘Rebellious’ depending on whether their Wnt signalling activities (*tcf* expression levels) are consistent or inconsistent with their NM gene expression states (Table S1). We quantified the proportion of Compliant, Rebellious and Congruent cells in the mesodermal and neural-fated domains of our three composite maps (Figure 5C). At 18ss and 28ss, most cells are Congruent. Also, more Incongruent cells are found in the mesoderm-fated domain than the neural-fated domain. However, at the 24ss, we find a greater number of Incongruent cells (Compliant and Rebellious) than Congruent cells in the mesoderm-fated domain. Specifically, the number of Rebellious cells in the mesoderm-fated domain peaks at this stage. Thus, consistent with the transition state model, we find a loosening of the relationship between cell state and fate as reflected by the increase in the number of rebellious NMPs at the 24ss prior to their commitment to the NM fate.

Discussion

Zebrafish tailbud NMPs have proven to be an attractive *in vivo* system to assess the transition state hypothesis. Specifically, we investigated whether a transient window of elevated stochasticity in gene expression precedes the NMP differentiation event at around 24ss.

Our discovery of rebellious cells in the ZebReg composite maps recapitulates the finding by Mojtahedi and colleagues in their *in vitro* study on the differentiation of a multipotent hematopoietic cell line (Mojtahedi et al., 2016). Rebellious cells emerge at day 3 post-treatment as cells that express an erythroid profile when stimulated with Granulocyte macrophage colony-stimulating factor /IL-3 or a myeloid profile when stimulated with erythropoietin. Eventually, cells disappear at day 6 post-treatment. In addition, ‘edge’ cells were identified in cancer cell lines as cells that adopt a gene expression profile that is different from the average profile in the population distribution (Li et al., 2016). This phenomenon is not unprecedented *in vivo*. In an earlier study on *Xenopus* embryonic development (Wardle & Smith, 2004), cells that express a lineage marker at the ‘wrong’ place, such as Goosecoid expressing cells in the ventral instead of the dorsal region of the embryo, were labelled as ‘rogue’ cells to indicate their abnormal expression profile. These cells appear more frequently in the early gastrula stage and reduce in frequency at the late gastrula stage. In both cases, these rebellious/rogue cells are proposed to ‘fit in or die trying’ - they would either die by apoptosis or transdifferentiate to adopt the appropriate gene expression profile if rescued by delivery of the appropriate signal or through interactions with neighbouring cells via the community effect.

In the context of zebrafish NMP differentiation, our finding that the proportion of rebellious cells (Figure 5C) is highest at 24ss within the mesoderm-fated domain extends a recent work on the connection between morphogenetic movements and mesoderm fate acquisition in the zebrafish NMps (Kinney et al., 2020). *Sox2* and canonical Wnt co-expression in mesoderm-fated NMps primes these cells towards both neural and mesodermal fates and acts as a developmental checkpoint that traps these cells in a poised, intermediate state. This intermediate state where EMT is delayed resembles a hybrid EMT transition state found in cells with high potential for metastasis (Yang et al., 2020). In fact, *tbxta* (Brachyury) is a driver of EMT in various tumors and is correlated with metastatic activity and the acquisition of a mesenchymal phenotype (Chen et al., 2020). Thus, our work emphasises a strong connection between the hybrid EMT transition state expressing multiple intermediate cell states (Sha et al., 2019) and the neural-mesodermal transition state (Steventon & Martinez-Arias, 2017). The correspondence between morphological fluctuations and the entry into a transitory state was also recently proposed in a study on hematopoietic stem and progenitor cells (Moussy et al., 2017). As NMps exit the transition state and differentiate into the neural or mesodermal fates, the heterogeneity in *sox2* and *tbxta* expression is resolved as NMps adopt either a high *sox2*/low *tbxta* (neural) or high *tbxta*/low *sox2* (mesodermal) expression profile (Figure 1D-F). This is consistent with the proposed role of Sox2 and Bra protein level ratios dictating the specific cell movements associated with each lineage (Romanos et al., 2021).

Recent work on multipotent zebrafish neural crest cells suggests that at least a portion of the neural biased trunk neural crest (NC) progenitors arise from early neural biased zebrafish NMps at 5-6ss (Lukoseviciute et al., 2021). A similar conclusion was reached in multiple studies of in vitro human pluripotent stem cell-derived axial progenitors, demonstrating that the generation of trunk NCs involves an obligatory NMP intermediate (Frith et al., 2018; Hackland et al., 2019). We identified *sp5l*, *cdh6*, *znf703* and *foxd3* as differentially expressed genes of the NMP cluster; all of which have important roles in neural crest specification. In addition, many differentially expressed genes identified from the NMP cluster are involved in signalling pathways (FGF, Wnt, BMP) and the synergistic action of these pathways play a critical role in neural crest differentiation (Sauka-Spengler & Bronner-Fraser, 2008). When we photolabelled the dorsal PW (NMP region) at 18ss and tracked these cells until 28ss, we noticed that the anterior photolabels in the dorsal neural tube appear to be emigrating away whilst the posterior labels do not show signs of migration (Figure S11D). 24 hours later, the photolabels were found to have spread more anteriorly, with the anterior labels appearing more dispersed ventrally. When we photolabelled the dorsal PW at 28ss, we also found similar photolabels in the dorsal neural tube 24 hours post-photolabeling (Figure S11E). The localisation of the labels in the dorsal neural tube alongside the anterior pattern of cell migration strongly suggest that the differentiation of the NMP-derived neural progenitors into the trunk NC progenitors continues throughout somitogenesis and occurs even as

we approach the end of somitogenesis. Therefore, we extend the observation made by Lukoseviciute and colleagues, providing support for an NMP to trunk NC progenitor lineage that occurs even in the later tailbud NMP population.

To the best of our knowledge, our work is the first to directly catalog the transient surge in heterogeneity in mRNA expression *in vivo* during an endogenous differentiation event in a wild-type vertebrate species. Whilst several studies have proposed mechanistic models to explain the relationship between transcriptional heterogeneity and cell fate commitment (Antolović et al., 2017; Pina et al., 2012) and even functional pluripotency (MacArthur & Lemischka, 2013), our study was not designed to discriminate between these causal models. Instead, we focused on assessing the association between cell fate commitment and the increase in gene expression heterogeneity *in vivo*. Future work, outside the scope of this paper, is necessary to fill in the mechanistic details that generate these heterogeneities during cell fate transitions *in vivo*.

Taken together, our work supports the existence of a transition state within an endogenous cell fate decision making event. Recognising the functional importance of transcriptional stochasticity and non-genetic heterogeneities during differentiation has important practical consequences. It drove the discovery that regulators of transcriptional noise may play a general role in the acquisition of malignancy by modulating the balance between proliferation and differentiation (Domingues et al., 2020), and may be an important dimension to consider when improving the efficacy of mesenchymal stem cell-based therapies (McLeod & Mauck, 2017; Pacini, 2014). Seen alongside the evidence presented from other systems, it becomes increasingly plausible that the transition state is not an idiosyncrasy of *in vitro* culture conditions or a peculiarity of cancer models. We await future developments on whether the critical behaviours predicted in the transition state model are a universal characteristic of cell state transitions in biological systems *in vivo*.

Acknowledgements

We would like to thank Nick Monk for advice on the computational modelling project. We also thank all members of the Steventon lab for comments on the manuscript. K.T. is funded by the Cambridge Commonwealth, European & International Trust under a Cambridge International Scholarship. D.S. was supported by the Wellcome Trust-funded Developmental Mechanisms PhD programme (220022/Z/19/Z). B.V. was supported by the Department of Zoology at the University of Oxford. B.J.S. was supported by a Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (109408/Z/15/Z).

Author contributions

K.T. and B.S. designed the study and wrote the manuscript; K.T. and D.S. jointly developed ZebReg; K.T. performed the experiments and analysed the data; B.V. provided advice on the computational modelling approach. All authors read and approved the final manuscript.

References

- Aires, R., de Lemos, L., Nóvoa, A., Jurberg, A. D., Mascres, B., Duboule, D., & Mallo, M. (2019). Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Developmental Cell*, 48(3), 383-395.e8. <https://doi.org/10.1016/j.devcel.2018.12.004>
- Antolović, V., Lenn, T., Miermont, A., & Chubb, J. R. (2019). Transition state dynamics during a stochastic fate choice. *Development (Cambridge)*, 146(12). <https://doi.org/10.1242/dev.173740>
- Antolović, V., Miermont, A., Corrigan, A. M., & Chubb, J. R. (2017). Generation of Single-Cell Transcript Variability by Repression. *Current Biology*, 27(12), 1811-1817.e3. <https://doi.org/10.1016/j.cub.2017.05.028>
- Arias, A. M., & Hayward, P. (2006). Filtering transcriptional noise during development: concepts and mechanisms. *Nature Reviews Genetics*, 7(1), 34-44. <http://www.nature.com/nrg/journal/v7/n1/full/nrg1750.html>
- Attardi, A., Fulton, T., Florescu, M., Shah, G., Muresan, L., Lenz, M. O., Lancaster, C., Huisken, J., van Oudenaarden, A., & Steventon, B. (2018). Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. *Development (Cambridge, England)*, 145(21). <https://doi.org/10.1242/dev.166728>
- Bargaje, R., Trachana, K., Shelton, M. N., McGinnis, C. S., Zhou, J. X., Chadick, C., Cook, S., Cavanaugh, C., Huang, S., & Hood, L. (2017). Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. *Proceedings of the National Academy of Sciences*, 114(9), 2271-2276. <https://doi.org/10.1073/pnas.1621412114>
- Besl, P. J., & McKay, N. D. (1992). <title>Method for registration of 3-D shapes</title>. *Sensor Fusion IV: Control Paradigms and Data Structures*, 1611(April 1992), 586-606. <https://doi.org/10.1117/12.57955>
- Binagui-Casas, A., Dias, A., Guillot, C., Metzis, V., & Saunders, D. (2021). Building consensus in neuromesodermal research: Current advances and future biomedical perspectives. *Current Opinion in Cell Biology*, 73(October), 133-140. <https://doi.org/10.1016/j.ceb.2021.08.003>
- Bouldin, C. M., Manning, A. J., Peng, Y.-H., Farr, G. H., Hung, K. L., Dong, A., & Kimelman, D. (2015). Wnt signaling and tbx16 form a bistable switch to commit bipotential progenitors to mesoderm. *Development*, 142(14), 2499-2507. <https://doi.org/10.1242/dev.124024>
- Bouldin, Courtney M., Snelson, C. D., Farr, G. H., & Kimelman, D. (2014). Restricted expression of cdc25a in the tailbud is essential for formation of the zebrafish posterior body. *Genes and Development*, 28(4), 384-395. <https://doi.org/10.1101/gad.233577.113>
- Brackston, R. D., Lakatos, E., & Stumpf, M. P. H. (2018). *Transition state characteristics during cell differentiation*. 1-22.
- Brown, J. M., & Storey, K. G. (2000). A region of the vertebrate neural plate in which neighbouring cells can adopt neural or epidermal fates. *Current Biology*, 10(14), 869-872. [https://doi.org/10.1016/S0960-9822\(00\)00601-1](https://doi.org/10.1016/S0960-9822(00)00601-1)
- Buganim, Y., Faddah, D. a, Cheng, A. W., Itskovich, E., Ganz, K., Klemm, S. L., & Oudenaarden, A. Van. (2012). Single-cell gene expression analyses of cellular reprogramming reveal a stochastic early and hierarchic late phase. *Cell*, 150(6), 1209-1222. <https://doi.org/10.1016/j.cell.2012.08.023>.Single-cell
- Busby, L., & Steventon, B. (2021). Tissue tectonics and the multi-scale regulation of developmental timing. *Interface Focus*, 11(3). <https://doi.org/10.1098/rsfs.2020.0057>
- Cambray, N., & Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. *Development*, 134(15), 2829-2840. <https://doi.org/10.1242/dev.02877>
- Cambray, Noemí, & Wilson, V. (2002). Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. *Development*, 129(20), 4855-4866. <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed%7B&%7Did=12361976%7B&%7Dretmode=ref%7B&%7Dcmd=prlinks>
- Campbell, P. D., Chao, J. a., Singer, R. H., & Marlow, F. L. (2015). Dynamic visualization of transcription and RNA subcellular localization in zebrafish. *Development*, 1368-1374. <https://doi.org/10.1242/dev.118968>
- Canham, M. A., Sharov, A. A., Ko, M. S. H., & Brickman, J. M. (2010). Functional Heterogeneity of Embryonic Stem Cells Revealed through Translational Amplification of an Early Endodermal Transcript. *PLoS Biology*, 8(5), e1000379. <http://dx.plos.org/10.1371/journal.pbio.1000379>
- Casey, M. J., Stumpf, P. S., & MacArthur, B. D. (2020). Theory of cell fate. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 12(2), 1-11. <https://doi.org/10.1002/wsbm.1471>
- Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., & Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453(7194), 544-547. <http://www.nature.com/doi/10.1038/nature06965>
- Chen, M., Wu, Y., Zhang, H., Li, S., Zhou, J., & Shen, J. (2020). The Roles of Embryonic Transcription Factor Brachyury in Tumorigenesis and Progression. *Frontiers in Oncology*, 10(June), 1-9. <https://doi.org/10.3389/fonc.2020.00961>
- Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust . *Development*, 145(12), dev165753. <https://doi.org/10.1242/dev.165753>
- Davis, R. L., & Kirschner, M. W. (2000). The fate of cells in the

- tailbud of *Xenopus laevis*. *Development*, 127(2), 255–267. <https://doi.org/10.1242/dev.127.2.255>
- Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, 51(6), 987–1000. [https://doi.org/10.1016/0092-8674\(87\)90585-X](https://doi.org/10.1016/0092-8674(87)90585-X)
- Domingues, A. F., Kulkarni, R., Giotopoulos, G., Gupta, S., Vinnenberg, L., Arede, L., Foerner, E., Khalil, M., Adao, R. R., Johns, A., Tan, S., Zeka, K., Huntly, B. J., Prabakaran, S., & Pina, C. (2020). Loss of KAT2A enhances transcriptional noise and depletes acute myeloid leukemia stem-like cells. *ELife*, 9, 1–29. <https://doi.org/10.7554/eLife.51754>
- Fischler, M. A., & Bolles, R. C. (1981). Random Sample Paradigm for Model Consensus: A Application to Image Fitting with Analysis and Automated Cartography. *Graphics and Image Processing*, 24(6), 381–395.
- Frith, T. J. R., Granata, I., Stout, E., Wind, M., & Thompson, O. (2018). *Human axial progenitors generate trunk neural crest cells*. <https://doi.org/10.1101/272591>
- Gillespie, D. T. (1996). The mathematics of Brownian motion and Johnson noise. *American Journal of Physics*, 64(3), 225–240. <https://doi.org/10.1119/1.18210>
- Gillespie, D. T. (2000). *The chemical Langevin equation The chemical Langevin equation*. 297(1), 297–306. <https://doi.org/10.1063/1.481811>
- Gont, L. K., Steinbeisser, H., Blumberg, B., & de Robertis, E. M. (1993). Tail formation as a continuation of gastrulation: the multiple cell populations of the *Xenopus* tailbud derive from the late blastopore lip. *Development (Cambridge, England)*, 119(4), 991–1004.
- Goto, H., Kimmey, S. C., Row, R. H., Matus, D. Q., & Martin, B. L. (2017). FGF and canonical Wnt signaling cooperate to induce paraxial mesoderm from tailbud neuromesodermal progenitors through regulation of a two-step epithelial to mesenchymal transition. *Development (Cambridge)*, 144(8), 1412–1421. <https://doi.org/10.1242/dev.143578>
- Gouti, M., Delile, J., Stamatakis, D., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V., & Briscoe, J. (2017). A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Developmental Cell*, 41(3), 243–261.e7. <https://doi.org/10.1016/j.devcel.2017.04.002>
- Goutsias, J., & Kim, S. (2004). A Nonlinear Discrete Dynamical Model for Transcriptional Regulation: Construction and Properties. *Biophysical Journal*, 86(4), 1922–1945. [https://doi.org/10.1016/S0006-3495\(04\)74257-5](https://doi.org/10.1016/S0006-3495(04)74257-5)
- Guillot, C., Djefal, Y., Michaut, A., Rabe, B., & Pourquie, O. (2021). Dynamics of primitive streak regression controls the fate of neuromesodermal progenitors in the chicken embryo. *ELife*, 10, 1–36. <https://doi.org/10.7554/elife.64819>
- Hackland, J. O. S., Shelar, P. B., Sandhu, N., Prasad, M. S., Charney, R. M., Gomez, G. A., Frith, T. J. R., & Garcia-Castro, M. I. (2019). FGF Modulates the Axial Identity of Trunk hPSC-Derived Neural Crest but Not the Cranial-Trunk Decision. *Stem Cell Reports*, 12(5), 920–933. <https://doi.org/10.1016/j.stemcr.2019.04.015>
- Halley, J. D., Burden, F. R., & Winkler, D. A. (2009). Stem cell decision making and critical-like exploratory networks. *Stem Cell Research*, 2(3), 165–177. <https://doi.org/10.1016/j.scr.2009.03.001>
- Hayashi, K., Lopes, S. M. C. de S., Tang, F., & Surani, M. A. (2008). Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell*, 3(4), 391–401. <https://doi.org/10.1016/j.stem.2008.07.027>
- Henrique, D., Abranches, E., Verrier, L., & Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development*, 142(17), 2864–2875. <https://doi.org/10.1242/dev.119768>
- Hu, M., Krause, D., Sharkies, S., Dexter, M., Heyworth, C., & Enver, T. (1997). Multilineage gene expression preceded commitment in the hemopoietic system. *Genes and Development*, 11, 774–785. <https://doi.org/10.1101/gad.11.6.774>
- Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23), 3853–3862. <http://dev.biologists.org/cgi/doi/10.1242/dev.035139>
- Kærn, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6(6), 451–464. <https://doi.org/10.1038/nrg1615>
- Kinney, B. A., Anber, A. Al, Row, R. H., Tseng, Y., Maxwell, D., Knaut, H., & Martin, B. L. (2020). *HHS Public Access*. 33(4). <https://doi.org/10.1016/j.celrep.2020.108311>. Sox2
- Kupiec, J. J. (1997). *A Darwinian theory for the origin of cellular differentiation*. 201–208.
- Li, Q., Wennborg, A., Aurell, E., Dekel, E., Zou, J.-Z., Xu, Y., Huang, S., & Ernberg, I. (2016). Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. *Proceedings of the National Academy of Sciences*, 113(10), 2672–2677. <https://doi.org/10.1073/pnas.1519210113>
- Li, W. M., Chan, C. M., Miller, A. L., & Lee, C. H. (2017). Dual functional roles of molecular beacon as a MicroRNA detector and inhibitor. *Journal of Biological Chemistry*, 292(9), 3568–3580. <https://doi.org/10.1074/jbc.M116.765776>
- Lukoseviciute, M., Mayes, S., & Sauka-Spengler, T. (2021). Neuromesodermal progenitor origin of trunk neural crest in vivo. *BioRxiv*, 2021.02.10.430513. <https://doi.org/10.1101/2021.02.10.430513>
- Lush, M. E., Diaz, D. C., Koenecke, N., Baek, S., Boldt, H., St Peter, M. K., Gaitan-Escudero, T., Romero-Carvajal, A., Busch-Nentwich, E. M., Perera, A. G., Hall, K. E., Peak, A., Haug, J. S., & Piotrowski, T. (2019). scRNA-Seq reveals distinct stem cell populations that drive hair cell regeneration after loss of Fgf and Notch signaling.

- ELife*, 8, 1–31.
<https://doi.org/10.7554/eLife.44431>
- Macarthur, B. D., & Lemischka, I. R. (2013). Statistical mechanics of pluripotency. *Cell*, 154(3), 484–489.
<https://doi.org/10.1016/j.cell.2013.07.024>
- Martin, B. L., & Kimelman, D. (2010). Brachyury establishes the embryonic mesodermal progenitor niche. *Genes & Development*, 24(24), 2778–2783.
<http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1962910>
- Martin, B. L., & Kimelman, D. (2012). Canonical Wnt Signaling Dynamically Controls Multiple Stem Cell Fate Decisions during Vertebrate Body Formation. *Developmental Cell*, 22(1), 223–232.
<https://doi.org/10.1016/j.devcel.2011.11.001>
- McLeod, C. M., & Mauck, R. L. (2017). On the origin and impact of mesenchymal stem cell heterogeneity: New insights and emerging tools for single cell analysis. *European Cells and Materials*, 34, 217–231.
<https://doi.org/10.22203/eCM.v034a14>
- Minelli, A., Pradeu, T., & Kupiec, J.-J. (2014). *Towards a Theory of Development. Edited Cell differentiation is a stochastic process subjected to natural selection.*
- Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J. C., & Reik, W. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Reports*, 20(5), 1215–1228.
<https://doi.org/10.1016/j.celrep.2017.07.009>
- Mojtahedi, M., Skupin, A., Zhou, J., Castaño, I. G., Leong-Quong, R. Y. Y., Chang, H., Trachana, K., Giuliani, J., & Huang, S. (2016). *Cell Fate Decision as High-Dimensional Critical State Transition.* 1–28.
<https://doi.org/10.1371/journal.pbio.2000640>
- Moris, N., Edri, S., Seyres, D., Kulkarni, R., Domingues, A. F., Balayo, T., Frontini, M., & Pina, C. (2018). Histone Acetyltransferase KAT2A Stabilizes Pluripotency with Control of Transcriptional Heterogeneity. *Stem Cells*, 36(12), 1828–1838.
<https://doi.org/10.1002/stem.2919>
- Moris, N., Pina, C., & Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*.
<https://doi.org/10.1038/nrg.2016.98>
- Moro, E., Ozhan-Kizil, G., Mongera, A., Beis, D., Wierzbicki, C., Young, R. M., Bournele, D., Domenichini, A., Valdivia, L. E., Lum, L., Chen, C., Amatruda, J. F., Tiso, N., Weidinger, G., & Argenton, F. (2012). In vivo Wnt signaling tracing through a transgenic biosensor fish reveals novel activity domains. *Developmental Biology*, 366(2), 327–340.
<https://doi.org/10.1016/j.ydbio.2012.03.023>
- Moussy, A., Cosette, J., Parmentier, R., da Silva, C., Corre, G., Richard, A., Gandrillon, O., Stockholm, D., & Páldi, A. (2017). Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. *PLoS Biology*, 15(7), 1–23.
<https://doi.org/10.1371/journal.pbio.2001867>
- Mukherjee, S., Heath, L., Preuss, C., Jayadev, S., Garden, G. A., Greenwood, A. K., Sieberts, S. K., De Jager, P. L., Ertekin-Taner, N., Carter, G. W., Mangravite, L. M., & Logsdon, B. A. (2020). Molecular estimation of neurodegeneration pseudotime in older brains. *Nature Communications*, 11(1), 1–12.
<https://doi.org/10.1038/s41467-020-19622-y>
- Muñoz-Descalzo, S., de Navascues, J., & Arias, A. M. (2012). Wnt-Notch signalling: An integrated mechanism regulating transitions between cell states. *BioEssays*, 34(2), 110–118.
<https://doi.org/10.1002/bies.201100102>
- Pacini, S. (2014). Deterministic and stochastic approaches in the clinical application of mesenchymal stromal cells (MSCs). *Frontiers in Cell and Developmental Biology*, 2(SEP), 1–13.
<https://doi.org/10.3389/fcell.2014.00050>
- Packer, J. S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., Waterston, R. H., & Murray, J. I. (2019). A lineage-resolved molecular atlas of *C. Elegans* embryogenesis at single-cell resolution. *Science*, 365(6459).
<https://doi.org/10.1126/science.aax1971>
- Paldi, A. (2020). Stochastic or deterministic? That is the question. *Organisms. Journal of Biological Sciences*, 4(1), 77–79.
<https://doi.org/10.13133/2532-5876/16963>
- Park, J., Zhou, Q. Y., & Koltun, V. (2017). Coloured Point Cloud Registration Revisited. *Proceedings of the IEEE International Conference on Computer Vision, 2017-Octob*, 143–152.
<https://doi.org/10.1109/ICCV.2017.25>
- Peláez, N., Gavalda-Mirallès, A., Wang, B., Navarro, H. T., Gudjonson, H., Rebay, I., Dinner, A. R., Katsaggelos, A. K., Amaral, L. A. N., & Carthew, R. W. (2015). Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *ELife*, 4(NOVEMBER2015), 1–23.
<https://doi.org/10.7554/eLife.08924>
- Petti, A. A., Khan, S. M., Xu, Z., Helton, N., Fronick, C. C., Fulton, R., Ramakrishnan, S. M., Nonavinkere Srivatsan, S., Heath, S. E., Westervelt, P., Payton, J. E., Walter, M. J., Link, D. C., DiPersio, J., Miller, C., & Ley, T. J. (2022). Genetic and Transcriptional Contributions to Relapse in Normal Karyotype Acute Myeloid Leukemia. *Blood Cancer Discovery*, 3(1), 32–49.
<https://doi.org/10.1158/2643-3230.bcd-21-0050>
- Pina, C., Fugazza, C., Tipping, A. J., Brown, J., Soneji, S., Teles, J., Peterson, C., & Enver, T. (2012). Inferring rules of lineage commitment in haematopoiesis. *Nature Publishing Group*, 14(3), 287–294.
<https://doi.org/10.1038/ncb2442>
- Richard, A., Boullu, L., Herbach, U., Bonnafoux, A., Morin, V., Vallin, E., Guillemin, A., Papili Gao, N., Gunawan, R., Cosette, J., Arnaud, O., Kupiec, J.-J., Espinasse, T., Sandrine, G.-G., & Olivier, G. (2016). Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biology*, *In press*, 1–35.
<https://doi.org/10.17605/OSF.IO>

- /K2Q5B
- Romanos, M., Allio, G., Roussigné, M., Combres, L., Escalas, N., Soula, C., Médevielle, F., Steventon, B., Trescases, A., & Bénazéraf, B. (2021). Cell-to-cell heterogeneity in *sox2* and *bra* expression guides progenitor motility and destiny. *eLife*, 10, 1–32. <https://doi.org/10.7554/eLife.66588>
- Row, R. H., Tsotras, S. R., Goto, H., & Martin, B. L. (2016). The zebrafish tailbud contains two independent populations of midline progenitor cells that maintain long-term germ layer plasticity and differentiate in response to local signaling cues. *Development*, 143(2), 244–254. <https://doi.org/10.1242/dev.129015>
- Rué, P., & Martinez Arias, A. (2015). Cell dynamics and gene expression control in tissue homeostasis and development [review]. *Molecular Systems Biology*, 11 VN-r(1), 792. <https://doi.org/10.15252/msb.20145549>
- Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. *Proceedings of International Conference on 3-D Digital Imaging and Modeling, 3DIM*, 145–152. <https://doi.org/10.1109/IM.2001.924423>
- Sambasivan, R., & Steventon, B. (2021). Neuromesodermal Progenitors: A Basis for Robust Axial Patterning in Development and Evolution. *Frontiers in Cell and Developmental Biology*, 8(January), 1–9. <https://doi.org/10.3389/fcell.2020.607516>
- Sauka-Spengler, T., & Bronner-Fraser, M. (2008). Evolution of the neural crest viewed from a gene regulatory perspective. *Genesis*, 46(11), 673–682. <https://doi.org/10.1002/dvg.20436>
- Schier, A. (2020). Single-cell biology: beyond the sum of its parts. *Nature Methods*, 17(January), 17–20.
- Semrau, S., Goldmann, J. E., Soumillon, M., Mikkelsen, T. S., Jaenisch, R., & Van Oudenaarden, A. (2017). Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-01076-4>
- Sha, Y., Haensel, D., Gutierrez, G., Du, H., Dai, X., & Nie, Q. (2019). Intermediate cell states in epithelial-to-mesenchymal transition. *Physical Biology*, 16(2). <https://doi.org/10.1088/1478-3975/aaf928>
- Smith, A. (2013). Nanog heterogeneity: Tilting at windmills? *Cell Stem Cell*, 13(1), 6–7. <https://doi.org/10.1016/j.stem.2013.06.016>
- Smith, R. C. G., Stumpf, P. S., Ridden, S. J., Sim, A., Filippi, S., Harrington, H. A., & MacArthur, B. D. (2017). Nanog Fluctuations in Embryonic Stem Cells Highlight the Problem of Measurement in Cell Biology. *Biophysical Journal*, 112(12), 2641–2652. <https://doi.org/10.1016/j.bpj.2017.05.005>
- Sorrells, S., Toruno, C., Stewart, R. A., & Jette, C. (2013). Analysis of apoptosis in zebrafish embryos by whole-mount immunofluorescence to detect activated caspase 3. *Journal of Visualized Experiments*, 82, 1–8. <https://doi.org/10.3791/51060>
- Steventon, B., Duarte, F., Lagadec, R., Mazan, S., Nicolas, J.-F., & Hirsinger, E. (2016). Species-specific contribution of volumetric growth and tissue convergence to posterior body elongation in vertebrates. *Development (Cambridge)*, 143(10). <https://doi.org/10.1242/dev.126375>
- Steventon, Ben, & Arias, A. M. (2017). Evo-engineering and the cellular and molecular origins of the vertebrate spinal cord. *Developmental Biology*, 432(1), 3–13. <https://doi.org/10.1016/j.ydbio.2017.01.021>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>
- Stumpf, P. S., Smith, R. C. G., Lenz, M., Howison, S. D., Arai, F., & MacArthur, B. D. (2017). Stem Cell Differentiation as a Non-Markov Stochastic Process. *Cell Systems*, 268–282. <https://doi.org/10.1016/j.cels.2017.08.009>
- Tanay, A., & Regev, A. (2018). *Single Cell Genomics: from phenomonology to mechanism*. 541(7637), 331–338. <https://doi.org/10.1038/nature21350>. Single
- Taniguchi, Y., Kurth, T., Weiche, S., Reichelt, S., Tazaki, A., Perike, S., Kappert, V., & Epperlein, H. H. (2017). The posterior neural plate in axolotl gives rise to neural tube or turns anteriorly to form somites of the tail and posterior trunk. *Developmental Biology*, 422(2), 155–170. <https://doi.org/10.1016/j.ydbio.2016.12.023>
- Teschendorff, A. E., & Feinberg, A. P. (2021). Statistical mechanics meets single-cell biology. *Nature Reviews. Genetics*, 0123456789. <https://doi.org/10.1038/s41576-021-00341-z>
- Tritschler, S., Büttner, M., Fischer, D. S., Lange, M., Bergen, V., Lickert, H., & Theis, F. J. (2019). Concepts and limitations for learning developmental trajectories from single cell genomics. *Development (Cambridge)*, 146(12). <https://doi.org/10.1242/dev.170506>
- Trott, J., Hayashi, K., Surani, A., Babu, M. M., & Martinez Arias, A. (2012). Dissecting ensemble networks in ES cell populations reveals micro-heterogeneity underlying pluripotency. *Molecular BioSystems*, 8(3), 744–752. <http://pubs.rsc.org/en/content/articlehtml/2012/mb/c1mb05398a>
- Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V., & Nicolas, J.-F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Developmental Cell*, 17(3), 365–376. <http://www.sciencedirect.com/science/article/pii/S1534580709003396>
- Verd, B., Clark, E., Wotton, K. R., Janssens, H., Jiménez-Guri, E., Crombach, A., & Jaeger, J. (2018). A damped oscillator imposes temporal order on posterior gap gene expression in *Drosophila*. *PLoS Biology*, 16(2), 1–24. <https://doi.org/10.1371/journal.pbio.2003174>
- Verd, B., Crombach, A., & Jaeger, J. (2014). Classification of

- transient behaviours in a time-dependent toggle switch model. *BMC Systems Biology*, 8(1), 43. <https://doi.org/10.1186/1752-0509-8-43>
- Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., & Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, 4362(April), eaar4362. <https://doi.org/10.1126/science.aar4362>
- Wardle, F. C., & Smith, J. C. (2004). Refinement of gene expression patterns in the early *Xenopus* embryo. *Development*, 131(19), 4687–4696. <https://doi.org/10.1242/dev.01340>
- Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M., & Klein, A. M. (2018). Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences of the United States of America*, 115(10), E2467–E2476. <https://doi.org/10.1073/pnas.1714723115>
- Westerich, K. J., Chandrasekaran, K. S., Gross-Thebing, T., Kueck, N., Raz, E., & Rentmeister, A. (2020). Bioorthogonal mRNA labeling at the poly(A) tail for imaging localization and dynamics in live zebrafish embryos. *Chemical Science*, 11(11), 3089–3095. <https://doi.org/10.1039/c9sc05981d>
- Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2), 307–319. <https://doi.org/10.1016/j.cell.2013.03.035>
- Wiesner, K., Teles, J., Hartnor, M., & Peterson, C. (2017). *Haematopoietic stem cells -- entropic landscapes of differentiation*. 1–11. <http://arxiv.org/abs/1712.07870>
- Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L., & Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, 20(1), 1–9. <https://doi.org/10.1186/s13059-019-1663-x>
- Wood, T. R., Kyrsting, A., Stegmaier, J., Kucinski, I., Kaminski, C. F., Mikut, R., & Voiculescu, O. (2019). Tissue Mechanics Determines Cell Fate in the Axial Stem Zones. *BioRxiv*, 1–30. <https://doi.org/https://doi.org/10.1101/622571>
- Wymeersch, F. J., Huang, Y., Blin, G., Wilkie, R., Wong, F. C. K., & Wilson, V. (2016). *Position-dependent plasticity of distinct progenitor types in the primitive streak*. 1–28. <https://doi.org/10.7554/eLife.10042>
- Wymeersch, F. J., Wilson, V., & Tsakiridis, A. (2021). Understanding axial progenitor biology in vivo and in vitro. *Development (Cambridge)*, 148(4). <https://doi.org/10.1242/dev.180612>
- Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., Dedhar, S., Derynck, R., Ford, H. L., Fuxe, J., García de Herreros, A., Goodall, G. J., Hadjantonakis, A. K., Huang, R. J. Y., Kalchauer, C., ... Sheng, G. (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, 21(6), 341–352. <https://doi.org/10.1038/s41580-020-0237-9>
- Ye, Z., & Kimelman, D. (2020). Hox13 genes are required for mesoderm formation and axis elongation during early zebrafish development. *Development (Cambridge)*, 147(22). <https://doi.org/10.1242/dev.185298>
- Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J. N., Beck, F., Mallo, M., & Deschamps, J. (2009). Cdx and Hox Genes Differentially Regulate Posterior Axial Growth in Mammalian Embryos. *Developmental Cell*, 17(4), 516–526. <https://doi.org/10.1016/j.devcel.2009.08.010>
- Zhang, L., Kendrick, C., Jülich, D., & Holley, S. A. (2008). Cell cycle progression is required for zebrafish somite morphogenesis but not segmentation clock function. *Development*, 135(12), 2065–2070. <https://doi.org/10.1242/dev.022673>
- Zhou, Q.-Y., Park, J., & Koltun, V. (2018). *Open3D: A Modern Library for 3D Data Processing*. <http://arxiv.org/abs/1801.09847>

Declaration of interests

The authors declare no competing interests.

Limitations of the study

The coloured ICP (cICP) algorithm employed in ZebReg will not be able to align point clouds exactly, as zebrafish tailbuds will inevitably differ from one another in their nuclei position and gene expression intensities. Instead, for each nucleus from the source image, ZebReg can, at best, map it to its most similar cell counterpart in the target image, based on their proximity to each other and similarity in expression of a reference gene. For multiply mapped and unmapped target points, ZebReg imputes their gene expression intensities by taking the average intensities of each point's k -nearest neighbors ($k=5$). This approach assumes a degree of spatial autocorrelation in gene expression intensities of a point with its neighbours. During the transition state where cell-cell correlation decreases, our approach may underestimate the extent of cellular heterogeneity in the population due to the application of an averaging procedure.

In our work, we adopted a descriptive, fixed imaging-based approach towards interrogating the level of gene expression heterogeneities during NMP differentiation. Whilst the peculiarities of the zebrafish NMP model enabled us to infer cellular fates from cellular positions, without adopting a live imaging approach, we were unable to document the details of the transcriptional dynamics around 24ss (Weinreb et al., 2017). Zebrafish embryos have been amenable to live RNA imaging using various techniques such as the MS2 labelling system (Campbell et al., 2015), 3' poly(A) tail labelling system (Westerich et al., 2020) and molecular beacon sensors (Li et al., 2017) due to its optical transparency. Thus, future work could perform live imaging of *sox2* and *tbxta* mRNAs to monitor the changes in transcription dynamics around 24ss as NMPs enter and exit the transition state.

STAR Methods

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal Anti-Histone H3 (phosphor S10)	Abcam	Ab14955
Rabbit polyclonal Anti-caspase3	Abcam	Ab13847
Secondary anti-mouse Alexa Fluor A488	ThermoFisher Scientific	A32723
Secondary anti-rabbit Alexa Fluor A633	Invitrogen	A21071
Chemicals		
4% paraformaldehyde (PFA)	Sigma	CAS no: 30525-89-4
Agarose, low gelling temperature	Sigma	A9414
Dulbecco's Phosphate Buffered Saline (PBS)	Sigma	D8537-500ML
DAPI	Sigma	CAS no: 28718-90-3
EDTA	Sigma	CAS no: 60-00-4
Triton-X	Sigma	CAS no:9002-93
Fetal bovine serum (heat-inactivated)	ThermoFisher Scientific	10437028
Bovine serum Albumin	Sigma	A7906-10G
Methylcellulose	Sigma	M0512
RNaseA	QIAGEN	19101
SSC buffer	Scientific Laboratory Supplies	S6639-1L
Methanol	ThermoFisher Scientific	CAS no: 67-56-1
VECTASHIELD Antifade mounting medium	Vector Laboratories	H-1000-10
Tween-20	ThermoFisher Scientific	AAJ20605A
Crucial Commercial assays		
In situ HCR v3.0	Molecular Instruments	N/A
QIAquick PCR purification kit	Qiagen	28104
SP6 mMessage mMachine kit	Invitrogen	AM1340
Plasmid		

Hsp70l:p2a-NLS kikGR	Benjamin Martin lab	ZDB-TGCONSTRCT-160321-3
HCR probes		
<i>Sox2</i>	This paper	Supplementary Information
<i>Tbxta</i>	This paper	Supplementary Information
Other		
NanoDrop 2000c Spectrophotometer	Thermo Scientific	13400411
35 mm glass bottom dish	MatTek	P35G-1.5-10-C
Inverted confocal Microscope	Leica	SP8
Inverted confocal Microscope	Zeiss	LSM700
KpnI-HF enzyme	NEB	R3142L

Resource Availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Ben Steventon (bjs57@cam.ac.uk).

Materials availability

HCR probe sequences for *sox2* and *tbxta* are documented in the supplementary file.

Data and code availability

- Code for establishing ZebReg can be found here: <https://github.com/DillanSaunders/ZebReg>
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experiment Model and Subject Details

Zebrafish husbandry

All zebrafish procedures were conducted under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012, following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB). Wild type lines used are either Tüpfel long fin (TL), AB/TL or AB. The Tg(7xTCF-Xla.Sia:GFP) reporter line (Moro et al., 2012) was provided by the Steven Wilson laboratory. All embryos obtained were obtained and raised in standard E3 media at 28°C. Embryos were staged according to Kimmel et al., 1995.

Method Details

Version 3 Hybridisation Chain Reaction (V3 HCR)

Zebrafish embryos at the required stages were fixed in 4% PFA in DEPC-treated, calcium and magnesium-free PBS at 4°C overnight. Embryos were then stained with V3 HCR (Choi et al., 2018). All hairpins were purchased from Molecular Instruments. All probes were purchased from Molecular Instruments except for *sox2* and *tbxta* which were manually designed. After the staining procedure, samples were counterstained with DAPI at a dilution of 1:1000 in 5xSSCT for 2 hours at room temperature. The tailbud region was cut out with a forceps and eyelash tool, and then mounted on a 35mm glass bottom dish (MatTek) with the Vectashield antifade mounting medium for confocal imaging.

Quantification of nuclear gene expression intensities in NMps

HCR images were processed in Imaris (Bitplane). Unless otherwise stated, all *sox2+tbxta+* HCR images were analysed for the number of NMps as described in Figure S1.

Immunostaining with anti-PH3 and anti-caspase 3 antibodies

Zebrafish embryos at the required stages were fixed in 4% PFA in DEPC-treated, calcium and magnesium-free PBS at 4°C overnight. Embryos were then co-stained with a 1:500 dilution of mouse anti-PH3 antibody (abcam, ab14955) and 1:500 dilution of rabbit anti-caspase3 antibody (abcam, ab13847), as described in Sorrells et al., 2013. Secondary anti-mouse Alexa Fluor 488-conjugated antibody and anti-rabbit Alexa Fluor 647-conjugated antibody were both diluted in 1:500 PDT solution and incubated with the samples overnight at 4°C. DAPI was added at the final step with a 1:1000 dilution in PDT and incubated for 2 hours at room temperature for nuclear detection. Images were quantified in the 3/4D Image Visualisation and Analysis Software Imaris 9.2.1 (Bitplane). The percentages of mitotic or apoptotic cells for each sample were calculated as the fraction of PH3+ or caspase3+ nuclei over the total number of nuclei in the tailbud, multiplied by 100.

Photolabeling with nuclear-targeted kikume

The hsp70l:p2a-NLS kikGR vector (Bouldin et al., 2015) was extracted from an overnight grown bacterial culture. Briefly, bacterial cells were collected via centrifugation and washed sequentially with the following 3 buffers: P1 containing 50 mM Tris-Cl at pH 8.2, 10 mM EDTA at pH 8.0, RNase A (QIAGEN); P2 (filter-sterilised) containing 0.8% NaOH and 1% SDS; P3 containing 3M KOAc that is adjusted to pH 5.5 with glacial acetic acid. Plasmid DNA was precipitated with 70% isopropanol and washed with 70% ethanol before resuspension in nuclease-free water.

The vector was linearised by restriction digestion with the KpnI-HF enzyme (NEB), and subsequently purified using the QIAquick PCR purification kit (Qiagen). The purified, linearised plasmid was transcribed at the SP6 promoter with the SP6 mMessage mMachine kit (Invitrogen), and lithium chloride precipitation was carried out for mRNA recovery. Quantification of the transcribed kikGR mRNA was performed on the NanoDrop instrument (ThermoFisher).

One-cell stage zebrafish embryos were injected with the NLS-kikGR mRNA and then embedded in low gelling point agarose (Sigma) at 1% w/v in E3 media at the bottom of a MatTek 35mm glass bottom dish. Photoconversion and image acquisition was performed on a Zeiss LSM 700 confocal microscope. Efficient, irreversible photoconversion of NLS-KikGR in the zebrafish embryos at mid-somitogenesis stages was carried out by scanning the 405 nm laser at 15% laser power for approximately 30 seconds in a region of interest.

Confocal microscopy imaging

Samples were imaged on either a Zeiss LSM700 inverted confocal or a Leica TCS SP8 inverted confocal at 10X, 20X or 40X magnification.

Analysis of scRNA-seq data

Downloading the 18hpf scRNA-seq dataset and preprocessing

The wild type 18hpf zebrafish scRNA-seq raw counts dataset and the associated clusterIDs were downloaded from GEO with the accession number GSM3067194 (Wagner et al., 2018). First, outlier cells with log-transformed library and feature sizes more than 3 median absolute deviations (MADs) from the respective median metric values were removed. Genes that were not expressed in the dataset were filtered out. At this quality control threshold, most genes and cells were retained for downstream analysis, resulting in a dataset with 30296 genes x 6954 cells (381 genes and 8 cells discarded). The data was then converted into a Seurat 3.0 object (Stuart et al., 2019) for subsequent analyses. Cell cycle scoring and regression were performed in Seurat 3.0 using a set of cell-cycle associated genes for zebrafish (Lush et al., 2019), with the S.Score and G2M.Score as inputs to the vars.to.regress argument in the *SCTransform* function. Data normalisation, scaling and the identification of the top 3000 most variable genes were also carried out using the *SCTransform* wrapper.

Low dimensional embedding and Louvain clustering

The normalised and scaled data was projected into low dimensional subspace via principal components analysis (PCA) with default settings for the *RunPCA* function. (Figure 2.2B). Following this, the uniform manifold approximation and projection

(UMAP) embedding was implemented via the *RunUMAP* function. To perform clustering, groups of similar cells on the UMAP embedding were identified by generating a shared nearest neighbor (SNN) graph of the dataset with the *FindNeighbors* function, and then clustered using the Louvain algorithm with the *FindClusters* function at various resolutions. Subclustering on the tailbud cells was performed in similar manner to the above clustering procedure, with a resolution of 1 set for the *FindClusters* function. To examine the clustering results, clustering trees were plotted with the *clustree* package whilst the adjusted rand index and clustering entropy were implemented in the *mclust* and *NMF* packages respectively.

Identification of differentially expressed genes

For each cluster, supervised annotation was carried out by examining the marker genes identified by a Model-based Analysis of Single-cell Transcriptomics (MAST) and a Wilcoxon Rank-Sum test. The tests were carried out using the *FindAllMarkers* function in Seurat that compares cells in each cluster against all other remaining clusters. The function is set to return only positive markers for each cluster (*only.pos* = TRUE). Differentially expressed genes with an adjusted p-value less than 0.05 were retained for analysis. They were then sorted in order of priority, based on the log fold-change of the average expression between the cluster under study and the remaining 7 tailbud subclusters (*avg_log2FC*).

Robustness analysis of tailbud clustering assignments

To assess the robustness of our selection of the zebrafish tailbud cells from the 18hpf dataset, we employed a different approach than Wagner and colleagues (Wagner et al., 2018) by embedding the 6,954 cells in the 18hpf dataset into a Uniform Manifold Approximation and Projection (UMAP) space and using the Louvain community detection algorithm to identify clusters (Figure S3A).

We first assessed the similarity between the two data clusterings using the Adjusted Rand index (ARI) and clustering entropy index. High ARI values and low entropy values are obtained across a wide range of clustering resolutions, apart from the initial resolution of 0.2 (Figure S4A,C). In addition, analysis of the clustering tree shows that at a resolution of 0.2, there are 11 clusters which continue to be split up gradually. At increasing resolutions, the number of in-proportion edges (edges with low transparency) remain low which indicates only minor changes in the clustering tree. At a clustering resolution of 0.8, we obtained 22 clusters (Figure S4B). When we re-examined the distribution of our tailbud labels against Wagner and colleagues' labels, we find that they are highly concordant (Figure S3B), suggesting that our selection of the zebrafish tailbud cells are robust across different analytical strategies. As the Louvain algorithm is stochastic, we re-ran the algorithm for 10

iterations and retained cells that are consistently located in the tailbud clusters for 9 and 10 iterations for downstream analyses (Figure S4D).

Computation of the critical index and transcriptional noise index

The critical index is defined as the ratio of two averaged Pearson correlation coefficients: the average correlations between all pairs of gene vectors over the average correlations between all pairs of cell state vectors (Mojtahedi et al., 2016). In the scRNA-seq analysis, to account for the differences in cell number between clusters, 200 cells from each cluster were randomly sampled with replacement to calculate the index, and the procedure was repeated for 10,000 times. We also assessed the robustness of the critical index to differences in cell number and number of marker genes used (Figure S5).

The transcriptional noise index was measured using the top 2000 highly variable genes of each cluster following the work of Mohammed and colleagues (Mohammed et al., 2017).

Tailbud image registration with ZebReg

Overview of pipeline

ZebReg is a 3D, non-landmark-based image registration Python tool which we developed to integrate cellular position and nuclear gene-expression information from confocal images of zebrafish tailbuds. Leveraging upon the open-source open3D library (Zhou et al., 2018), ZebReg implements a set of rigid body, point-based registration algorithms that are popular in the field of geometric registration to align a 3D point cloud (source cloud) into a reference point cloud (target cloud).

At present, we have tested ZebReg on zebrafish tailbuds ranging from 18ss to 30ss. Briefly, confocal images were first preprocessed in Imaris to obtain segmented DAPI-stained nuclei (Figure S1). Next, to ensure a consistent field of view, all nuclei posterior to the tip of the developing notochord for all the images were retained for analysis. ZebReg performs the alignment by first importing the 3D centroid coordinates and gene expression intensities (if present) of the segmented nuclei and converting each image into a point cloud. Then, given a set of source clouds and a reference point cloud (target cloud), ZebReg finds the best linear transformation (no shearing, stretching or other deformations) between each source cloud and the target cloud. Additionally, if colour intensities of the source clouds are provided, ZebReg can map them onto the target cloud by imputing the gene expression intensities in the target point cloud and thus generate a composite image (See Figure S7)

Imputation of gene expression intensities

There are three possible sets of outcomes during the imputation procedure:

- i) First, the mapping of the source point to the target point may be unique, in which case the target point simply adopts the intensity value of the corresponding source point.
- ii) In cases where there is not a single source point corresponding to the target point, ZebReg provides the user with several options to resolve the discrepancy. If multiple source points map to the same target point, the target point adopts either the mean or median of these source intensity values (default: 'median').
- iii) Alternatively, if there is no source point that corresponds to the target point, ZebReg provides three options to impute the gene expression intensity of this target point: 'null', 'complete' or 'knn' (default : 'knn'). 'null' sets the intensity of the target point to 0, whilst 'complete' can be used if the target point cloud already has an intensity channel for that gene, in which case the point simply retains the original target intensity value. In the default 'knn' case, regression is performed based on the k-nearest neighbors of the point (default : $n = 5$) as implemented in the *sklearn* package. The target point takes on the mean intensity value of the closest target points.

Notably, in cases ii) and iii), ZebReg imputes the expression intensities of the target points by borrowing information from multiple source or neighboring target points.

Point set registration algorithms

To conduct the image alignment, ZebReg employs the following point set registration algorithms:

- i) Random Sample Consensus (RANSAC)

The RANSAC algorithm is a non-deterministic global alignment algorithm that is used in ZebReg to provide the initial coarse alignment for the ICP and cICP local algorithms (Fischler & Bolles, 1981).

- ii) Iterative Closest Point (ICP)

In the vanilla ICP algorithm, the algorithm repeatedly updates the transformation required to map the source to target cloud by minimizing the distance between points (Besl & McKay, 1992). In ZebReg, we use the point-to-plane ICP variant due to its increased speed of convergence (Rusinkiewicz & Levoy, 2001).

- iii) Coloured Iterative Closest Point (cICP)

For images with a colour channel in common, it is advantageous to consider their colour on top of geometry during point set registration. In these cases, ZebReg uses

cICP, a modified version of ICP implemented in open3D, which optimises a joint geometric and photometric objective (Park et al., 2017).

ZebReg carries out all alignments by first performing a coarse global alignment with RANSAC, followed by either a finer alignment with ICP (if no colour channel is supplied) or cICP (if a common colour channel is present in the source and target image).

In silico validation of ZebReg

First, we constructed a mean absolute error (MAE) metric which quantifies the average difference in normalised signal intensities of the shared colour channel between the source and target image pair after image registration. To assess the accuracy of ZebReg's image alignment with the cICP algorithm, we selected a source and target point cloud of the zebrafish tailbud and used the MAE as the test statistic in a permutation test which tests the following hypotheses:

H_0 : ZebReg cICP registration has no effect on the colour intensity residuals between source and target clouds.

H_1 : ZebReg cICP registration reduces the colour intensity residuals between source and target clouds.

In the permutation test, the sampling distribution under the null hypothesis was constructed by randomly rearranging the order of the target colour array, and then calculating the MAE using the permuted and original target colour arrays over 10,000 iterations. In effect, the null distribution provides the range of MAE estimates under the condition where ZebReg's reported correspondence mapping between the source and target colour arrays is random. The null distribution was then fit to a gaussian distribution for the computation of the 95% confidence interval (Figure S9A).

Next, to assess the effectiveness of the various point cloud registration algorithms, we registered a point cloud with its rotated counterpart using 3 algorithms that are implemented in open3D: RANSAC, ICP and cICP, and assessed whether they can successfully recover the correspondence map. (Figure S9B)

We then compared the cICP's performance across the different datasets. The datasets we have chosen for comparison were (Figure S9C-D):

- *Test sample*: Images of two separate zebrafish tailbuds at 18ss. Test sample exemplifies the performance of the algorithm on an actual use case in practice.

- *Lateral halves*: Images of two lateral halves of a single 18ss tailbud image. Since the point clouds do not overlap, any correspondence between the points in Lateral halves are spurious.
- *AP*: Images of the anterior and posterior ends of a single 18ss tailbud image. Like Lateral halves, any correspondence found between points in AP are spurious.

In the absence of ground truth data or an alternative image registration method, to achieve a better grasp of ZebReg's performance, we benchmarked the registration results of these three datasets onto a noise calibration curve, which we obtained from registering noise-shifted versions of the source cloud onto its original copy.

Zero-mean gaussian distributions with standard deviations ranging from 0 to 30 were sampled to construct an array of noise matrices. These noise matrices were added to the positions of the source clouds to generate an array of noise-shifted point clouds. Conceptually, each noise-shifted point cloud is an *in silico* analogue of a tailbud that differs from its idealised, identical twin in nuclei position by a prespecified level of noise. To construct the noise-calibration curve, all noise-shifted point clouds were registered against the original source cloud, which returned the values for the fitness, inlier RMSE and inlier MAE metrics. For the heavily noise-shifted point clouds, many points are classified as outliers and therefore, the inlier metrics overestimate the registration quality by omitting these points. To correct this, we scaled the inlier RMSE and inlier MAE metrics by the corresponding fitness and plotted the scaled inlier RMSE and scaled inlier MAE values instead. Comparing our results (Figure S9E-F), we conclude that ZebReg's registration of Test Sample outperforms the Lateral halves and AP datasets and returns acceptable fitness, inlier RMSE and MAE scores in its alignment of the Test Sample dataset.

Validation of ZebReg against HCR data

To assess whether the imputation procedure alters the gene expression distributions, we constructed Q-Q plots of the original and imputed gene expression distributions of 12 genes at 18ss (Figure S10A). For the purposes of the comparison of Q-Q plots, we also analysed the expression of four additional genes (*wnt8a*, *thbs2*, *id3* and *depdc7a*) that were not used in constructing the composite maps.

We also assessed the extent to which ZebReg maintains the quantitative relationships between genes in the NMps by comparing the pairwise linear correlations of imputed genes with the original correlations from the HCR datasets (Figure S10B). A total of 17 gene pairs were compared. As a measure of how close the original and imputed correlations are to each other, we computed the minimum difference between the imputed correlation and the associated original correlations (Figure S10C). The minimum difference was computed by calculating the differences

between the correlations obtained from the HCR images and the correlation from the composite map, and then taking the minimum value of the differences.

Construction of *in silico* composite maps

To construct the composite maps for each stage, we first selected images across different samples to be used for the imputation. Each image consists of *sox2* and *tbxta* stained alongside one or two additional genes and belongs to an image group. Specifically, there are a total of five image groups that correspond to particular HCR experiments (Table S2): STT (*sox2*, *tbxta*, *tcf*), STHC (*sox2*, *tbxta*, *hes6*, *cdh6*), STSC (*sox2*, *tbxta*, *sp5l*, *cdh6*), STTC (*sox2*, *tbxta*, *tagln3b*, *cdh6*) and STZC (*sox2*, *tbxta*, *znf703*, *cdh6*). For each of the three composite maps (18ss, 24ss, 28ss), five images of the same stage, one from each of the five image groups, were mapped onto a chosen target image using *sox2* as the common colour channel for cICP alignment (Figure S8). These six images for each composite map were chosen to best reflect the number and spatial distributions of the *in silico* NMps in the resultant composite maps. In summary, each composite map combines information across six images (one for the target image and five for the source images) to generate an eight-dimensional (*sox2*, *tbxta*, *cdh6*, *hes6*, *sp5l*, *tagln3b*, *tcf*, *znf703*) point cloud image.

Analysis of the composite maps

Identification of *in silico* NMps

Following the construction of the composite map with ZebReg, imputed intensity values of all genes below the 0.7 quantile threshold were set to 0 and rescaled by min-max normalization (Figure S12A). Amongst the *sox2+tbxta+* cells that were identified in the composite map, most were found within the approximate NMP spatial regions (Figure S12Bi-ii). Of the *sox2+tbxta+* points that reside outside of the NMP regions and are thus excluded from being NMps, they fall into two groups (Figure S12Bii). The first group corresponds to the hypochord cells that constitute the bulk of these *sox2+tbxta+* non-NMps in the 18ss (47/146: 32%), 24ss (43/203: 21%) and 28ss (29/56: 52%) composite maps. The second group of cells are found in the 18ss composite map only (17/146 : 11%) and are a small population of aberrant cells that have likely arisen from technical errors in either ZebReg's alignment or mapping procedure. These cells flank the hypochord and floor plate and thus may have been mistakenly assigned above-background levels of *sox2* due to their proximity to these two *sox2*-expressing structures. After the removal of these two groups of cells, the resultant 78, 183 and 27 *in silico sox2+tbxta+* cells in the 18ss, 24ss and 28ss composite maps are defined as the *in silico* composite map NMps. (Figure S12Biii)

Construction of the Neural-Mesodermal (NM) Index

The neural-mesodermal index, NM_j for the j^{th} cell is defined as:

$$NM_j = N_j - M_j \quad \text{Equation 1}$$

where NM_j , N_j , M_j are the neural-mesodermal index, neural index and mesodermal index of the j^{th} cell, respectively, and $j = 1, 2, \dots, C$ for a total of C NMps.

The neural index, N_j , for the j^{th} cell is defined as:

$$N_j = sox2_j + \sum_{k=1}^G \widetilde{\rho}_k (1 - \epsilon_k) Gene_{kj} \quad \text{Equation 2}$$

where $Gene_{kj}$ is the min-max normalised expression intensity of the k^{th} gene in the j^{th} cell; $\widetilde{\rho}_k$ is the median of the Pearson's correlation coefficients of $Gene_k$ and $sox2$, computed from the NMps segmented from all the HCR images of the same somite stage; ϵ_k is the interquartile range of $Gene_k$'s correlation coefficients. The $(1 - \epsilon_k)$ term penalises $Gene_k$'s contribution to the neural index if it displays large variability in its correlation coefficients between all the HCR images of that somite stage. The summation is applied to all the genes, G , minus $tbxta$ and $sox2$. The total number of genes is $G + 2$.

The mesodermal index, M , is defined symmetrically but with $tbxta$ replacing $sox2$ and the correlation coefficients calculated with respect to $tbxta$ instead. We further verified that the NM index provides a sensible summary of the NMp's neural/mesodermal potential (Figure S13B-D) and is not systematically biased towards either neural or mesodermal indices (Figure S13E-E').

Construction of the NMp probability map

Tailbud images (source images) were aligned to an arbitrarily chosen target image tailbud. The NMp nuclei in the source images were pre-segmented prior to alignment in Imaris (Bitplane) and hence, it is possible to keep track of the number of times each target cell receives a mapping from a source NMp cell. Target cells with a large count number is assigned a high probability of being an NMp. For visualisation purposes, in my probability maps, we displayed only target cells with a minimum count number of 2 for each probability map (Figure S14).

Estimation of the standard error of empirical entropy

The standard error was estimated by the leave-one jackknife resampling method and is implemented using the R *bootstrap* package (Efron and Tibshirani, 1993; Wiesner et al., 2017). In this method, the entropy was repeatedly estimated but with one of the data points randomly removed during each computation.

Dynamical systems modelling of a toggle switch model

Deterministic model formulation

The deterministic model consists of two continuous ODEs describing the rate of change of protein concentrations (Figure 4E, Figure S14). As shown in Equation 3 and Equation 4, the regulatory interactions between genes are summarized by Hill functions, whereas the decay rate of each product is set to be linearly proportional to the protein concentration.

$$\frac{dx}{dt} = \left(\alpha_1 + \frac{\max_1 x^n}{a^n + x^n} \right) \left(\frac{b^n}{b^n + y^n} \right) - \lambda_1 x \quad \text{Equation 3}$$

$$\frac{dy}{dt} = \left(\alpha_2 + \frac{\max_2 y^n}{c^n + y^n} \right) \left(\frac{d^n}{d^n + x^n} \right) - \lambda_2 y \quad \text{Equation 4}$$

where:

- x and y are the concentrations of the two proteins X and Y.
- α_1 and α_2 are the rates of production of gene X/Y, in the absence of activation by X/Y.
- n is the Hill coefficient and is set to 4 for all simulations.
- a and c are the concentrations for the half-maximal activation by genes X/Y on genes X/Y. The inverses, $\frac{1}{a}$ and $\frac{1}{c}$ represents the efficiency of the activator in other equivalent formulations (Goutsias & Kim, 2004).
- b and d are the concentrations for the half-maximal repression by genes X/Y on genes Y/X. The inverses, $\frac{1}{b}$ and $\frac{1}{d}$ represents the efficiency of the repressor in other equivalent formulations.
- λ_1 and λ_2 are the protein degradation rates.
- $\max_1$ and $\max_2$ are the maximum rates of protein production by activators X / Y. To see this, take $\lim_{a \rightarrow 0} \frac{\max_1 x^n}{a^n + x^n}$.

Classification of stability of steady states

The stability of the steady states was determined by linear stability analysis (Strogatz, 2000). Briefly, the Jacobian matrix for the ordinary differential equations (Equation 3 and Equation 4) was constructed to obtain an approximate description of the phase portrait around the fixed points. By neglecting the quadratic and higher order terms in the Taylor series expansion, one can describe the dynamics of the linearised system with the Jacobian matrix. To assess the stability of the fixed points, we computed the determinant of the Jacobian at each equilibrium point. A negative determinant corresponds to a saddle, while a positive determinant corresponds to a

sink (stable node), source (unstable node) or center. In our model, all equilibria with positive determinants correspond to sinks.

Stochastic toggle switch model formulation

As per the deterministic formulations, the rate of production of X and Y can be written as:

$$\rho_x = \left(\alpha_1 + \frac{max_1 x^n}{a^n + x^n} \right) \left(\frac{b^n}{b^n + y^n} \right) \quad \text{Equation 5}$$

$$\rho_y = \left(\alpha_2 + \frac{max_2 y^n}{c^n + y^n} \right) \left(\frac{d^n}{d^n + x^n} \right) \quad \text{Equation 6}$$

The chemical langevin equation (CLE) approximation for the process leads to the following white noise form of the Langevin equation:

$$\frac{dx}{dt} = \rho_x - \lambda_1 x + (\sqrt{\rho_x + \lambda_1 x}) \Gamma_x(t) \quad \text{Equation 7}$$

$$\frac{dy}{dt} = \rho_y - \lambda_2 y + (\sqrt{\rho_y + \lambda_2 y}) \Gamma_y(t) \quad \text{Equation 8}$$

Here, the $\Gamma_i(t)$ terms correspond to Gaussian white noises, which are formally defined as (Gillespie, 1996, 2000):

$$\Gamma_i(t) \equiv \lim_{dt \rightarrow 0} N\left(0, \frac{1}{dt}\right) \quad \text{Equation 9}$$

The two averaged properties of Gaussian white noise processes are:

$$\langle \Gamma_i(t) \rangle = 0 \quad \text{Equation 10}$$

$$\langle \Gamma_i(t) \Gamma_j(t') \rangle = \delta_{ij} \delta(t - t') \quad \text{Equation 11}$$

The first property indicates that the white noise process has zero mean. In the second, the first delta function is Kronecker's delta and the second is Dirac's delta, which indicates that white noise processes are statistically independent and that the white noise process is temporally uncorrelated.

Numerical simulation with the Euler-Maruyama method

The standard form of the Langevin equation is written as the following difference equation, where Ω is the volume parameter (Gillespie, 2000):

$$X_{t+dt} = X_t + (\rho_x - \lambda x)dt + (\sqrt{\rho_x + \lambda x})N(0,1)\sqrt{\frac{dt}{\Omega}} \quad \text{Equation 12}$$

In the limit $\Omega \rightarrow \infty$, we obtain the deterministic result.

Numerical simulation of the CLE was carried out with the Euler-Maruyama algorithm, which has the following numerical scheme that approximates the infinitesimal dt with finite Δt (Wilkinson, 2020):

$$X_{t+\Delta t} = X_t + \Delta x = X_t + (\rho_x - \lambda x_t) + (\sqrt{\rho_x + \lambda x_t})\sqrt{\frac{\Delta t}{\Omega}}\xi \quad \text{Equation 13}$$

If the molecule count number dips below zero, it is automatically set back to zero to prevent the molecule numbers from assuming negative values.

Modelling non-autonomy via a stepwise approximation approach

The incorporation of non-autonomy was carried out following the work of Verd and colleagues (Verd et al., 2014), where the parameter value is kept constant over a short time step and incremented at regular intervals.

Quantification and Statistical Analysis

For all boxplots in

Figure 2, the lower and upper hinges correspond to the first and third quartiles. In addition, the upper whisker extends from the hinge to the largest value no further than 1.5 times the interquartile range. Outlier samples are coloured in red. Wilcoxon-Mann-Whitney unpaired two-sample test ****: p-value < 0.0001; * p-value < 0.01; ns = not significant.

Figures

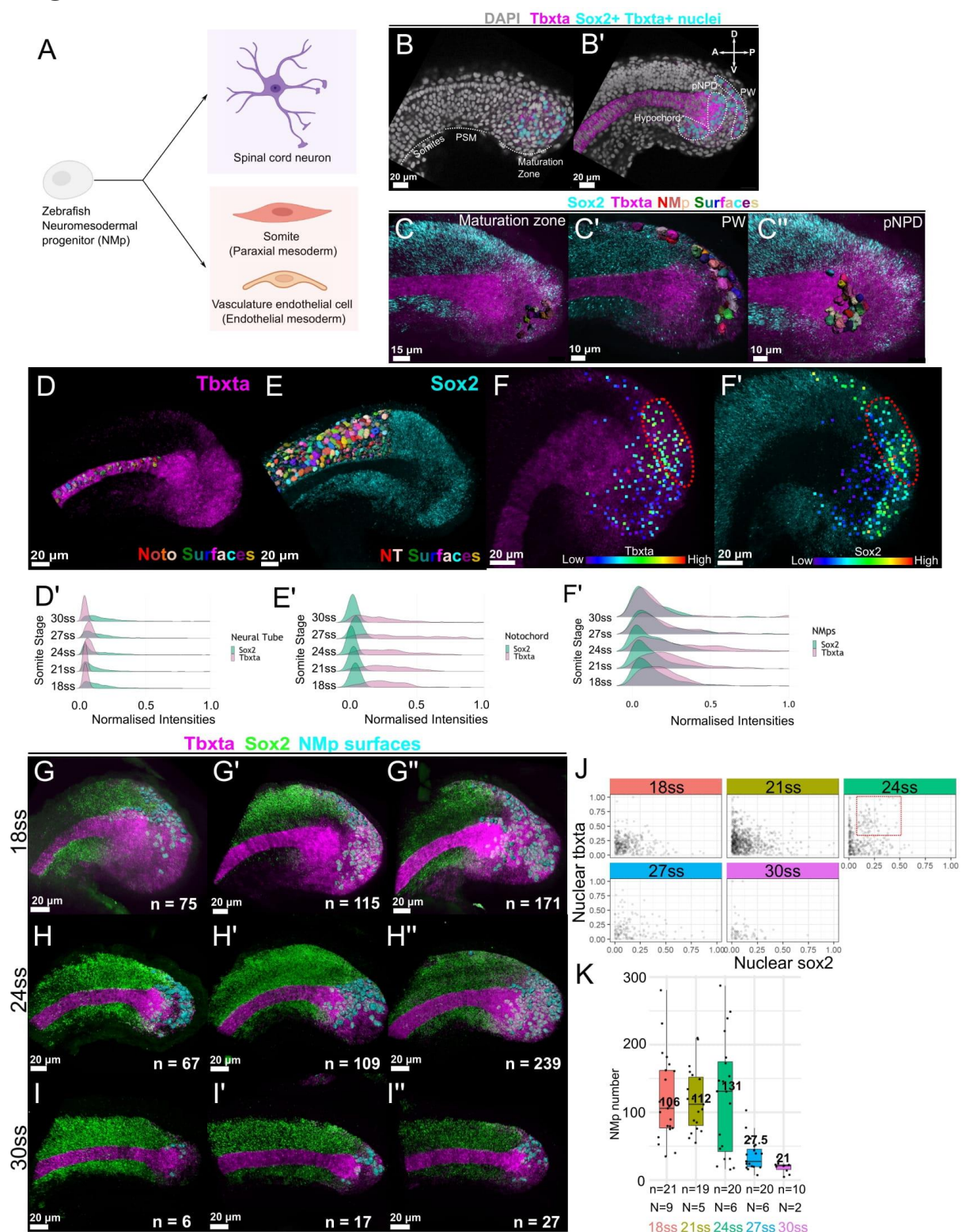


Figure 1. Heterogeneity in *sox2* and *tbx1a* expression and variability in the number and locations of NMPs peak at 24ss.

(A) Zebrafish NMps undertake a binary fate decision to differentiate into the posterior neural and mesodermal fates.

(B) 2D lateral slice showing *sox2+tbxta+* nuclei (cyan surfaces) in the maturation zone. (B') 2D medial slice showing *sox2+tbxta+* nuclei in the hypochord, pNPD and PW. pNPD: posterior notochord progenitor domain; PW: posterior wall.

(C-C'') Segmented NMp surfaces located in the (C) maturation zone (C') PW (C'') pNPD.

(D, E, F-F') Maximum intensity projections of *tbxta* and *sox2* shown alongside segmented surfaces of the posterior notochord (D) and posterior neural tube (E). NMps are shown as points colored according to their (F) *tbxta* and (F') *sox2* expression levels. The red regions highlight the NMps in the posterior wall that co-express intermediate levels of (F') *tbxta* and (F'') *sox2*, which is also highlighted with a red region in (J).

(D', E', F'') Ridge plots from 18ss to 30ss depicting the expression distributions of nuclear *sox2* and *tbxta* distributions in the (D') posterior notochord (E') posterior neural tube (F'') NMp populations.

(G-I'') HCR-stained samples at (G-G'') 18ss, (H-H'') 24ss and (I-I'') 30ss with three representative images per set. n: number of segmented NMps in each sample.

(J) Scatterplots of *sox2* and *tbxta* expression of NMps from 18ss to 30ss at three-somite intervals. Each point corresponds to the normalised nuclear *sox2* and *tbxta* intensities of a single NMp. The red region at 24ss highlights the NMps with intermediate levels of both genes.

(K) Box and whisker plots of the number of NMps from 18ss to 30ss at three-somite intervals. Each point corresponds to the number of NMps in a single sample. The median NMp number is indicated in bold. n: total number of samples analysed for each stage (biological replicate). N: number of distinct imaging experiments, where different biological samples imaged on the same day are considered a single imaging experiment.

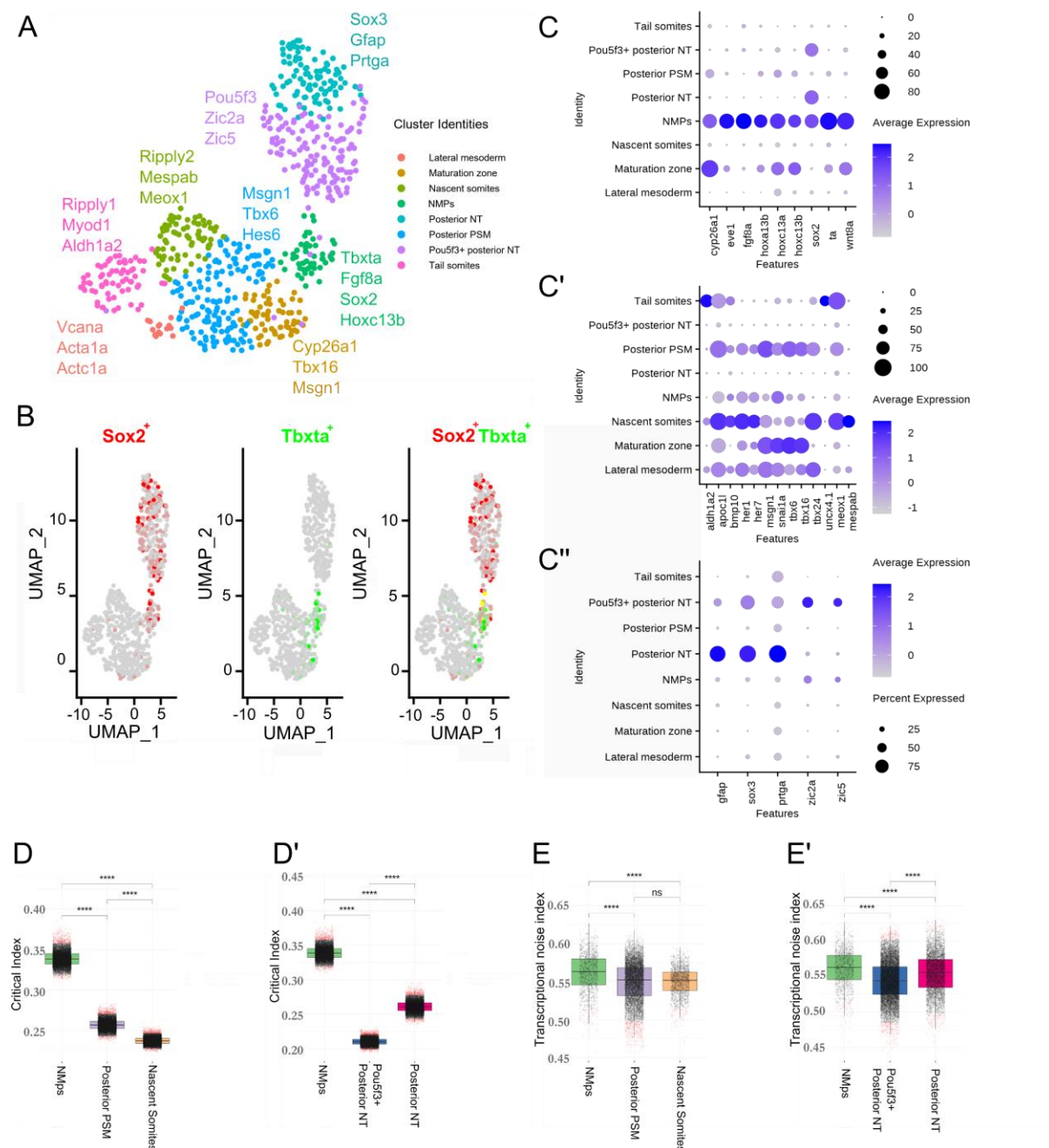


Figure 2. Analysis of 18ss scRNA-seq data reveals a peak in the critical index and transcriptional noise index in the NMP population relative to its derivatives

(A) UMAP embedding showing the 8 tailbud clusters at 18ss alongside the key differentially expressed genes used for manual annotation. UMAP: uniform manifold approximation and projection.

(B) UMAP embedding in (A) coloured by *sox2* and *tbxta* expression and *sox2-tbxta* together to illustrate co-expression.

(C-C'') Dot plots displaying the expression of differentially expressed genes from the tailbud (C) NMP cluster, (C') Mesodermal clusters and (C'') Neural clusters.

(D-D') Distribution of the critical transition index I_c calculated using marker genes of each cluster along the mesodermal (D) and neural (D') branches. A bootstrapping procedure was applied in calculating I_c to account for the differences in cell number between cell clusters.

(E-E') Distribution of the transcriptional noise index calculated using marker genes of each cluster along the mesodermal (E) and neural (E') branches. A bootstrapping procedure was applied in calculating the transcriptional noise index to account for the differences in cell number between cell clusters.

(E-E') Distribution of pairwise cell-to-cell distances / transcriptional noise along the mesodermal (E) and neural (E') branches.

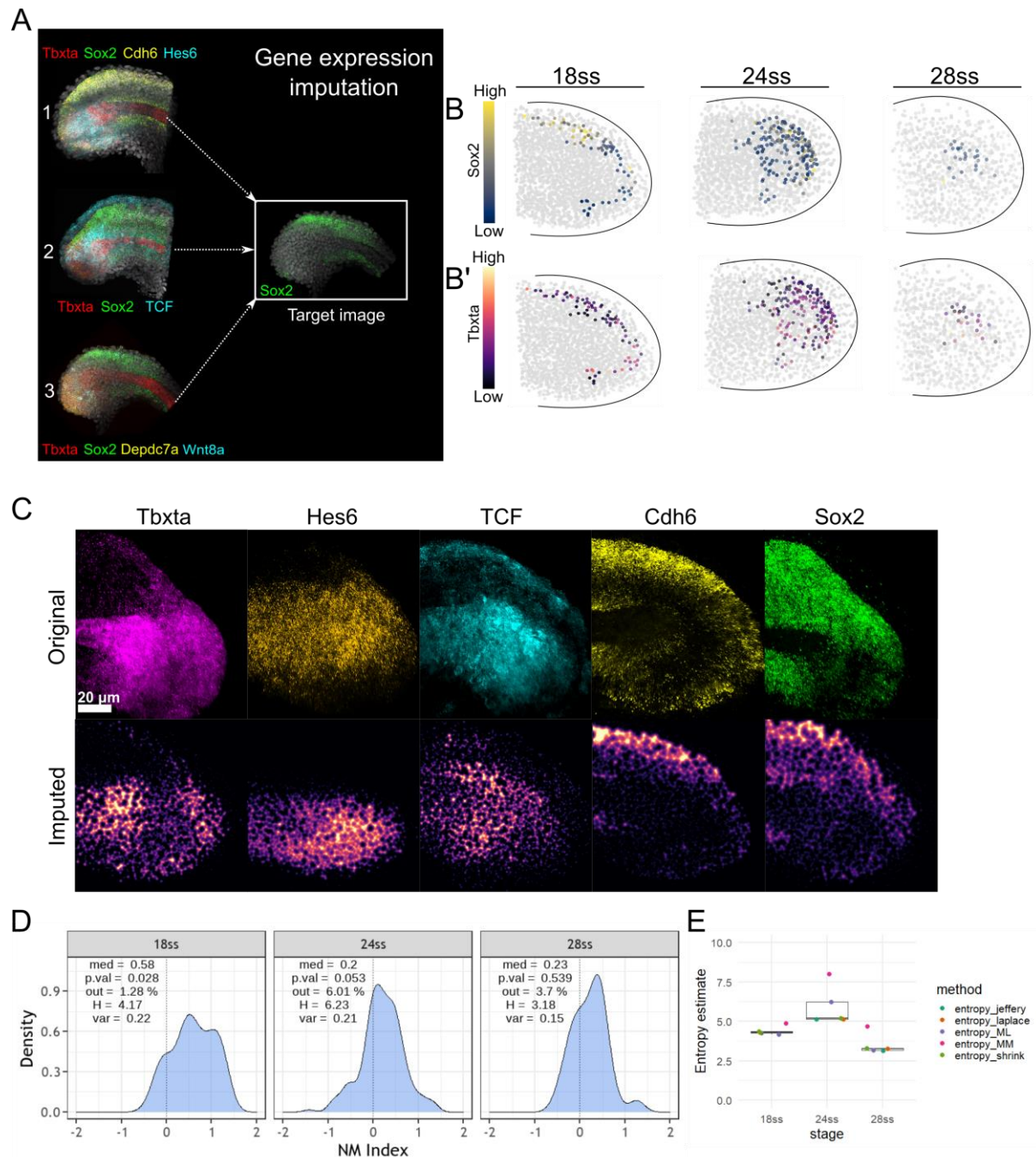


Figure 3. Gene expression imputation and the construction of a composite map via ZebReg demonstrates a peak in the NM index entropy at 24ss.

(A) Application of ZebReg for the imputation of multiple genes onto a target composite image. In the panel, *tbxta*, *cdh6*, *hes6*, *tcf*, *depdc7a* and *wnt8a* are imputed onto a target image that is stained only for *sox2*. *sox2* is the common colour channel used to assist the alignment of the source images onto the target image. In this example, the resultant target image has 7 distinct colour channels.

(B-B') Colouring *in silico* NMps in the 18ss, 24ss and 28ss composite maps by (B) *sox2* expression (B') *tbxta* expression levels.

(C) The top 'Original' row depicts the 2D projections of the HCR data at 18ss. The bottom 'Imputed' row depicts the corresponding expression of these genes in the target composite image.

(D) NM index density distributions computed from the 8-gene composite maps at 18ss, 24ss and 28ss. med: median; p.val: p-value for the Shapiro-Wilk test; out: outlier percentage; H: empirical entropy estimate; var: variance.

(E) Entropy estimates of the NMP index, with the estimation of the standard error obtained via jackknife resampling. The entropy estimates consistently peak at 24ss. entropy_jeffrey: Dirichlet-multinomial pseudocount entropy estimator (Dirichlet) with Jeffrey's prior; entropy_laplace: Dirichlet with Laplace's prior; entropy_ML: empirical maximum likelihood estimator; entropy_MM: Miller-Madow estimator; entropy_shrink: James-Stein type shrinkage estimator.

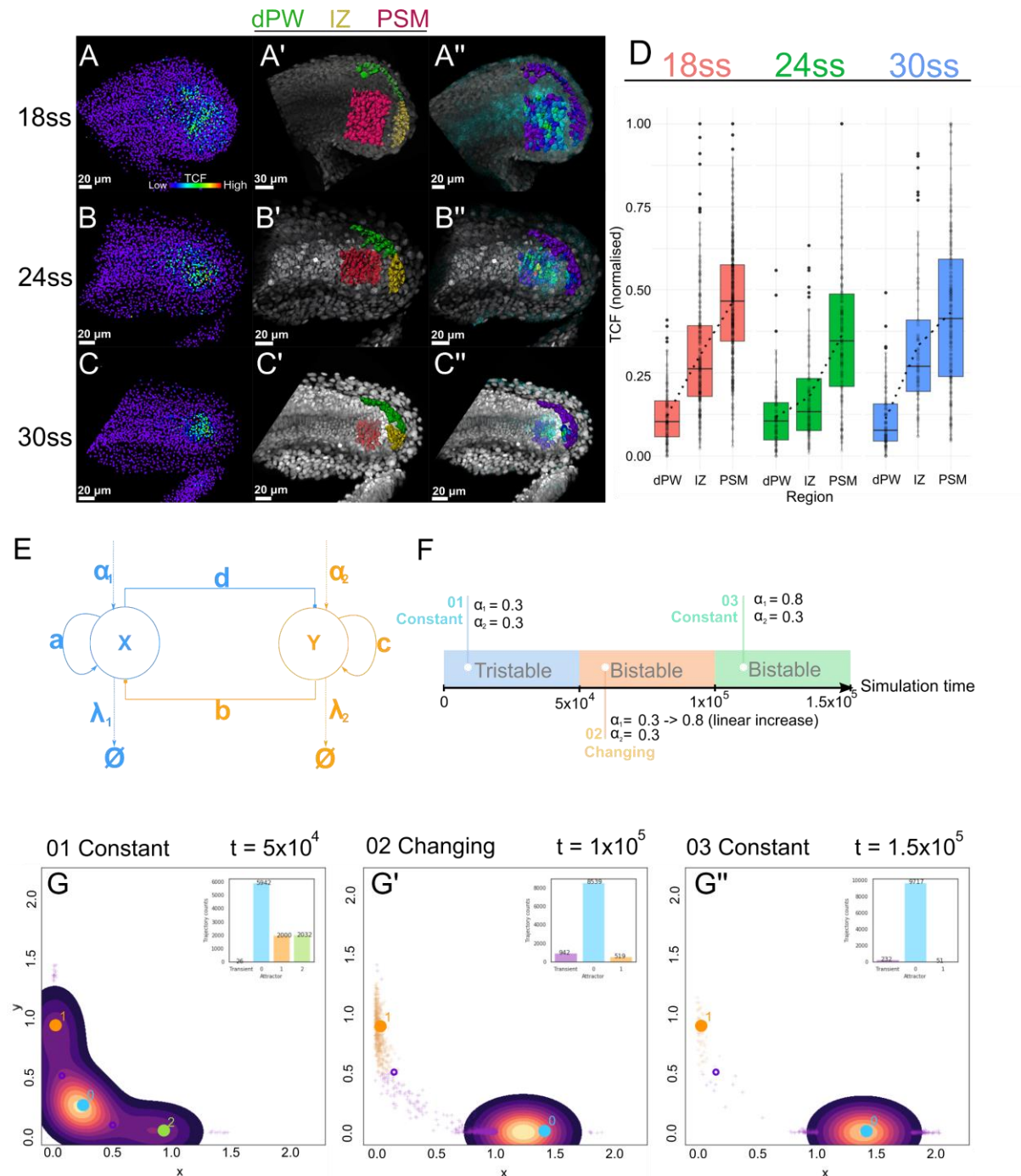


Figure 4. Stochastic modelling of a genetic toggle switch with a time-varying parameter uncovers 'rebellious' cells.

(A-A'') (A) Segmented 18ss tailbud nuclei coloured according to *tcf* expression. (A') Nuclei residing in the dPW (green), IZ (yellow) and PSM (red) are depicted. (A'') Same nuclei from (A') coloured according to *tcf* expression. dPW: dorsal posterior wall, IZ: intermediate zone, PSM: presomitic mesoderm.

(B-B'') Similar to (A-A'') showing tailbuds at 24ss.

(C-C'') Similar to (A-A'') showing tailbuds at 30ss.

(D) Quantification of normalised *tcf* expression levels in the dPW, IZ and PSM at 18ss, 24ss and 30ss.

(E) Schematic representation of the toggle switch model.

(F) Schematic of the computational experiment. The experiment has three stages: first (01 constant), the simulations are allowed to equilibrate in the tristable dynamical regime from $t = 0$ to $t = 5 \times 10^4$. Next (02 changing), from $t = 5 \times 10^4$ to $t = 1 \times 10^5$, the external activation parameter, α_1 is increased stepwise and reaches a value of 0.8 at $t = 1 \times 10^5$. Finally (03 constant), α_1 is kept constant for the remaining simulation duration from $t = 1 \times 10^5$ to $t = 1.5 \times 10^5$.

(G-G'') Combined results of 10,000 stochastic simulations. Each '+' symbol corresponds either to the location of a 'transient' trajectory with transient behaviour (purple) or to a 'rebellious' trajectory residing in the alternative attractor (orange). By $t = 1.5 \times 10^5$ (end of 03 Constant), most trajectories have entered the primary attractor (attractor 0).

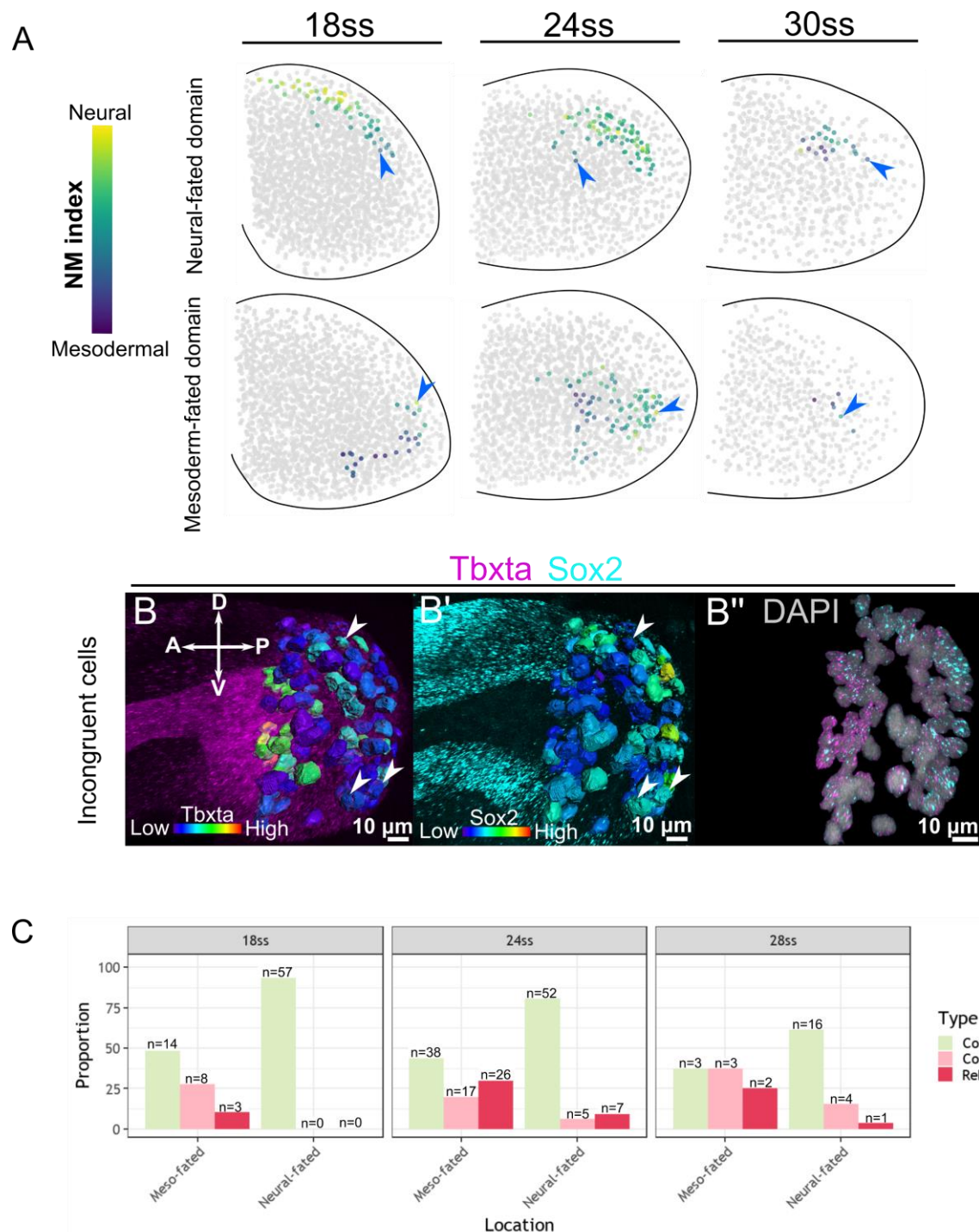


Figure 5. ZebReg's composite maps reveal that the number of Rebellious cells peak at 24ss.

(A) Demarcation of the neural-fated and mesoderm-fated domains in the composite maps. Non-NMps are coloured grey, whereas NMps are coloured according to their NM index levels. Blue arrows mark incongruent cells.

(B-B'') HCR stains of a representative zebrafish tailbud at 24ss for *tbxta* and *sox2*. Segmented surfaces correspond to NMps which are coloured by the expression levels of *tbxta* (B) and *sox2* (B'). Nuclear signals for *sox2* and *tbxta* are shown to illustrate co-expression (B''). Arrow heads mark Incongruent cells.

(C) Proportion of Congruent, Compliant and Rebellious cells in the mesoderm-fated and neural-fated domains at 18ss, 24ss and 28ss. At each stage, summing up the number of Compliant and Rebellious cells yields the number of Incongruent cells.