Morphogenetic coupling leads to pattern emergence in the presomitic mesoderm.

Timothy Fulton[1], Seongwon Hwang[1], Yuxuan Wang[1],
Lewis Thomson [1], Bethan Clark [1,2], Berta Verd,[1,3*] Benjamin Steventon[1*]

[1]Department of Genetics, University of Cambridge, Cambridge, UK
[2]Department of Zoology, University of Cambridge, Cambridge, UK
[3]Department of Zoology, University of Oxford, UK

*To whom correspondence should be addressed;
E-mail: berta.verdfernandez@zoo.ox.ac.uk, bjs57@cam.ac.uk

Pattern formation in development has been principally studied in tissues that are not undergoing extensive cellular rearrangement. However, in most developmental contexts, gene expression domains emerge as cells re-arrange their spatial positions within the tissue, providing an additional, and seldom explored, level of complexity to the process of pattern formation *in vivo*. To investigate this issue, we addressed the regulation of TBox expression in the presomitic mesoderm (PSM) as this tissue develops in zebrafish embryos. Here, cells must differentiate in a manner that leads to well-defined spatial gene expression domains along the tissue while undergoing rapid movements to generate axial length. We find that *in vivo*, mesoderm progenitors undergo TBox differentiation over a broad range of time scales while *in vitro* their differentiation is simultaneous. By reverse-engineering a gene regulatory network (GRN) to recapitulate TBox gene expression, we were able to predict the population-level differentiation dynamics observed in culture, but not *in vivo*. In order to address this discrepancy in differentiation dynamics we developed a 'Live Modelling' framework that allowed us to simulate the GRN on 3D tracking

data generated from large-scale time-lapse imaging datasets of the developing PSM. Once the network was simulated on a realistic representation of the cells' morphogenetic context, the model was able to recapitulate the range of differentiation time scales observed *in vivo*, and revealed that these were necessary for TBox gene expression patterns to emerge correctly at the level of the tissue. This work thus highlights a previously unappreciated role for cell movement as a driver of pattern formation in development.

As an embryo takes shape, two processes must occur in perfect harmony: cell fate decision making and cell movements that drive morphogenesis. At the tissue level, these two processes appear ordered and predictable leading to the emergence of well-defined gene expression domains throughout the embryo. While the role played by gene regulatory networks (GRNs) and morphogens in patterning the embryo has been extensively studied in multiple contexts (*1*), much less is known about the role of cell movements during this process. Here, we aim to determine how extracellular signals, cell-intrinsic gene-regulatory networks and cell movements act together to generate stable TBox expression domains along the zebrafish presomitic mesoderm (PSM). Within this tissue, cells transition from a *tbxta* positive progenitor state, upregulate *tbx16* as they become specified to the pre-somitic mesoderm, and then switch to a *tbx24* postive state as they enter the PSM prior to somitogenesis. Both FGF and Wnt signalling are known regulators of TBox gene expression within the zebrafish PSM (*2–4*). Response of the TBox GRN to Wnt and FGF inputs must be tightly coordinated in space, as PSM maturation ultimately directly links with oscillations of Notch signal activity which are known to be important for somitogenesis (*5*). To assess how TBox gene expression dynamics, Wnt/FGF signal pathway activity, and cell movements act together during PSM maturation we first obtained quantitative measures of each of these components.

To quantify the pattern of TBox expression across the PSM, we performed *in situ* Hybridisation Chain Reaction (HCR) stains at the 20 somite stage for *tbxta*, *tbx16* and *tbx24* mRNA,

that show the form spatially well-defined domains along the PSM (Figure 1A-B; D-G). In order to only capture the PSM expression of *tbxta*, the notochord and notochord progenitors were masked prior to analysis to exclude them from the measurements (Figure S1A-D). The normalised mean expression of each of the three TBox transcription factors was plotted on a normalised posterior to anterior axis, between 0 at the posterior end of the PSM and 1 at the posterior boundary of the most recently formed somite (Fig. 1C). Expression was normalised across multiple embryos between 0 and 1, with 0 representing the lowest mean expression, and 1 the highest mean expression. This describes a transition from an initial domain of *tbxta*/*tbx16* co-expression in the PSM, to only *tbx16*, followed by only *tbx24* expression in the anterior-most end of the tissue.

We next sought to quantify the spatial distribution of Wnt and FGF signalling activity, as these signals have been demonstrated to have a role in spatiotemporal regulation of somite formation. Wnt/TCF signalling was spatially quantified using a 7XTCF promoter fusion to a GFP reporter (*6*), with HCR performed against the *gfp* mRNA to give an immediate readout of the reporter activity (Figure 1H). FGF activity was quantified using an antibody against diphosphorylated ERK, as a readout of FGF ligand binding the FGF receptor (Figure 1I). These signalling profiles across the PSM were plotted along the same normalised axes as the TBox transcription factors and show a posterior to anterior gradient of Wnt, followed by an overlapping domain of FGF (Figure 1J)

While these profiles describe the gene expression domains' spatial arrangements, it is important to note that the rate of cell rearrangement varies significantly along the PSM. In particular, previous studies have characterised this tissue as undergoing a fluid-to-solid transition close to the *tbx16*/*tbx24* expression boundary (*7, 8*), where cell rearrangements quickly go from being extensive to practically non-existent. As a consequence of extensive cell mixing in the posterior progenitor zone, individual cells must somehow control their differentiation dynamics to match the timing of entry into the anterior PSM. To directly visualise the position at which the fluid-to-solid transition occurs, a number of photolabels were placed in the posterior, mid and anterior PSM using the photoconvertible nuclear protein, KikumeGR (*9–11*) (Figure 1K-M). Over the

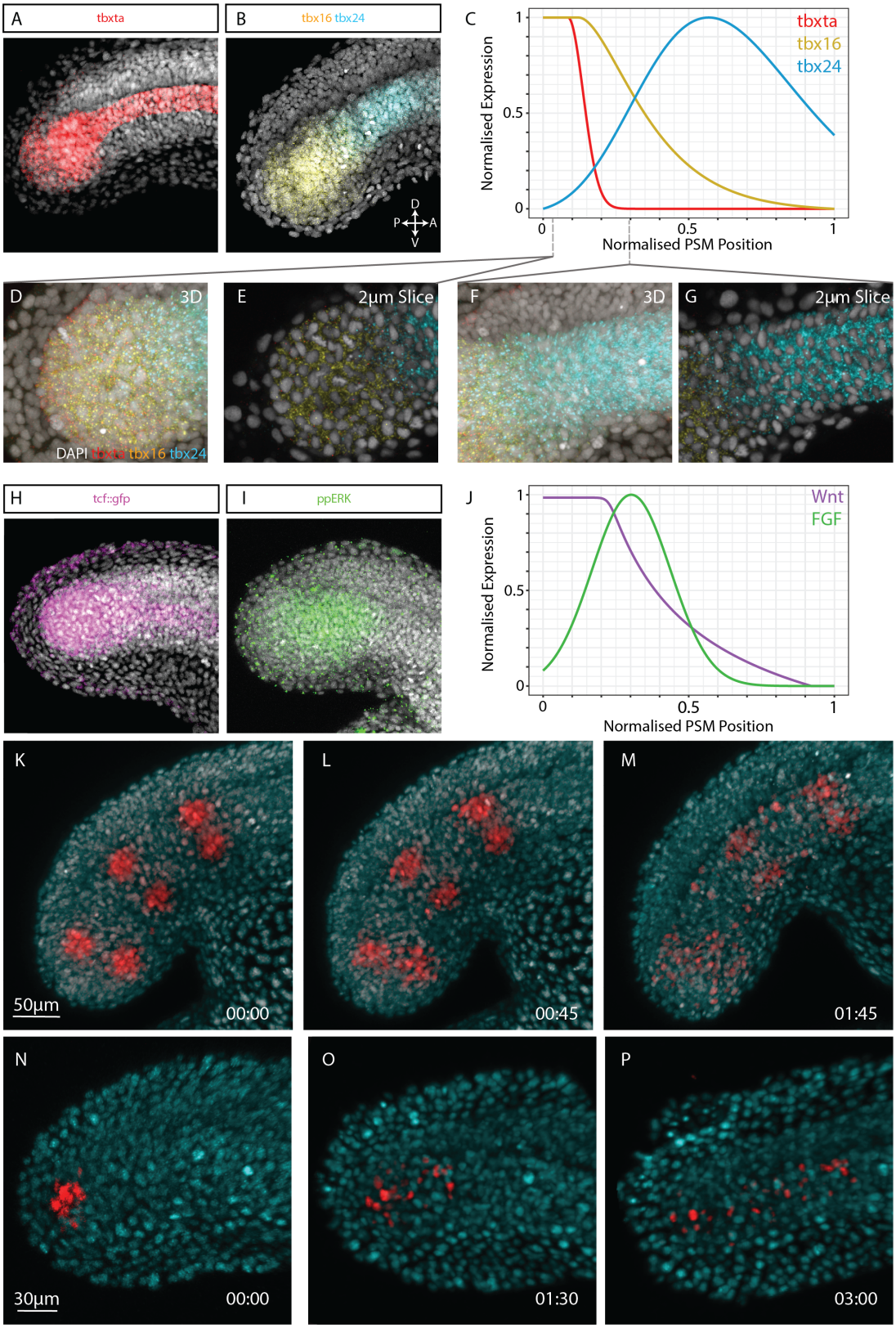

Figure 1: **Mesoderm progenitors differentiate into mature PSM over a range of time scales.** (Continues on the next page)

Figure 1: (From the previous page) Using HCR, the domains of gene expression for the three TBox transcription factors (A) Tbxta and (B) Tbx16 and Tbx24 (also known as Tbx6) were identified. (C) These were quantified along a normalised PSM length where 0 is posteriormost PSM and 1 is the posterior boundary of the most recently formed somite. Normalised expression is taken by mean intensity of each HCR signal divided by the maximum mean value from multiple embryos. Line plotted is the mean fitted curve taken from multiple fixed samples. (D-G) Zoom in of multiplex HCR showing all three TBox transcription factors in the posterior as a 3D rendering of a confocal image and 2µm slice. Wnt signalling was quantified using (H) a 7XTCF::GFP reporter with HCR used against *gfp* mRNA to give an immediate readout of transcription. (I) FGF signalling was quantified using an antibody against diphosphorylated ERK and (J) plotted on the same normalised axis. Using (K) small round labels placed along the length of the PSM using a photoconvertible nuclear protein, Kikume-GR, the degree of label spread (L-M) can be observed over time. Most spread is observed in the posterior label, with the least spread in the mid- and anterior label. (N) labels placed in the most posterior region of the PSM demonstrate that (O-P) the label spreads significantly over 90 minutes with the (P) first cell entering a newly formed somite in 3 hours. Plots in C and J show trend line compiled from 9 embryos. Labels in K-M representative of 2 labelled embryos. Labels in N-P representative of 4 embryos.

course of two hours, the posterior label becomes fully intermixed with unlabelled cells, whereas the middle and anterior labels remain unmixed, showing that these labelled cells travelled anteriorly as a single cohort (Supplementary Movie 1). Labels such as these can also me used to infer the minimum length of time taken by a mesoderm progenitor to differentiate into mature mesoderm and form a somite. We placed smaller posterior labels in the progenitor zone of a 20 somite stage zebrafish embryo and measured how long before the first labelled cell reached a somite. The fastest cell from these posterior labels reproducibly entered the most recently formed somite in three hours, while the rest of the cells remain spread throughout the PSM (Figure 1N-P) revealing that progenitors differentiate into mature PSM over a broad range of time scales. We assume that the longest time period over which a cell can take to differentiate is at least six hours as a small proportion of labelled cells are still located in the progenitor domain many hours after labelling. Together, these labels provide the range of timescales over which a cell can transition through the PSM (Figure 2A).

If we assume that the fastest cell traverses the PSM at constant speed, taking the most direct route from posterior to anterior in three hours (Figure 2B), we can translate space into time and approximate its gene expression and signalling dynamics from the previously obtained

quantitative gene expression and signalling profiles (Figure 1C; 1J) where the normalised PSM length on the x-axis now represents three hours of development. Our labelling experiments have shown, in agreement with previous studies, that cells spend varying amounts of time in the progenitor region, and progress towards the somites at a relatively constant pace once they have entered the PSM. To account for the slower differentiation dynamics of cells which remain longer within the posterior tailbud, we assume that they experience the high Wnt, low FGF and express high *tbxta*/*tbx16* levels while they are in the tailbud (Figure 2D). Using these generalised gene expression and dynamics, we set out to reverse engineer a GRN model to address the question of how spatio-temporal coordination of cell fate decisions may be regulated in this system.

We formulated a GRN model of TBox transcription factors and their external signals Wnt and FGF in terms of a dynamical system (See Materials and Methods). Gene products of the three TBox genes considered (*tbxta*, *tbx16* and *tbx24*) constitute the network's nodes and are represented by the state variables of the system, while the interactions between the genes and from the external signals (Wnt and FGF) are represented by model parameters (Figure 2C). As we have seen, mesoderm progenitors take anywhere between three and six hours to make their way into a somite from the tailbud, expressing a stereotypical progression of TBox genes (Figure 1A-C; Figure 2A-B). A suitable GRN will therefore be able to tune the onset of TBox differentiation according to differing lengths of time spent in the tailbud. We used a Markov Chain Monte Carlo (MCMC) algorithm to infer candidate networks which recapitulate the range of TBox gene expression dynamics observed *in vivo*. Out of 100 independent optimisation runs, 56 networks were able to recapitulate the expression dynamics of the full spectrum of differentiation timescales observed in progenitor cells (Figure 2D). All 56 networks share the sign of more than half of their parameters (13/24), and pairwise comparison between the networks shows that on average they differ by three interactions. Although the networks do not form clear distinct clusters (Figure S2) they resemble each other in topology. We decided to select a network for further study by filtering based on the sign of three network interactions that are very well-established in the literature: the activation of *tbx24* by Tbx16 (*2, 12*), the activation of

tbx16 by FGF (*3, 13*) and the repression of *tbx24* by Wnt (*12*) (Figure 2C, Table S1).

In order to understand how the inferred network controls the onset of differentiation in response to Wnt and FGF signalling dynamics we turned to phase space analysis. The network has been formulated as a non-autonomous dynamical system to accommodate the signalling dynamics, which causes the underlying phase portraits to also be dynamic. We use instantaneous phase portraits (*14, 15*) at every time point to characterise how the changing dynamical regime shapes the cells' trajectories (Supplementary Movie 2-5). We find that regardless of the time spent in the progenitor state (tailbud) the progression of dynamical regimes is always the same: cells start at a high *tbxta*/*tbx16* stable steady state, then a bifurcation at Wnt = 0.98 and FGF = 0.22 annihilates this steady state, releasing the trajectory which makes its way towards low *tbxta*/*tbx16* values, before converging towards a newly appeared stable steady state at high *tbx24* values. According to the model, the variation in differentiation timescales between different cells stems from the different amounts of time spent at the initial stable steady state, which is sustained by the initial concentrations of (high) Wnt and (low) FGF. However, the model also predicts that this steady state will disappear, and therefore that differentiation will begin, upon exposure of the cells to small amounts of FGF (FGF = 0.36) while still in a high Wnt signalling environment (Wnt = 0.98), suggesting that the observed signalling dynamics are not strictly required for the correct up-regulation of *tbx24* (Figure 2E) once a relatively low threshold of FGF has been met. This threshold in the embryo is already encountered by cells at 11% posterior-anterior (the tip of the tailbud being 0% and the most posterior somite boundary 100%) position. Note that a 11% posterior-anterior position is well embedded in the tailbud region.

The observation that posterior progenitor cells should differentiate into *tbx24* positive paraxial mesoderm in the absence of exposure to the full dynamics of Wnt and FGF signals is surprising given the known role of FGF in regulating PSM development and maturation. To test this prediction experimentally, we made use of an *in vitro* culture method that has previously been shown so support the autonomous oscillations of the segmentation clock (*16*). We explant the posterior 30% of the progenitor zone at the 18 somite stage and cultured them as single cells in a fully defined L15 medium without any signal supplementation (Figure 3E-F). Our assumption

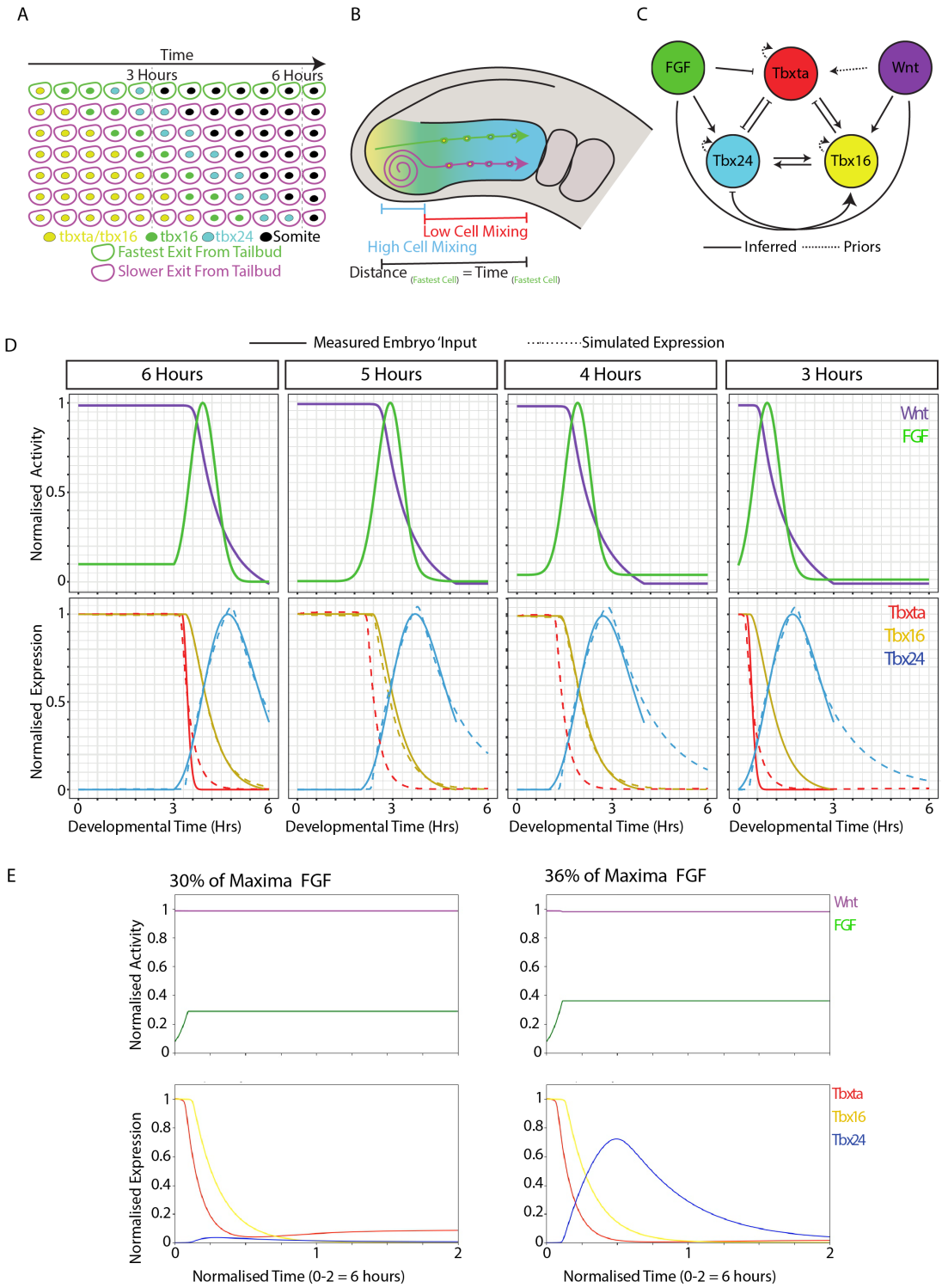

Figure 2: **A reverse engineered GRN that tunes the onset of differentiation in response to external signals.** (A) Cells differentiate along a range of timescales, with the fastest cell entering a somite in three hours with slower cells taking a range of time scales to differentiate. The timescale over which a cell takes to differentiate (B) is given by the period of time taken spent in the posterior of the PSM. (Continues on the next page).

Figure 2: (From the previous page.) (C) Using these timescales and patterns of gene expression in space, a reverse engineered gene regulatory network was produced. This model was demonstrated to reproduce (D) the gene expression pattern across a range of timescales between the fastest and slowest assumed time scales. The signalling dynamic of (E) FGF is shown to be essential for initiating the differentiation of progenitors to *tbx24* postitive, with only a small increase in FGF, up to 36% of the maximum *in vivo*, required to initiate differentiation

is that once removed from their signalling environments, cells' Wnt and FGF levels will remain constant, and that, in line with model predictions, only those cells that have already experienced the FGF threshold will differentiate. Those cells below the threshold of FGF will remain in a progenitor state. To determine the activity of Wnt and FGF pathways in dissociated cells we first performed qPCR against *axin2* and *sprouty4* (Figure 3A-B), which function as transcriptional readouts of Wnt and FGF signalling respectively. Both *sprouty4* and *axin2* maintained the same concentration present in the tailbud explants as measured immediately after dissection and prior to dissociation. To follow the profile of Wnt and FGF in single cells, we performed HCR against the *gfp* of the 7XTCF::GFP reporter and an antibody stain against diphosphorylated ERK (Figure 3C-D; 3M). These stains were both imaged at the single cell level and confirmed no significant spike in FGF or Wnt activity occurred across the time course, with a modest decrease in FGF signalling in the final hours whilst maintaining the population average (Figure 3C).

So far, we have used the model to simulate TBox expression dynamics in a generalised cell. In order to better compare model predictions to the experimental data obtained from both qPCR experiments and single cell HCR stains we simulated TBox expression dynamics in a population of *in silico*-dissected tailbud progenitors. We segmented the nuclei and quantified the levels of *tbxta*, *tbx16* and *tbx24* of all the cells in the posterior 30% region of the PSM of an HCR-stained 18ss embryo; a total of 736 cells. These concentrations were used as the initial conditions from which to simulate the GRN model in every cell. Wnt and FGF values were inferred from the Wnt and FGF expression profiles (Figure 1J) depending on a cells relative antero-posterior position and kept constant throughout. Simulated qPCR results were obtained by pooling the concentrations of the TBox genes in every cell at given time points. Qualitatively, we see a good correspondence between simulated (Figure 3G) and experimental (Figure

3H) qPCR profiles for both *tbx16* and *tbx24*, particularly over the first four hours post dissociation with more discrepancy at six hours, possibly due to increased cell death. If we now turn to look at single cell expression dynamics in the populations of *in silico*-dissected tailbud progenitors, the population-level model predicts, in agreement with the model of the generalised cell, that a bi-modal distribution will emerge reflecting the fact that some cells differentiate (up-regulate *tbx24* and down-regulate*tbx16*), namely those that have been exposed to FGF levels above the threshold, while others don't (do not up-regulate *tbx24* and maintain *tbx16*), namely those that haven't been exposed to the minimum FGF threshold, and that those that do differentiate will do so synchronously (Figure 3I; 3K). This same synchronous differentiation was observed experimentally when cells were imaged individually using HCR (Figure 3J; 3L; 3N; 3O). The non-differentiating cells which do not up regulate *tbx24* remain high in *tbxta*, and do not express the ectodermal marker (*keratin18*) (Figure 3O). Taken together, these results show that PSM specification is already established within the posterior progenitor domain and that differentiation takes place synchronously for all the cells over a defined time-period of 4-6 hours. These *in vitro* gene expression dynamics differ starkly from the range of differentiation time scales that we described *in vivo* (Figure 1N-P; 2A-B), suggesting that an additional layer of regulation must exist to slow down the differentiation of cells remaining in the progenitor domain. As morphogenesis is notoriously missing from our *in vitro* set up, we reasoned that cell movements themselves may be responsible for tuning the dynamics of FGF exposure, hence controlling the onset of TBox expression.

We hypothesised that the cell movements driving PSM elongation might themselves regulate when cells surpass the minimal FGF threshold required for differentiation - thereby creating a feedback between tissue morphogenesis and gene regulatory networks from which pattern at the tissue level could emerge. To explore this idea, we next aimed to determine how our GRN model would predict Tbox gene expression when simulated in the real morphogenetic context of PSM developement. We generated 3D cell tracking datasets in which the movements of individual cells within PSM are followed for three hours (Figure 4A). Using multi-photon imaging, the tracks of single cells could be generated using automatic nuclei tracking, with a track ac-

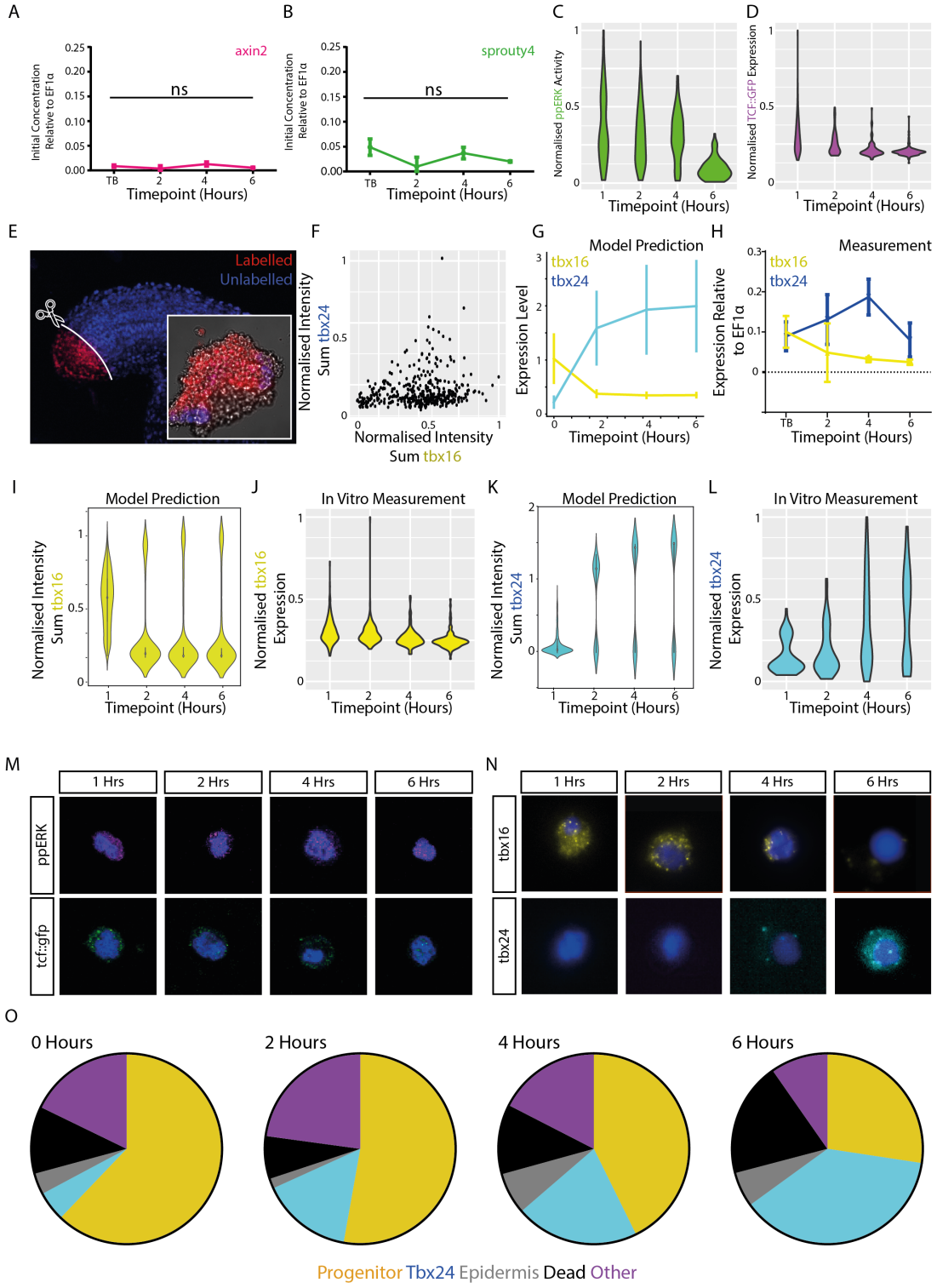

Figure 3: **Single progenitor cells *in vitro* differentiate synchronously.** Cells of the posterior 25% of the zebrafish PSM were explanted and, using bulk qPCR were demonstrated not to upregulate (A) Wnt or (B) FGF signalling as measured by *axin2* and *sprouty4* expression. (Continues on the next page).

Figure 3: (From the previous page.) Values compared to explanted, non-dissociated tailbud (TB). This was also demonstrated by staining for (C) *gfp* mRNA on the 7XTCF::GFP reporter (D) or antibody staining using ppERK. Explants were (E) labelled using photoconvertible KikumeGR and explanted then reimaged, demonstrating the majority of the explant comes from the posterior 25% of the PSM. Staining this region using HCR and segmenting out individual nuclei (F) confirms the majority of cells contain high levels of nuclear *tbx16* transcripts and low *tbx24* transcripts. (G) Simulation of culture of these cells given signalling dynamics given by their initial position within the embryo, and maintained throughout culture predicts an increase in expression in *tbx24* over time. This prediction is confirmed (H) by bulk qPCR analysis of isolated single cells *in vitro*. (I) Plotting the spread of predicted expression levels from single cells predicts down regulation of *tbx16* and (K) an increase in *tbx24* expression, forming two populations with high and low expression. These predictions are tested *in vitro* and confirmed to result in a (J) reduction in *tbx16* and (L) an increase in *tbx24* over time. We observe a population expressing high *tbx24* and low *tbx24* at the final timepoint. Example images of fixed cells at each timepoint showing (M) signal activity and (N) TBox trascription factor expression. (O) Cells were classified as progenitors if they expressed *tbx16*, *tbxta* or both, as *tbx24* positive, as *keratin18* expressing epidermis or other. Dead cells were classified based on observation of nuclear shape. Over the time-course, the proportion of progenitor cells decreased, with an increase in *tbx24* positive cells. (O) proportions taken from three independent experiments with a minimum of 174 cells per experiment. Progenitor defined as a cell expressing *tbx16* and/or *tbxta*.

curacy of approximately 80% (Supplementary Movie 6 & 7). In this time, three somites were formed. We extracted cell tracking data from only the unsegmented PSM and at each time-point normalised the PSM length, removing each new somite as it is formed. We projected the Wnt and FGF signalling profiles measured from fixed embryos (Figure 1J; Supplementary Movie 8) onto the tracks. Thus each cell in the simulation continually updates its signal input depending on position within the tissue, and uses this input to simulate TBox gene expression at the next time-point (Figure 4B-C; Figure S3). We name this framework 'Live Modelling', since it simulates GRNs on cell tracking data obtained from live-imaging. Unlike what we observe in the *in vitro* simulations, live modelling simulations do not predict the bi-modal and synchronous differentiation of progenitor cells, which would lead to a *tbx24* expression domain starting at 11% posterior-anterior position (where the FGF threshold is positioned in the embryo) and reaching the anterior of the PSM. Instead, live-modelling recapitulates the asynchronous differentiation of the progenitors observed *in vivo*, and shows that it is required for correct patterning at the level of the tissue (Figure 4B). As signalling domains and TBox transcription factor expression domains were demonstrated to accurately scale during axial elongation (Figure S1E-F) we

were able to project the same signalling domains onto multiphoton movies, taken from different stage embryos. Correct pattern formation was also obtained when different tracks were used demonstrating that this approach recapitulates TBox gene expression profiles at different stage embryos, yielding accurate spatial patterning despite subtly different morphogenetic processes occurring (Figure S3).

When following the output of live modelling simulations over time (Figure 4B-C; Figure S3; Supplementary Movie 9), we observed two deviations from the mean expression profiles obtained for TBox gene expression. Firstly, the anterior *tbx24* expression did not decline in the simulations as observed *in vivo*, likely due to the absence of additional known anterior repressors such as Ripply1 (*17*) from our model. Secondly, the model predicts that single cells in the posterior progenitor domain display an increased level of heterogeneity in *tbx16* and *tbx24* expression compared to the anterior domain (Figure 4C). To assess whether this relates to the direction of cell movement, we re-sized each point according to cell velocity, as a proxy for directional movement, with small points representing cells that displace towards the posterior, and large points those that displace in the anterior direction (Figure 4D-E). This reveals a correlation between heterogeneous TBox expression with an increased posterior displacement, and suggest that the maintenance of a *tbx16*/*tbx24* boundary is in part a consequence of cells being displaced out of the progenitor domain only once they have passed the threshold of FGF required for *tbx24* expression.

Our model also predicts the existence of an increased level of *tbx24* heterogeneity in the posterior PSM, likely due to the random posterior displacement of cells that have already begun to differentiate. To identify whether there is indeed a posterior bias of ectopic expression of *tbx24* as predicted by the live modelling, the nuclei within the PSM were segmented (Figure 4F) and then classified as either anterior or posterior relative to the intersection between the *tbx16* and *tbx24* expression profiles within each embryo. Following this, cells in each domain were identified which had aberrant gene expression, defined as expressing *tbx24* in the posterior, or *tbx16* in the anterior domain. The frequency of aberrant gene expression was measured and confirmed a posterior bias in erroneous gene expression, as predicted by the live modelling

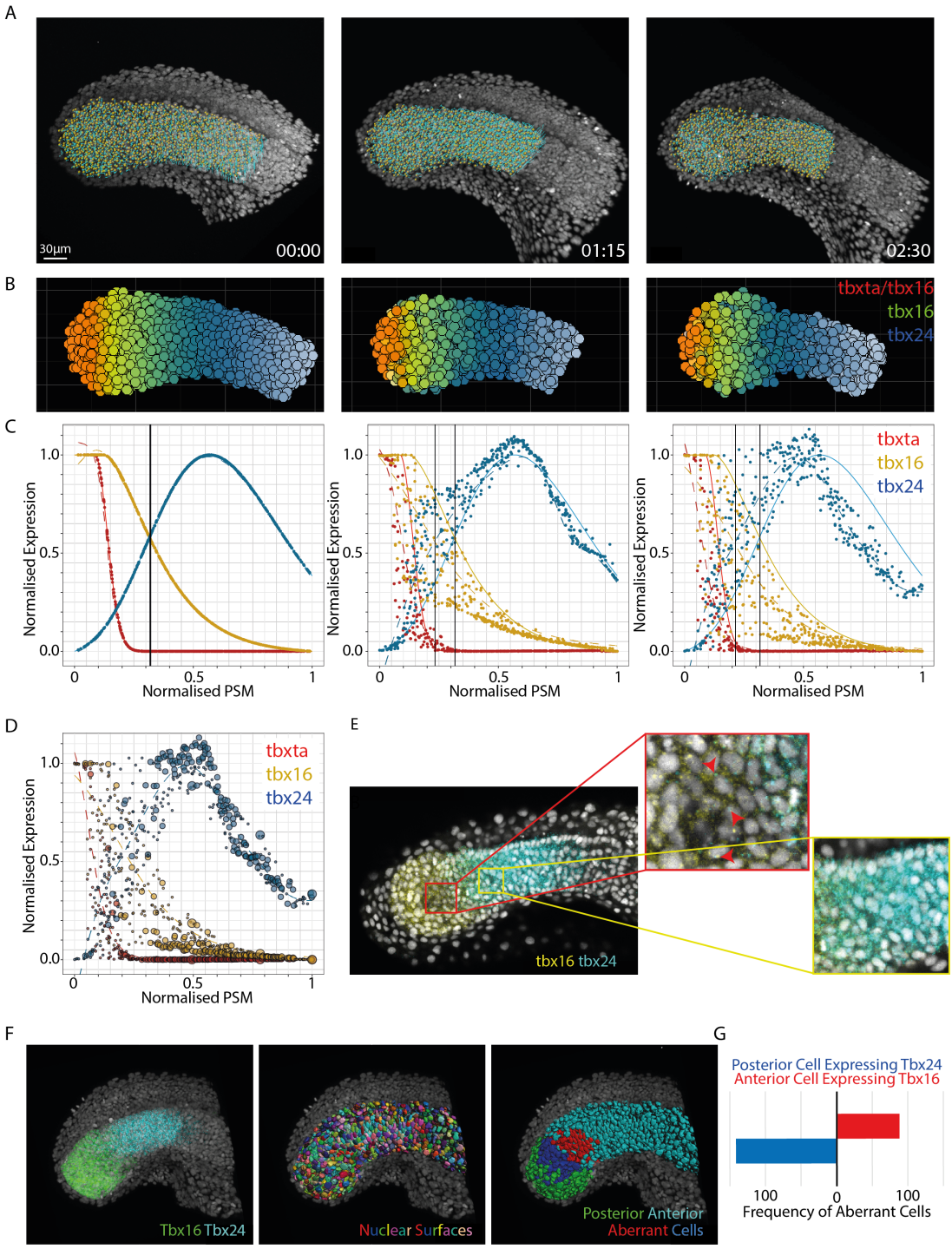

Figure 4: **Live modelling couples gene regulatory networks to morphogenesis.** (A) Live imaging of the zebrafish PSM over the period of 2.5 hours using two-photon microscopy with a 150 second time step allows cell tracking of individual cells within the PSM. These tracks are used (B) to model the PSM GRN on by cells reading in their levels of FGF and Wnt signalling at each time point. (Caption continues on the next page.)

Figure 4: (From the previous page.) Using these signal inputs, each cell predicts its transcription factor gene expression profile given by the colour map. This is quantified (C) by maximum projection with each cell positioned along a normalised PSM axis. This demonstrates good correlation between the predicted pattern (points and dotted trend line) and measured *in vivo* values shown by the solid line. Solid bar represents position along the PSM where *tbx16* expression is less than *tbx24* expression on average. (D) Deviation from the *in vivo* trend line is observed in the posterior and correlation between cell movement direction and gene expression state demonstrates posteriorly moving cells are more prone to aberrant gene expression. Large points correlate with highly posterior to anterior movement of cells. (E-G) These aberrant cells can be observed *in vivo*, with more erroneous gene expression identified in the posterior than the anterior. Bars in (G) represent mean number of aberrant cells in each domain from five embryos. Boundary between anterior and posterior define per embryo, where *tbx16* expression exceeds *tbx24* as measured in Figure 1C.

(Figure 4G).

Taken together we propose a mechanism whereby dynamic signalling driven by cell movements in the mesoderm progenitors tunes the timing of differentiation onset in single cells, leading to the emergence of well-defined gene expression domains along the PSM while at the same time maintaining the undifferentiated progenitor pool. In this way, our work reveals a previously unappreciated generative role of cell movements in pattern emergence, showing that signalling cues are coupled to cell-intrinsic GRNs by the process of morphogenesis itself. In addition, we propose that the fluid-to-solid transition (*8*) might be responsible for the observed heterogeneity in the progenitor domain, causing a minority of differentiated cells to be transiently pushed backwards as the PSM solidifies. Therefore, tissue morphogenesis acts as a direct regulator of pattern formation, both in terms of enabling coordinated differentiation towards the anterior, as well as generating the observed gene expression heterogeneity within posterior progenitors. This heterogeneity might play a role in the maintenance of uncommitted progenitors, or as our work seems to suggest, might not be functional.

Across the presomitic mesoderm (PSM), a posterior to anterior gradient of FGF has been demonstrated to coordinate the spatiotemporal organisation of PSM cells into somites in the anterior, by interactions with Notch Signalling (*16, 18–21*). In addition to this, FGF has also been demonstrated to regulate the early EMT process which generates the mesenchymal population of progenitor cells within the posterior of the PSM. This mesenchyme has also been demonstrated to be responsible for elongating the embryonic axis (*3, 7*). Finally FGF signalling has

been demonstrated to regulate cell fate decisions of mesodermal progenitor cells (*3, 12*). FGF is therefore able to regulate cell movements, tissue morphogenesis and also cell fate decisions, however how one signal is able to be accurately interpreted in such a variety of ways within a single tissue remains unclear. Our model proposes that the function of FGF in regulating PSM differentiation may be limited only to the posterior most region of the PSM, perhaps allowing for a spatial compartmentalisation of its function to allow for multiple different cell responses to a single signalling profile. Alternatively, single cell dynamic profiles of FGF pathway activity may be essential for providing more precise interpretation to control multiple aspects of cell function (*22*) . Further work is required to incorporate precise spatial manipulation of FGF activity across the PSM, within the context of simultaneously monitoring the impact on cell movements through the use of Live Modelling approaches.

The observation that cells *in vitro* differentiate simultaneously is highly reminiscent of the differentiation of stem cells when responding to external signal sources in culture. Here too, cells often undergo direct differentiation, without experiencing the range of dynamic differentiation profiles that we have seen are required to generate gene expression patterns within a 3D morphogenetic context. Our results provide insight into the manner in which the cell movements of morphogenesis themselves control the temporal regulation of differentiation, via displacing cells relative to the sources or sinks of extracellular signals. Such a mechanism represents an example of downward causation (*23*) (information flow from the tissue to the single cell level i.e. in the global rate of tissue morphogenesis and accompanying cell rearrangements) and illustrates its central role patterning the embryo. Further investigation will be key in order to improve our understanding of how gene expression patterns emerge within developing multi-cellular systems, and will help identify which control parameters, at various levels of biological organisation, must be modified and how in order to reliably bioengineer organoids for therapeutic applications.

Acknowledgements

We would like to thank Dillan Saunders and Meagan Hennessy for feedback on this manuscript. We would like to thank the Steven Wilson lab for sharing the Tg(7xTCF-Xla.Sia:GFP), used here as 7XTCF::GFP reporter. Thanks to the Cambridge Advanced Imaging Centre (CAIC) for imaging support. T.F., S.H. and B.S. are supported by a Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (109408/Z/15/Z) and T.F. by a scholarship from the Cambridge Trust, University of Cambridge. L.T. is supported by a scholarship from the BBSRC. Y.W. is supported by a summer vacation stipend from St Catharine's College, University of Cambridge. B.C. is supported by a Stipend from the Bedford Fund, King's College, University of Cambridge and a scholarship from the Wellcome Trust. B.V. was supported by a Herschel Smith Postdoctoral Fellowship, University of Cambridge and Department of Zoology, University of Oxford.

Methods

Animal Husbandry

This research was regulated 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). Embryos were obtained and raised in standard E3 media at 28°C. Wild Type lines are either Tüpfel Long Fin (TL), AB or AB/TL. The Tg(7xTCF-Xla.Sia:GFP) reporter line (*6*) was provided by the Steven Wilson laboratory. Embryos were staged as in (*24*)

Primary Culture of Tailbud Progenitor Cells

Cells were explanted from the tailbud as in (*16*). Effort was made to remove the ectoderm prior to dissection. Cells were dissected in calcium and magnesium free PBS in order to promote cell dissociation. Cells were cultures in 8 well Ibidi Micro-Slides under the fully defined L15 media supplemented with PenStrep solution to limit bacterial growth.

RNA extractions were made in triplicate, from independent experiments, using Trizol Reagent

(Ambion LifeTechnologies) following a standard protocol and reverse transcription using Superscript III (Invitrogen). Resultant cDNA was quantified using SYBRGreen with liquid handling robot (Qiagility, Qiagen) and analysed on a RotorGeneQ thermocycler (Qiagen). Primer sequences: *axin2* 5'-TACCCTCGGACACTTCAAGG-3' and 5'-TGCCCTCATACATTGGCAGA-3'; *sprouty4* 5'-CACGCGCCCTAGTATCAAAC-3' and 5'-GGGATCTTGGTGAAGTGTGC-3'; *EF1a* 5'-GGAGACTGGTGTCCTCAA-3' and 5'-GGTGCATCTCAACAGACTT-3'. Concentration of cDNA was estimated using an in-house MAK2 analysis method, as described previously (*25*).

In Situ Hybridisation Chain Reaction (HCR)

Embryos were raised to the required stage then fixed in 4% PFA in DEPC treated PBS without calcium and magnesium at 4°C overnight. Embryos were then stained using HCR following the standard zebrafish protocol found in (*26*). Probes, fluorescent hairpins and buffers were all purchased from Molecular Instruments. After staining, samples were counterstained with DAPI and mounted under 80% glycerol.

Immunohistochemistry

Embryos were raised to the required stage then fixed in 4% PFA in DEPC treated PBS without calcium and magnesium at 4°C overnight. Embryos were then blocked in 3% goat serum in 0.25% Triton, 1% DMSO, in PBS for one hour at room temperature. Diphosphorylated ERK was detected using the primary antibody (M9692-200UL, Sigma) diluted 1 in 500 in 3% goat serum in 0.25% Triton, 1% DMSO, in PBS. The samples were incubated at 4°C overnight then washed in 0.25% Triton, 1% DMSO, in PBS. Secondary Alexa 647nm conjugated antibodies were diluted 1 in 500 in 3% goat serum in 0.25% Triton, 1% DMSO, 1X DAPI in PBS and applied overnight at 4°C.

Imaging and Image Analysis

Samples were imaged using a Zeiss LSM700 inverted confocal at 12 bit, 20X or 40X magnification, with an image resolution of 512x512 pixels. Single cell HCR was imaged using a Nikon

Ti inverted widefield microscope at 63X magnification.

Image analysis of confocal images was done using the line drawing tool on Fiji (*27, 28*) set to a width of 50 pixels. Lines were drawn following the curve of the embryo, through the centre of the PSM from posterior PSM to the posterior most clear somite boundary. Profiles were normalised to the length of the PSM and signal intensity as individual embryos by dividing the measured value by the maximum value of that embryo.

Nuclear segmentation of whole embryos stained using HCR was conducted using a tight mask applied around the DAPI stain using Imaris (Bitplane) with a surface detail of 0.5µm. Touching surfaces were split using a seed size of 4µm. Values were exported as X, Y, Z coordinates relative to the posteriormost tip of the PSM where X, Y, Z were equal to (0, 0, 0). The PSM was then segmented by hand by deleting nuclear surfaces outside of the PSM, including notochord, spinal cord, anterior somites and ectoderm. Only the PSM closest to the imaging objective, therefore of highest imaging quality was measured with the distal PSM also removed.

Intensity mean values of each transcription factor HCR signal were exported and normalised between 0 and 1 by dividing each cell's mean signal intensity by the maximum measured within that sample, per gene. PSM length was normalised individually between 0 and 1 by division of the position in X by the maximum X value measured in each embryo.

Single cell image analysis was conducted using Imaris (Bitplane) by generating loose surface masks around the DAPI stain to capture the full nuclear region and a small region of cytoplasm. Surface masks were then filtered to remove any masks where two cells joined together or small surfaces caused by background noise, or fragmented apoptotic nuclei. The intensity sum of each channel was measured and normalised by the area of the surface, as surface area and transcript intensity had been demonstrated to correlate. Expression level was then normalised between 0 and 1 using the maximum value measured for each gene, in each experiment.

Live imaging datasets of the developing PSM was created using a TriM Scope II Upright 2-photon scanning fluorescence microscope equipped Insight DeepSee dual-line laser (tunable 710-1300 nm fixed 1040 nm line). Embryo was imaged with a 25X 1.05 NA water dipping objective. Time step and frame number as per figure legend. Embryos laterally in low melting

agarose with the entire tail cut free to permit normal development (*29*).

Model formulation

We formulated the Tbox gene regulatory network using a dynamical systems formulation. The models aim is to recapitulate the dynamics of Tbox gene expression for any cell, or rather a general cell, in the developing zebrafish PSM. We use a connectionist model formulation previosly used to model other developmental patterning processes (*30*).

The mRNA concentrations encoded by the Tbox genes *tbxta*, *tbx16* and *tbx24* are represented by the state variables of the dynamical system. For each gene, the concentration of its associated mRNA a at time t is given by $g^a(t)$. mRNA concentration over time is governed by the following system of three coupled ordinary differential equations:

$$\frac{dg^a(t)}{dt} = R^a\phi(u^a) - \lambda^a g^a(t) \tag{1}$$

where R^a and λ^a respectively represent the rates of mRNA production and decay. ϕ is a sigmoid regulation-expression function used to represent the cooperative, saturating, coarse-grained kinetics of transcriptional regulation and introduces non-linearities into the model that enable it to exhibit complex behaviours:

$$\phi(u^a) = \frac{1}{2}\Big(\frac{u^a}{\sqrt{(u^a)^2+1}} + 1\Big), \tag{2}$$

where

$$u^a = \sum_{b\in G} W^{ba} g^b(t) + \sum_{s\in S} E^{sa} g^s(t) + h^a. \tag{3}$$

$G = \{tbxta, tbx16, tbx24\}$ is the set of Tbox genes while $S = \{\mathrm{Wnt}, \mathrm{FGF}\}$ is the set of external regulatory inputs provided by the Wnt and FGF signalling environments. The concentrations of the external regulators g^s are interpolated from quantified spatial mRNA expression data (Figure 1J) and translated into time as explained in the main text to used as dynamic inputs to the model. Changing Wnt and FGF concentrations over time renders the parameter term $\sum_{s\in S} E^{sa} g^s(t)$ time-dependent and therefore the model non-autonomous (*14, 15*).

The interconnectivity matrices W and E house the parameters representing the regulatory interactions among the TBox genes, and from Wnt and FGF to the Tbox genes, respectively.

Matrix elements w^{ba} and e^{sa} are the parameters representing the effect of regulator b or s on target gene a. These can be positive (representing an activation from b or s onto a), negative (repression), or close to zero (no interaction). h^a is a threshold parameter representing the basal activity of gene a, which acknowledges the presence of regulators absent from our model. Model parameters are detailed in Table S1.

Model fitting and selection

We reverse-engineered values for parameters R^a, λ^a, W, E, and h^a by fitting the model to the inferred gene expression dynamics of a cell that took 6 hours to enter a somite from an initial progenitor state. Priors were set on 4 of the parameters: parameters representing auto-activations must be positive as should be the parameter representing the regulation of *tbxta* from Wnt, to accommodate the ample evidence supporting the nature of this interaction in the literature (*4*). No other constraints were imposed on the parameters.

Markov Chain Monte Carlo was used to infer the parameter values that best recapitulated the experimental data. We used the emcee package for Python and ran the fitting algorithm independently 100 times. Each run returned an estimated 'best' a parameter set. The 100 parameter values were then simulated, compared to the data and re-simulated for cells that took 3, 4 and 5 hours respectively to undergo differentiation across the PSM from a progenitor state. Of the 100 initial networks, 56 recapitulated accurately the gene expression dynamics all the different scenarios.

We selected a network for our analysis by further imposing the sign of three interactions which are strongly supported by the literature. These are the activation of *tbx24* by Tbx16 (*2, 12*), the activation of *tbx16* by FGF (*3, 13*) and finally the inhibition of *tbx24* by Wnt (*12*) . These interactions weren't initially imposed as priors in order to not further bias the fitting, but were introduced at this stage to help with model selection to avoid choosing a network with interactions that might be knowingly wrong in sign. It is well known that networks with different topologies are often able to generate the same dynamics (*31*), making it difficult to differentiate between models without the aid of further evidence.

Out of these 56 networks, only one satisfied these three conditions imposed a posteriori:

network 11. In order to investigate how similar and therefore representative C11 is compared to the other 56 networks we used clustering of the 56 networks (Figure S2) demonstrating that they do not cluster well, and in fact are all broadly similar to one another.

Model analysis

The gene regulatory network model was formulated as a system of coupled ordinary equations. As such, it is amenable be analysed using the established tools and concepts from dynamical systems theory which we use to elucidate the mechanisms driving the observed dynamics (*31*) In addition, our system is rendered non-autonomous by the changing signalling Wnt and FGF environments, which we accommodate by calculating instantaneous phase portraits (*14, 15*).

Live modelling framework

In brief, our live-modelling framework simulates the gene regulatory network previously inferred from static tissue level expression pattern quantifications on every cell in the tracks, updating Wnt and FGF values dynamically as the cells' positions changes during PSM morphogenesis.

Cell tracks spanning four somite stages were obtained from live-imaging a developing zebrafish PSM using a multiphoton microscope. The position of each cell in the PSM was recorded in 3D space as an (x, y, z) coordinate every 2.5 minutes. The coordinate system was such that the origin (0,0,0) was set at the posterior-most tip of the PSM. The positive x-axis runs in the posterior to anterior direction through the middle of the PSM, the positive y-axis runs from centre to dorsal PSM and the positive z-axis, from centre to lateral PSM. The raw tracking data were manually modified to terminate tracks once these had entered a newly formed somite so that our dataset would only contain the tracks of cells within the PSM. Newly formed somites can be identified in the tracks by eye as soon as their posterior boundary forms.

The simulations are initialised using the same TBox gene expression and signalling profiles (Figure 1C; 1J) that were used to fit the gene regulatory network model. At the first time point, Tbox expression profiles (Figure 1C) are projected onto the normalised length of the PSM, and cells are assigned *tbxta*, *tbx16* and *tbx24* values according to their posterior to anterior position

(x-axis). These values will be used as the initial conditions from which to simulate the gene regulatory network model in every cell. The same methodology is used to assign each cell with initial Wnt and FGF values.

The model is allowed to run in every cell for the duration between time points in the tracks using the Wnt and FGF values previously assigned. At the next time point, the position of the cell is updated to the tracking data. The new length of the PSM is again normalised, signalling profiles are the projected onto the newly normalised PSM and cells will update their Wnt and FGF values again according to their new relative positions in the PSM. The model will run again simulating the time between tracking time points using the last simulated *tbxta*, *tbx16* and *tbx24* values as the initial conditions and the recently updated values for Wnt and FGF. This will be done iteratively until the last tracking time point. Cells which appear between time points, following cell division or a new track for example, are initialised in the same way cells were at the first time point.

References

1. J. Briscoe, S. Small, *Development* **142**, 3996 (2015).

2. C. M. Bouldin, *et al.*, *Development* **142**, 2499 (2015).

3. H. Goto, S. C. Kimmey, R. H. Row, D. Q. Matus, B. L. Martin, *Development* **144**, 1412 (2017).

4. B. L. Martin, D. Kimelman, *Developmental cell* **15**, 121 (2008).

5. O. F. Venzin, A. C. Oates, *Developmental biology* **460**, 40 (2020).

6. E. Moro, *et al.*, *Developmental Biology* **366**, 327 (2012).

7. B. Bénazéraf, *et al.*, *Nature* **466**, 248 (2010).

8. A. Mongera, *et al.*, *Nature* **561**, 1 (2018).

9. I. Ridelis, *et al.*, *FEBS Letters* **586**, 784 (2012).

10. H. Tsutsui, S. Karasawa, H. Shimizu, N. Nukina, A. Miyawaki, *EMBO reports* **6**, 233 (2005).

11. S. Habuchi, H. Tsutsui, A. B. Kochaniak, A. Miyawaki, A. M. van Oijen, *PLoS ONE* **3**, e3944 (2008).

12. R. Fior, *et al.*, *Development (Cambridge, England)* **129**, 3311 (2012).

13. K. Griffin, S. Amacher, C. Kimmel, D. Kimelman, *Development* **122**, 4119 (1998).

14. J. R. Collier, N. A. Monk, P. K. Maini, J. H. Lewis, *Journal of theoretical Biology* **183**, 429 (1996).

15. B. Verd, A. Crombach, J. Jaeger, *BMC systems biology* **8**, 1 (2014).

16. A. B. Webb, *et al.*, *eLife* **5**, e08438 (2016).

17. A. Kawamura, *et al.*, *Developmental Cell* **9**, 735 (2005).

18. A. Goldbeter, O. Pourquié, *Journal of Theoretical Biology* **252**, 574 (2008).

19. J. Dubrulle, M. J. McGrew, O. Pourquié, *Cell* **106**, 219 (2001).

20. S. A. Holley, H. Takeda, *Seminars in Cell Developmental Biology* **13**, 481 (2002).

21. A. Aulehla, O. Pourquié, *Current Opinion in Cell Biology* **20**, 632 (2008).

22. A. De Simone, *et al.*, *Nature* **590**, 129 (2021).

23. L. Busby, B. Steventon, *Preprints* (2020).

24. C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, *Developmental Dynamics* **203**, 253 (1995).

25. D. a. Turner, *et al.*, *Development* **141**, 4243 (2014).

26. H. M. Choi, *et al.*, *Development (Cambridge)* **145** (2018).

27. J. Schindelin, *et al.*, *Nature Methods* **9**, 676 (2012).

28. C. a. Schneider, W. S. Rasband, K. W. Eliceiri, *Nature Methods* **9**, 671 (2012).

29. E. Hirsinger, B. Steventon, *Journal of Visualized Experiments* **2017**, e55210 (2017).

30. E. Mjolsness, D. H. Sharp, J. Reinitz, *Journal of theoretical Biology* **152**, 429 (1991).

31. S. H. Strogatz, *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering* (CRC press, 2018).