

## Recurrent hyper-motif circuits in developmental programs

Miri Adler<sup>1,2,5,\*</sup> and Ruslan Medzhitov<sup>3,4,5,\*</sup>

<sup>1</sup> Department of Genetics, Silberman Institute of Life Science, Edmond J. Safra Campus, The Hebrew University of Jerusalem, Jerusalem, Israel

<sup>2</sup> Department of Immunology and Cancer Research, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

<sup>3</sup> Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA;

<sup>4</sup> Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut, USA

<sup>5</sup> Tananbaum Center for Theoretical and Analytical Human Biology, Yale University School of Medicine, New Haven, Connecticut, USA

\*Correspondence to:

[miri.adler@mail.huji.ac.il](mailto:miri.adler@mail.huji.ac.il)

[ruslan.medzhitov@yale.edu](mailto:ruslan.medzhitov@yale.edu)

### Abstract

During embryogenesis, homogenous groups of cells self-organize into stereotypic spatial and temporal patterns that make up tissues and organs. These emergent patterns are controlled by transcription factors and secreted signals that regulate cellular fate and behaviors through intracellular regulatory circuits and cell-cell communication circuits. However, the principles of these circuits and how their properties are combined to provide the spatio-temporal properties of tissues remain unclear. Here we develop a framework to explore building-block circuits of developmental programs. We use single-cell gene expression data across developmental stages of the human intestine to infer the key intra- and inter-cellular circuits that control developmental programs. We study how these circuits are joined into higher-level hyper-motif circuits and explore their emergent dynamical properties. This framework uncovers design principles of developmental programs and reveals the rules that allow cells to develop robust and diverse patterns.

## Introduction

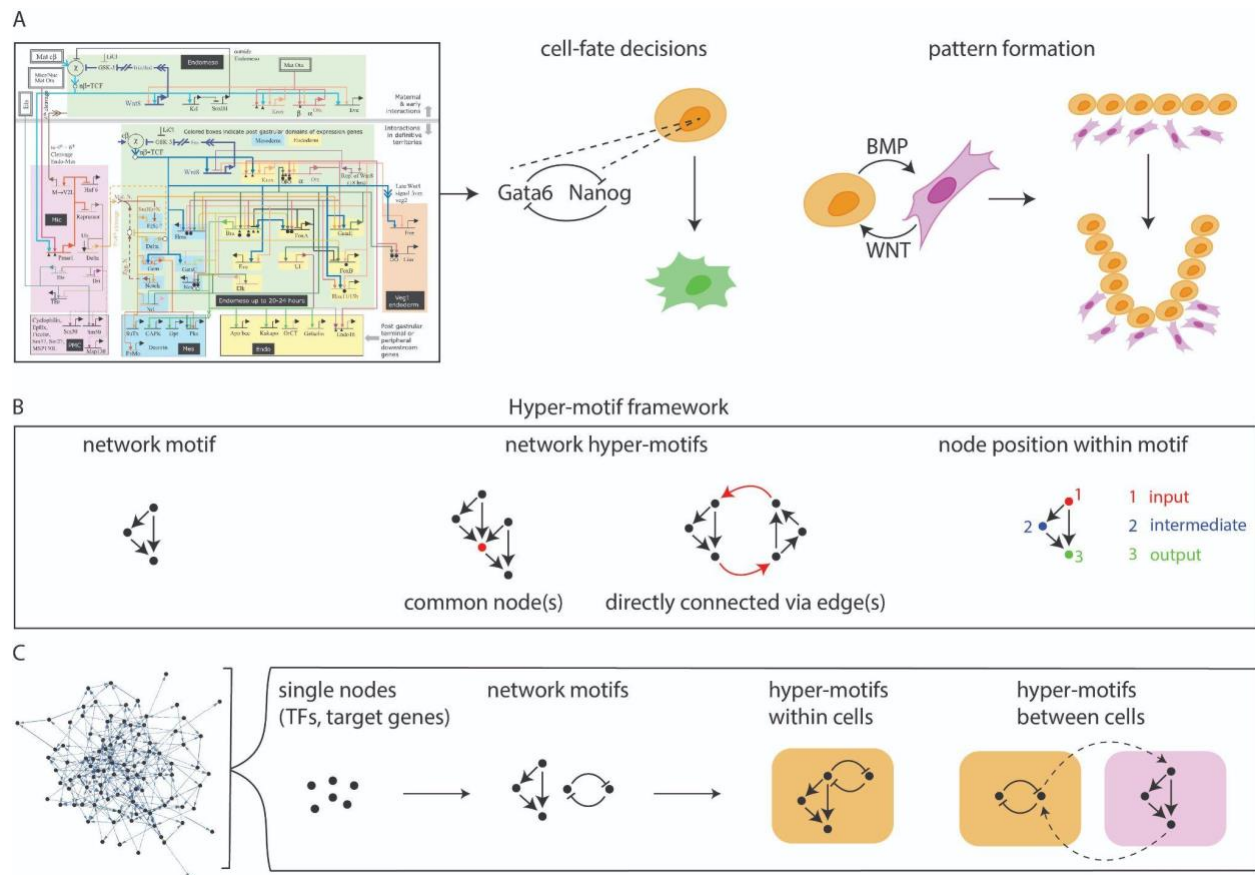
Embryonic development is a well-orchestrated process in which cells divide, migrate, and differentiate into the diversified tissues that form the embryo. Starting from a single fertilized egg, cells use communication signals that they secrete and sense to regulate their cell fate decisions<sup>1</sup>. These signals, also known as morphogens, produce gradients and spatial patterns that inform the cells of their position and allow them to respond accordingly<sup>2–6</sup>. There are several families of morphogens that play key roles in shaping the spatial patterns in the embryo including the bone morphogenic proteins (BMPs)<sup>7,8</sup>, fibroblast growth factors (FGFs)<sup>9</sup>, Hedgehog (HH)<sup>10</sup>, WNT<sup>11,12</sup> and Retinoic acid<sup>13</sup>. The production of these morphogens and their receptors is regulated by several transcription factor (TF) families that play critical roles in development, including HOX, GATA, ETS, SOX, PAX, TBOX, FOX, and E-proteins<sup>14–22</sup>. These morphogens and TFs regulate each other's expression through intracellular regulatory circuits as well as cell-cell communication circuits, forming a complex network of interacting genes<sup>23–29</sup> (Fig 1A). However, the rules that govern these interactions in regulating cell-fate decisions and forming tissue-level patterns remain poorly understood. Two complementary conceptual frameworks provide guiding principles governing cell fate decisions: Turing reaction-diffusion model and the concept of positional information<sup>30–35</sup>.

Studies in model organisms have shed light on key pathways in developmental processes, where a detailed transcription network can be inferred in specific contexts from experimental data (as exemplified in Fig 1A). However, it is unclear which circuits out of the detailed transcription network are the 'building blocks' required for the proper development of the organism, and how their dynamical properties explain the wide range of observed tissue-level patterns<sup>36</sup>.

Previous theoretical and experimental work has defined network motifs, which are recurrent patterns, as the building blocks of biological networks<sup>37,38</sup>. Network motifs such as the feedforward loop (FFL) have been extensively studied and their dynamical properties were elucidated<sup>39,40</sup>. Seminal previous work has uncovered several families of network motif designs that can provide distinct behaviors in the interpretation of morphogen gradients<sup>41</sup>. How the integration of the individual network motifs within the complex regulatory network influences their functionality is not yet understood, however. Integration of small network motif patterns may generate emergent properties that cannot be explained by each motif alone. Understanding complex systems such as developmental regulatory networks therefore require the understanding of the rules of interaction between the systems' building-block components. To address this, we recently defined hyper-motifs - higher-level network modules that describe how network motifs are

embedded in large networks<sup>42</sup>. Given two network motifs, we consider two ways in which they can be directly joined to form a hyper-motif: having at least one node that is common to both motifs and being directly linked in the network through at least one edge (Fig 1B). Hyper-motifs can generate novel properties that are not observed in each motif alone and can also be used to silence certain motifs when not needed<sup>43</sup>. Such properties are critical to ensure accurate choice of cell fate, developmental timing and spatial patterning during embryogenesis. It is therefore important to understand how network motifs form these higher-level functional modules.

Here, we develop a network analysis framework to study the underlying principles of regulatory and cell-cell interactions that control developmental programs. Going beyond previous studies that focused on network motifs, we explore how the different network motifs in developmental programs are wired together within cells and between different cells through morphogen signaling pathways (Fig 1C). This framework can help explain how cellular interactions that are governed by a minimal set of regulators show diverse, complex and precise emergent properties.



**Figure 1: A network analysis framework to explore developmental programs.** (A) Out of the multitude of regulatory interactions in gene regulatory networks (GRNs), what are the building block circuits of developmental programs for cell-fate decisions and pattern formation? (GRN

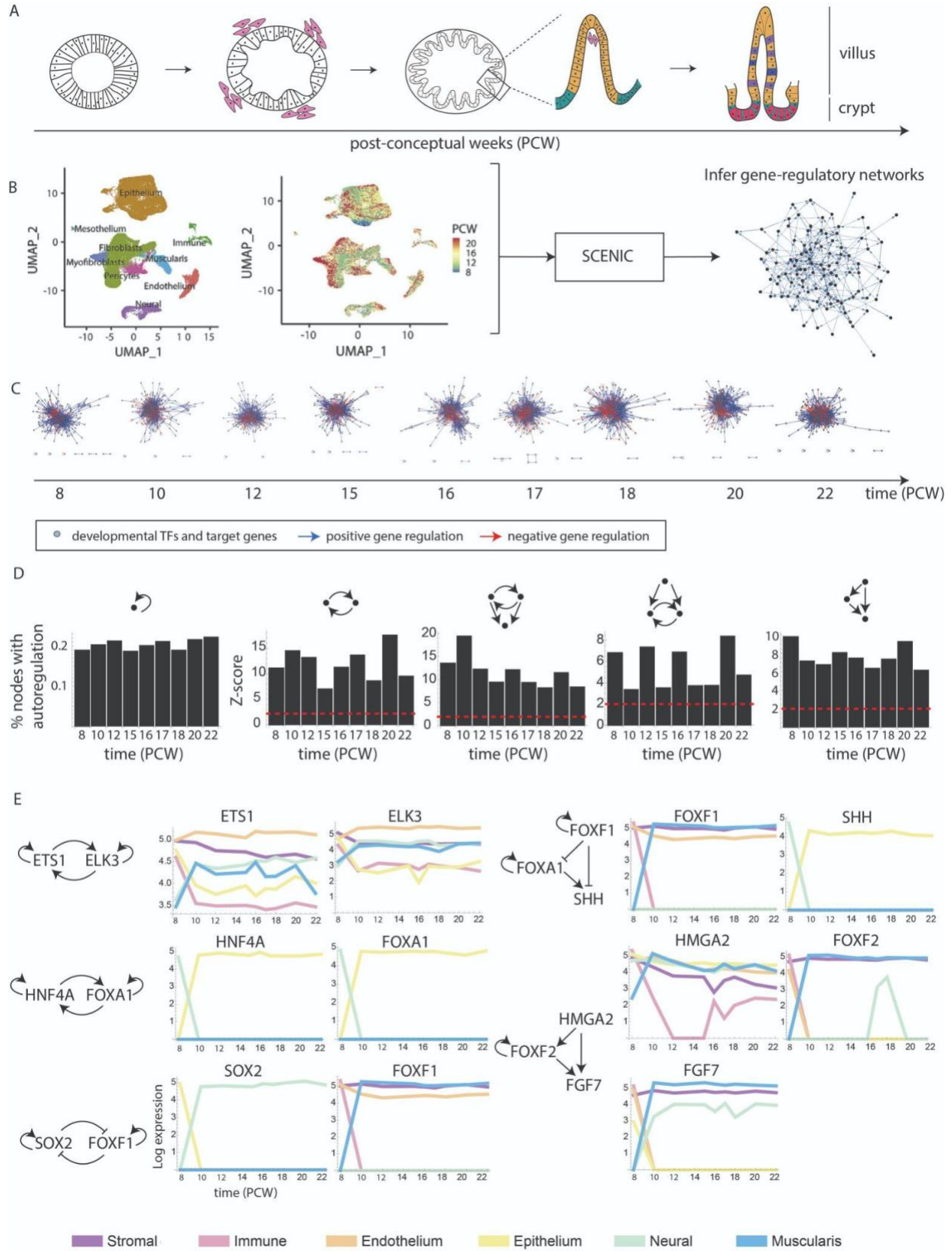
adapted from Davidson et al., 2002<sup>23</sup>). (B) Definition of the hyper-motif framework: given two network motifs, their hyper-motifs are defined as combinations of motifs where the motifs share at least one node or interactions of motifs where the motifs are directly linked through at least one edge. Distinguishing different node positions within the motif is important to explore different hyper-motif structures. (C) A computational framework to analyze regulatory networks by identifying network motifs as building-block circuits and how they are joined together to form hyper-motifs within and between communicating cells.

## Results

### Enriched network motifs in the development of the human intestine

We analyze single-cell RNA-sequencing (scRNAseq) data from the human intestine across 9 time points, spanning 8-22 post-conceptual weeks (PCWs), from Fawcner-Corbett et al.<sup>44</sup> (Fig 2A-B). We focus on developmental-related genes including developmental TFs, morphogens ligands, their antagonists, receptors and co-receptors. We used SCENIC<sup>45</sup> to infer regulatory interactions between the developmental transcription factors and their target genes (Fig 2C, Methods). The networks we inferred are such where we sum over all regulatory interactions between genes that are expressed across different cell types. The size of the networks remains roughly constant across time, with the number of genes ranging between 220 and 275 and number of edges between 752-1045 (Fig S1).

We find five network motifs (up to 3-node circuits, Methods) that are continuously enriched across all time points. The network motifs are: (1) autoregulation (mostly positive), (2) mutual feedback loop, (3) a regulated feedback loop, (4) a regulating feedback loop, and (5) a feedforward loop (FFL) (Fig 2D). These network motifs are also found in developmental programs of other organisms<sup>40</sup>. The dynamical properties of these 5 network motifs, including bi-stability and delayed response time, are crucial for developmental processes<sup>46,47</sup>. The network motifs we find in intestinal developmental programs can give rise to diverse gene expression patterns from ubiquitous genes expressed in all major cell types to genes expressed in specific cell types (Fig 2E).





**Figure 2: Five network motifs enriched in gene regulatory networks of developing intestine.** (A) Schematic of intestine development that is spanned by the data in Fawkner-Corbett et al.<sup>44</sup>. (B) We analyze the single-cell data and use SCENIC to infer gene regulatory interactions between developmental-related TFs and target genes. (C) The GRNs inferred from the data across 9 developmental timepoints. Nodes are TFs and target genes, edges represent positive (blue) and negative (red) interactions. (D) The most statistically enriched network motifs are shown with their Z-score. The red dashed line is a threshold for enrichment significance. For the autoregulation motif (left-most panel) the fraction of nodes in the network that are autoregulated is plotted. (E) Examples of network motifs inferred from 8 PCW. Log mean expression of genes in 6 major cell type categories is plotted as a function of time (PCW).

### **Developmental genes can be categorized based on their position in the network motifs**

To explore patterns in the network structure over time, we categorize the genes at each time point according to their position in the network motifs (or their motif role, Fig 1B) and explore whether genes change their positions across time or keep the same motif roles. Considering the 5 network motifs we find in the intestinal developmental programs, there are 7 unique motif roles that we partition the genes into. These motif roles are: autoregulated TFs; TFs in mutual feedback; input, intermediate, and output of an FFL; input to a regulated feedback; and output of a regulating feedback (Fig 3A). Categorizing the genes into these 7 motif roles, we find that the most abundant role in each network is the FFL's output (Fig 3B). Considering the proportions of genes that generally participate in either input, intermediate or output roles show that intermediate genes form a much smaller group than input and output genes, suggesting a bowtie structure in the networks<sup>48</sup>.

Next, we explore the robustness of categorizing genes according to their role in motifs by looking for changes in the identity of the TFs and genes participating in each motif role. To do so, we compute the overlap (Jaccard index, Methods) between the lists of genes participating in a certain motif role from two successive time points (Fig 3C). We find that autoregulation is the most robust motif role, where about 60% of the TFs that are autoregulated at a certain time point remain so in the next time point in the data. Other motif roles such as the input to a regulated feedback and the output of a regulating feedback are more variable where on average only 30% of the genes participating in these roles maintain their role in successive time points (Fig 3C).

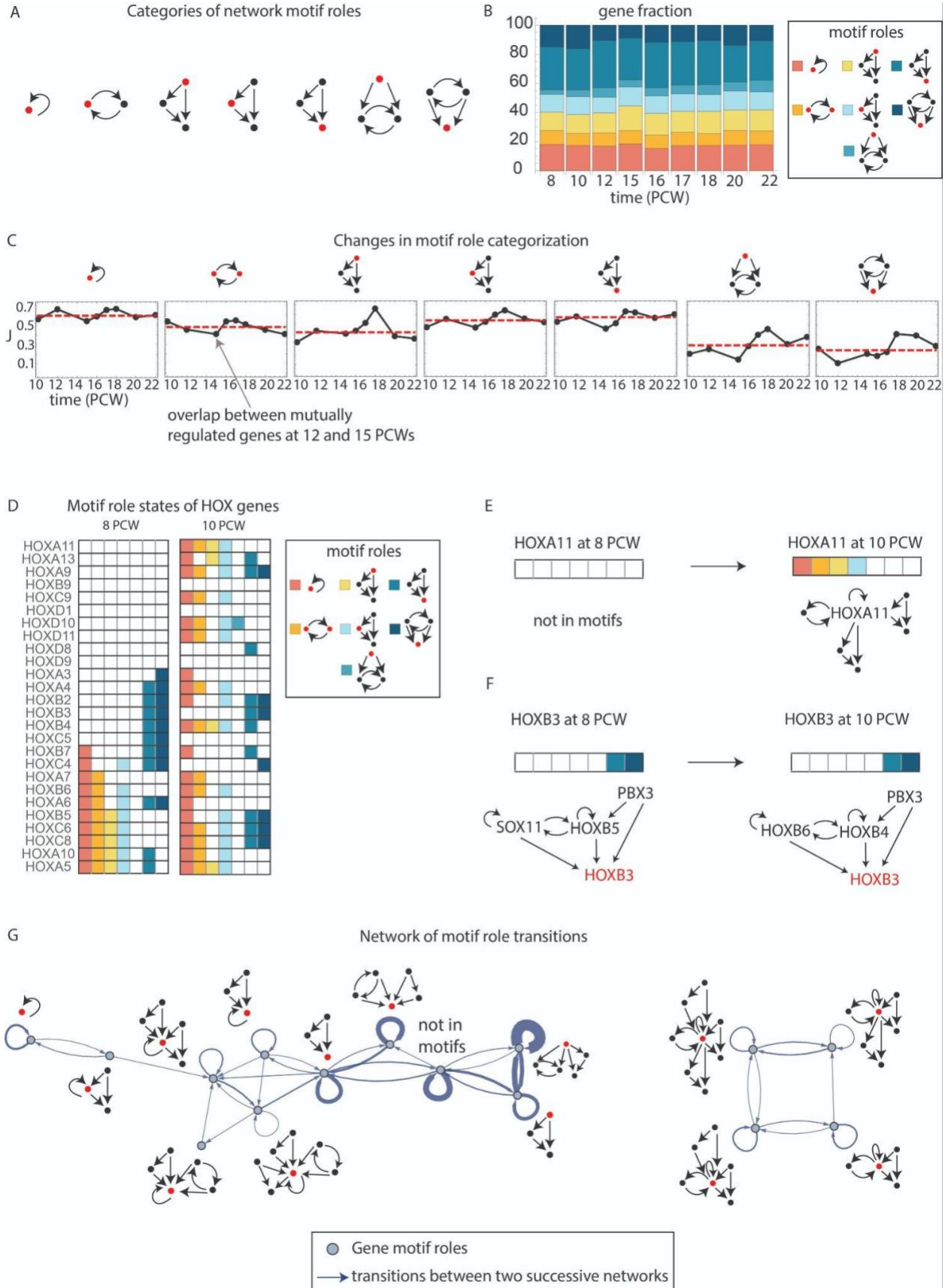
This analysis highlights time points where there are major changes in the identity of genes that participate in a certain motif role. We quantitatively define major changes by considering time points where the Jaccard index is smaller than the median across all

time points (red dashed line in Fig 3C). For example, the largest changes in the regulated feedback's input occur at 10, 12, and 15 PCWs (second panel to the right, Fig 3C).

Next, we analyze the TFs and genes that are added in those transition time points to a certain motif role to explore whether these genes are indeed associated with specific expression patterns or are thought to be important for the differentiation of a new cell type. Among the genes added at 10 PCW, we find BMP2, BMP4, and BMP5, which were reported to be expressed in pre-cryptal S2 fibroblasts (telocytes) and to promote epithelial proliferation and differentiation<sup>44</sup>. We also find FOXD1 that controls the terminal ileum, TCF21 which is important for fibroblast early development, and TWIST2 that is involved in key TF networks delineating muscularis cells. FOXF2 which is important for muscle cell differentiation appears in a new motif role at 12 PCW. At 15 PCW, we find several epithelial-restricted genes including LEFTY1 and IHH which are expressed in an epithelial-specific module. At 16 PCW we find WNT5B, which is expressed by stromal, neuronal and muscle cells and supports epithelial cells. At 20 PCW we find FOXL1 which is a marker for S2 fibroblast differentiation, and WNT2B which is expressed in a myofibroblast module. At 22 PCW we find SOX13 which is important for intestinal angiogenesis<sup>44</sup>.

Next, we asked whether changes to gene motif roles across time are random or follow certain rules. To answer this, we define a gene state based on its motif roles (hereby called motif role state) where a gene can participate in one or multiple roles simultaneously (forming a hyper-motif as discussed next). For each gene participating in network motifs at least once, we follow how its motif role state changes over time.

For example, considering HOX genes in our inferred networks (Fig 3D), we find that some HOX genes maintain the same motif role state overtime (Fig 3F) and some transition into new roles (Fig 3E, Fig S2). To visualize the possible transitions between motif role states, we build a network of motif role state transitions (Fig 3G). In this network, the nodes represent a motif role state, and edges represent transitions between states from two successive time points. The weights of the edges are based on the frequency of the transitions. Exploring the patterns of motif role transitions, we find that there are specific rules regarding which role a specific gene can transition into based on its current role within the network. For example, we find that input genes do not transition directly into output genes and vice versa. We also find that TFs that participate in mutual feedback are usually found to be hubs, where they serve as common nodes of multiple network motifs (Fig 3G).

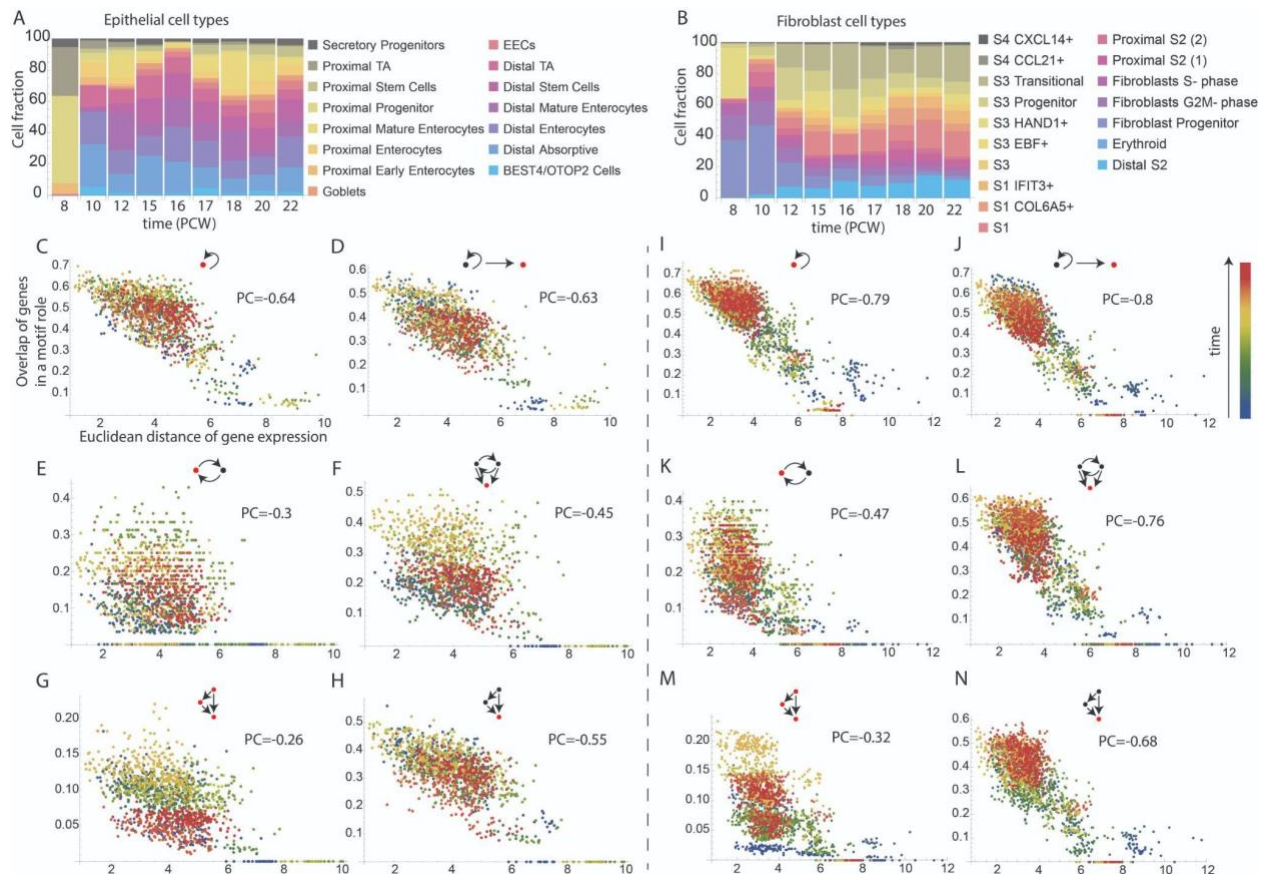




**Figure 3: Categorizing developmental genes by motif roles highlights developmental transitions.** (A) Developmental genes can be categorized to 7 motif roles (highlighted in red). (B) Fraction of genes participating in each motif role at each time point. (C) In each panel the black points are the Jaccard index between the two sets of genes participating in a particular motif role from two successive time points. The red dashed line is the median Jaccard index across all time points. (D) Table of motif role states of HOX genes at 8 and 10 PCWs. (E) HOXA11 transitions from not participating in motifs to a node common to multiple network motifs combined. (F) HOXB3 maintains the same motif role state between 8 and 10 PCWs despite other genes in the motif that change their regulatory roles. (G) Network of motif role transitions. Nodes in the network represent motif role states and edges represent transitions between two successive networks. Only transitions that occur more than 2 times are considered.

### Network motifs as predictors of gene expression profile

To assess the role of the network motifs we find in developmental programs in regulating gene expression, we next explore the relationship between the identity of network motifs and similarity in gene expression profiles. To do so, we focus on the epithelial and fibroblast cell type populations and consider the specific cell types epithelial and fibroblast progenitor cells diversify into (Fig 4A-B). Within each cell type category, we calculated the Euclidean distance between every pair of cell types (e.g. secretory progenitors and goblet cells, proximal enterocytes and proximal stem cells, etc.) in gene expression space considering their average gene expression profiles across the entire transcriptome (Methods). We then compare this gene expression distance measure to the similarity in the network motifs that we inferred regulating these cells' gene expression. We estimate the similarity of network motifs by computing the overlap of lists of genes participating in or regulated by a specific network motif between every pair of cells (Methods). We find a large correlation between gene expression distance and overlap in network motif genes where the more network motifs two cell types share, the more similar their gene expression profiles are (Fig 4C-N). This correlation is remarkable given the fact that examining whether a group of ~20 TFs are autoregulated or not, provides a strong prediction on the similarity in the gene expression state of ~20,000 genes (Fig 4C, I). We find that some network motifs are better predictors than others. For example, for epithelial cells, autoregulation is more correlated with gene expression than mutual feedback and feedforward loops (Fig 4E,G). However, focusing on effector genes - the identity of genes that are directly regulated by the motifs, improves the correlation (Fig 4D,F,H). Additionally, we find that fibroblast cell types show larger correlations than epithelial cells across all network motifs and their effector genes we examined.



**Figure 4: High correlations between network motif identity and gene expression distance in epithelial and fibroblast cell types.** (A-B) Cell fraction of epithelial (A) and fibroblast (B) cell types across time. (C-N) Every point is a pair of epithelial (C-H) or fibroblast (I-N) cell types, where the y axis is the overlap (Jaccard index) between genes participating in a network motif or its output gene in the two cell types and the x axis is the Euclidean distance between the gene expression profiles of the two cell types. The overlap in genes is computed for the nodes marked in red in each panel. The color of the points in the plots represents the developmental stage from early stage (8 PCW - blue) to late stage (22 PCW - red). PC=Pearson correlation.

### The recurrent network hyper-motifs show common patterns

We next compute the most statistically enriched hyper-motifs<sup>42</sup>, where multiple motifs are combined through at least one common node (Methods). The most enriched hyper-motifs show several common patterns. Autoregulation is usually coupled to feedback circuits or to FLLs through their intermediate and output nodes. Feedback circuits are combined with FLLs through their intermediate nodes. Motifs share a common input or output node (Fig 5A).

Some combinations are statistically enriched across all developmental time points, while others are significantly enriched only at certain time points. Assuming that some combinations of motifs may lead to new properties, time points with enrichment of hyper-motifs suggests emergence of new cell types or tissue-level patterns. The most enriched hyper-motif topologies appear at 10 PCW, which is indeed the time where the crypt to villus axis is formed (Fig 5A).

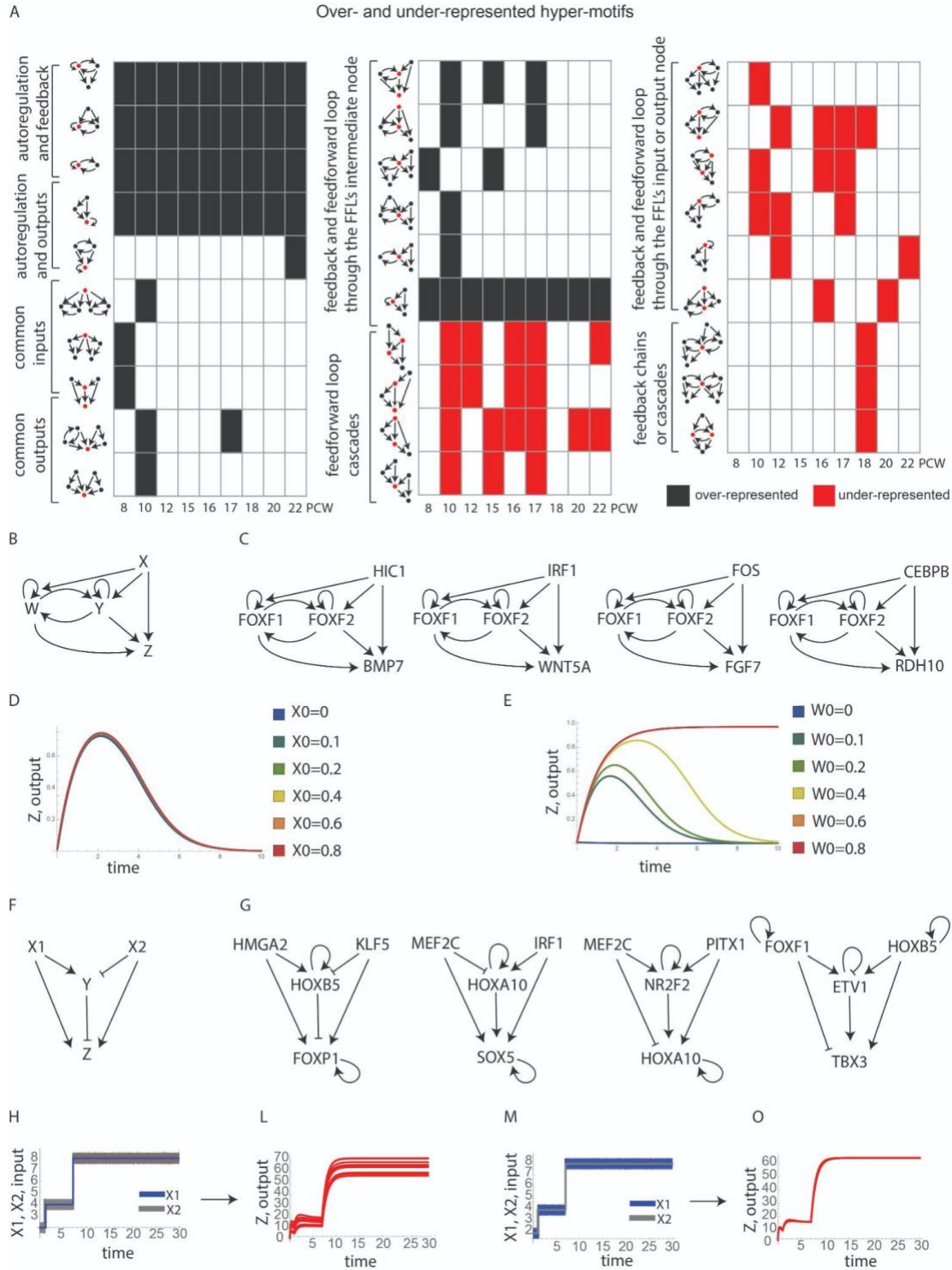
The under-represented network hyper-motifs also follow certain patterns. Autoregulation and mutual feedback do not combine with the FFL's input node, combinations of multiple feedback circuits are avoided only at 18 PCW, and cascades of FFLs are also excluded in multiple time points (Fig 6B).

### **Dynamical properties of enriched hyper-motif regulatory circuits**

We consider several hyper-motif regulatory circuits that we find to be statistically enriched in the networks and mathematically model their dynamical behavior. One of the enriched hyper-motifs is a combination of a feedback circuit and an FFL through the intermediate node of the FFL (Fig 5B). This hyper-motif is very frequent in the regulatory networks of stromal and epithelial cells throughout the developmental process where it directly controls the expression levels of morphogen ligands and receptors as well as various development-related TFs (Fig 5C). We developed a mathematical model for the dynamical behavior of this hyper-motif (Methods). Our model suggests an interesting property of this hyper-motif where there are two types of input signals affecting the output gene. The first input signal is the FFL's input  $X$ , and the second class of input signal are the  $Y/W$  TFs which mutually regulate each other. While both inputs are assumed to be required for the transcription of the output gene ( $Z$ ) in our model, they differ in their effect on the dynamical behavior of  $Z$ . The  $X$  input signal is required to turn on  $Z$ , but its initial or final levels do not influence the dynamical behavior of  $Z$ . This means that  $Z$  is insensitive to variations in the levels of  $X$  (Fig 5D). In contrast, the initial levels of  $Y$  or  $W$  have a large effect on the dynamical behavior of  $Z$  (Fig 5E). This sensitivity to the  $Y/W$  levels provides a way of generating different responses to varying levels of inputs using a simple regulatory circuit, thus fulfilling the requirement for the interpretation of morphogen gradients, for example.

We also consider a hyper-motif which joins two types of FFL circuits: coherent and incoherent FFLs (Fig 5F). This circuit is seen for example at 8 PCWs in fibroblast cells where the HMGA2-HOXB5-FOXP1 circuit is an incoherent FFL (HMGA2 activates FOXP1 directly but inhibit it through HOXB5), whereas the KLF5-HOXB5-FOXP1 circuit is a coherent FFL (Fig 5G). This hyper-motif can exhibit known properties of each motif when the input signals vary separately. When the two input signals are changing simultaneously, this motif presents an interesting behavior where the final level of the

output Z is robust to variations in the input signal regulating the incoherent FFL (X1) but is sensitive to noisy inputs stimulating the coherent FFL (X2) (Fig 5H-O).



**Figure 5: Enriched network hyper-motifs and their dynamical behavior.** (A) Table of over- and under-represented hyper-motifs across the developmental process. Rows are different hyper-



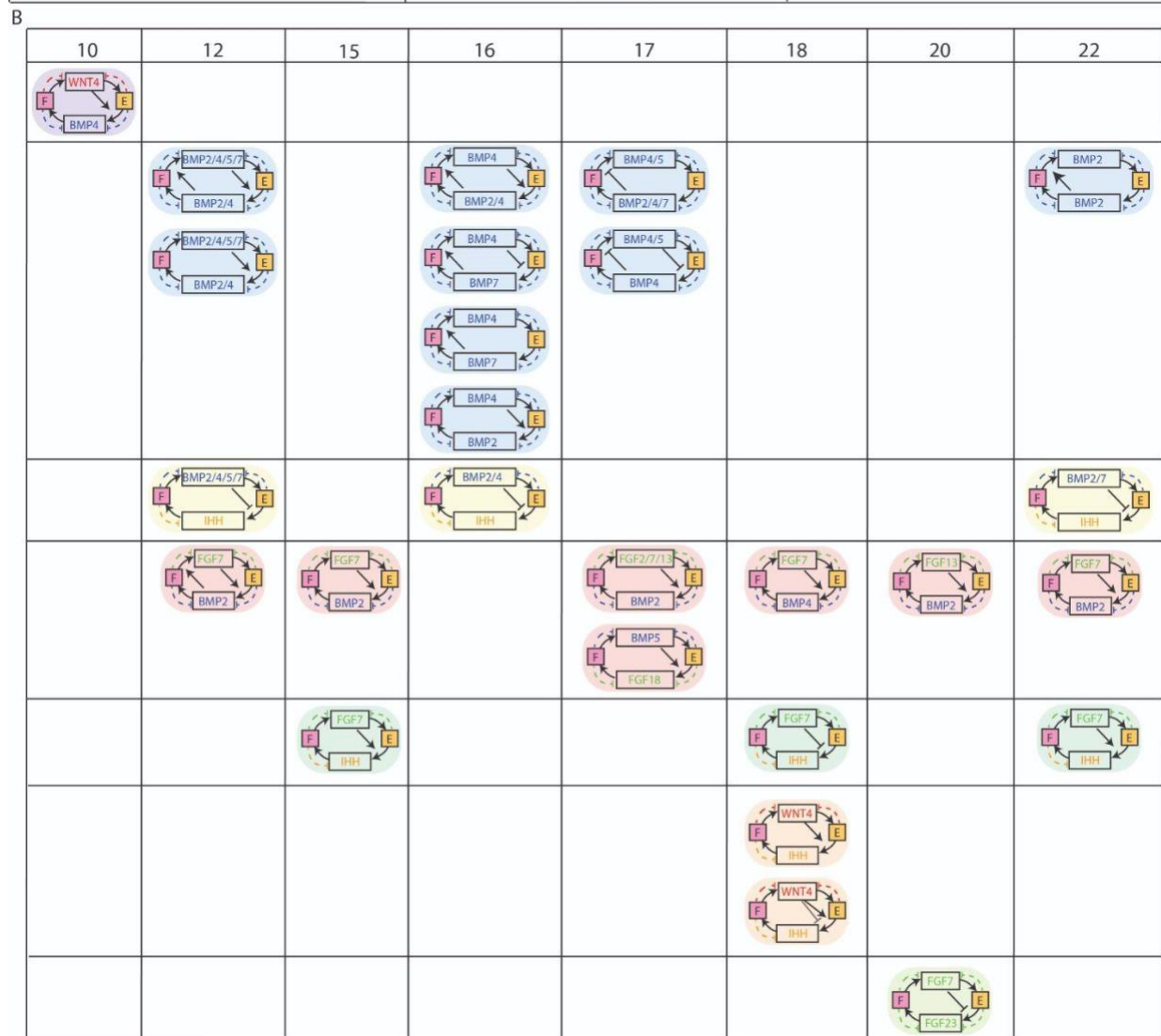
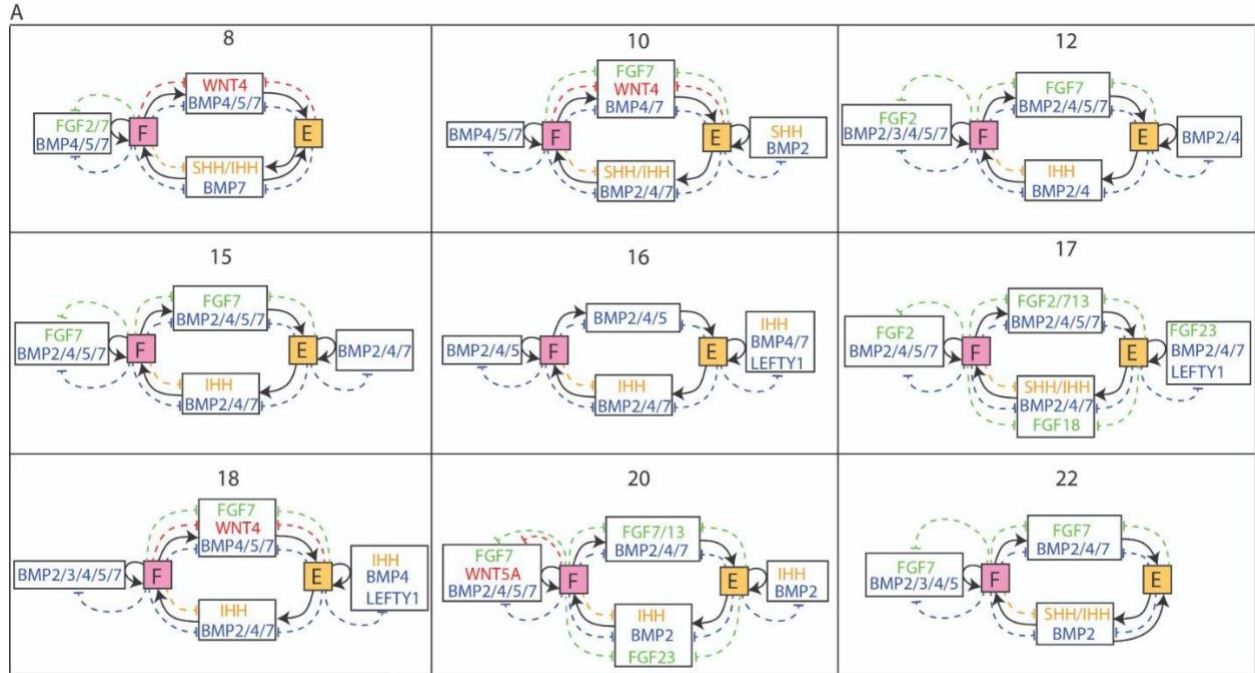
motif topologies; columns are different time points in the data. Black squares mark positive enrichment, and red squares mark negative enrichment. (B) An enriched hyper-motif circuit where input signal X activates output gene Z and the mutually regulating TFs Y and W which both activate Z as well. (C) Examples of the hyper-motif in the developmental networks. (D-E) Dynamical behavior of the output gene Z where the different curves are for different initial conditions of the input signal X (D) or the TF W (E). (F) Hyper-motif circuit that combines a coherent FFL and an incoherent FFL. (G) Examples of the hyper-motif in the developmental networks. (H) Simultaneous step-like changes in X1 (noisy, gray) and X2 (precise, blue) (L) Dynamical behavior of output Z in response to the input signals plotted in H. (M-O) Same as H-L only with X1 as the noisy signal and X2 as the precise signal.

### **Cell-cell hyper-motif circuits emerge through morphogen signaling pathways**

Next, we explore how regulatory circuits within cells are linked through morphogen signaling pathways creating hyper-motif circuits between communicating cells. For the purpose of cell communication circuit analysis, we focus on fibroblast and epithelial cell type populations, the two cell types playing important roles in governing intestinal development<sup>49</sup>. First, we identify pairs of interacting cells where we consider interactions through ligand-receptor pairs of the developmental morphogen families (BMP, WNT, HH, FGF, and retinoic acid). To identify pairs of interacting cells, we used NicheNet, a computational method that infers ligand-receptor interactions considering expression patterns of ligands, receptors and downstream signaling genes in the receiving cells<sup>50</sup> (Methods). Using NicheNet, we highlight the top ligands that are predicted to link fibroblast and epithelial cell types (Fig 6A). Throughout the developmental process, we find that fibroblasts and epithelial cells communicate reciprocally through various BMP ligands; whereas IHH and SHH production is restricted to epithelial cells and its receptors and antagonists are produced by fibroblasts. Other morphogens, such as WNTs, appear at specific time points (Fig 6A).

Next, we use these cell communication circuits to infer regulatory networks linking perception of certain morphogens to production of other morphogens. We focus on reciprocal circuits composed of provisions of pairs of signaling pathways. Out of these circuits, we look for situations where gene regulatory interactions that are downstream to a certain morphogen signaling pathway in the receiving cell directly upregulate or downregulate expression of a morphogen ligand (Fig 6B). Following these morphogen circuits over time, we find several patterns that recur in multiple time points. For example, downregulation of IHH by BMP signaling appears at 12, 16, and 22 PCWs.

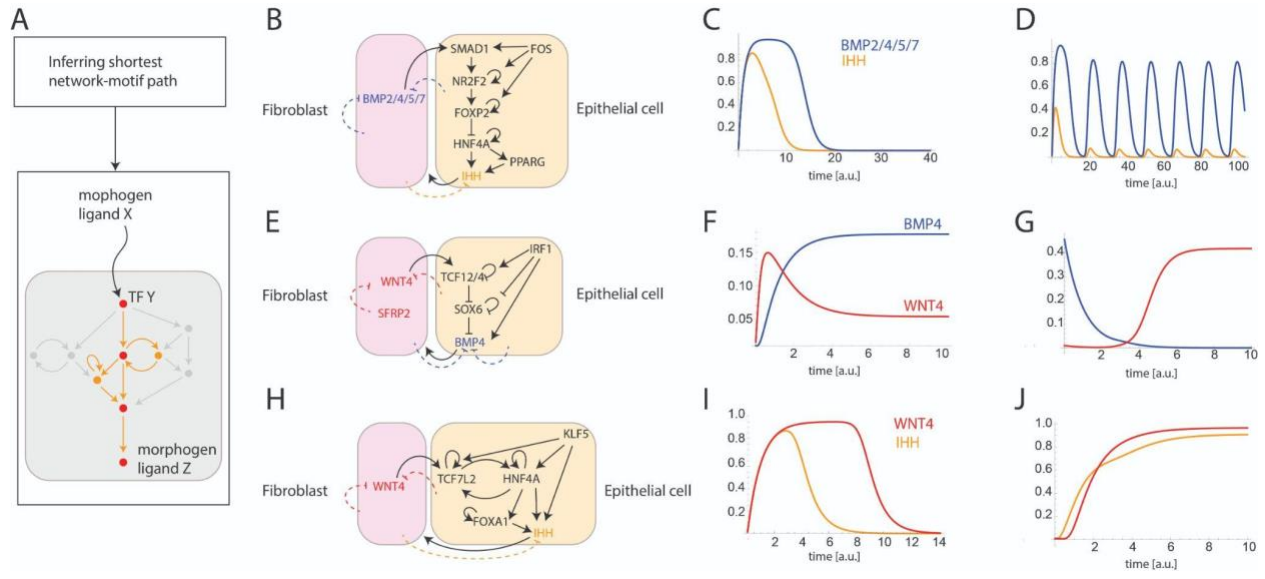




**Figure 6: Inferring hyper-motif circuits between communicating cells through ligand-receptor interactions.** (A) Cell-circuits between the epithelial (E, orange square) and fibroblast (F, pink square) cell type populations through morphogen and growth factor (GF) provision from 8-22 PCWs. Black arrows from cells to GFs represent ligand production. Black arrows from GFs to cells represent binding to cognate receptors. Dashed colored inhibitory arrows represent inhibition of GFs through antagonist expression. (B) Table of cell circuits where expression of morphogen ligands is directly regulated by a regulatory circuit downstream to a morphogen signaling pathway. Black arrows pointing from GF1 to another GF's production arrow, GF2, represent activation or inhibition through a gene regulatory circuit that is downstream to the GF1 signaling pathway in the cell type producing the GF2 signal.

### Cell-cell hyper-motif circuits can provide emergent dynamical properties

We consider three examples of hyper-motif regulatory circuits between communicating cells and examine their potential dynamical behavior in more detail. To do so, we consider the shortest path in the gene regulatory network linking the TF (Y) downstream to the input morphogen signaling pathway (X) received by the cell and the output morphogen ligand (Z) that is produced by the same cell (Fig 7A). Since the genes participating in this shortest path are part of a larger complex gene regulatory network, we developed a new method to extract the circuit topology controlling the expression of the output gene Z. Instead of simply considering a cascade of genes following the shortest path between X and Z in the network, we are inferring the '*shortest network motif path*'. Using the shortest network motif path approach, we follow the network motifs in which the genes linking X and Z participate in (Fig 7A, Methods). After inferring the shortest network motif path between a pair of morphogen signals, we developed a mathematical model to simulate the dynamical behavior of the resulting hyper-motif circuit (Methods). We find that interactions between members of BMPs and IHH (Fig 7B) can provide pulsatile response (Fig 7C) and pure oscillations (Fig 7D). Interactions between BMP4, WNT4 and its inhibitor SFRP2 (Fig 7E) can yield antagonistic expression patterns for BMP4 and WNT4 depending on model parameters (Fig 7F-G). The hyper-motif in epithelial cells that regulates IHH and is activated by WNT4 (Fig 7H) can provide pulsatile (Fig 7I) or delayed and sustained (Fig 7J) response.

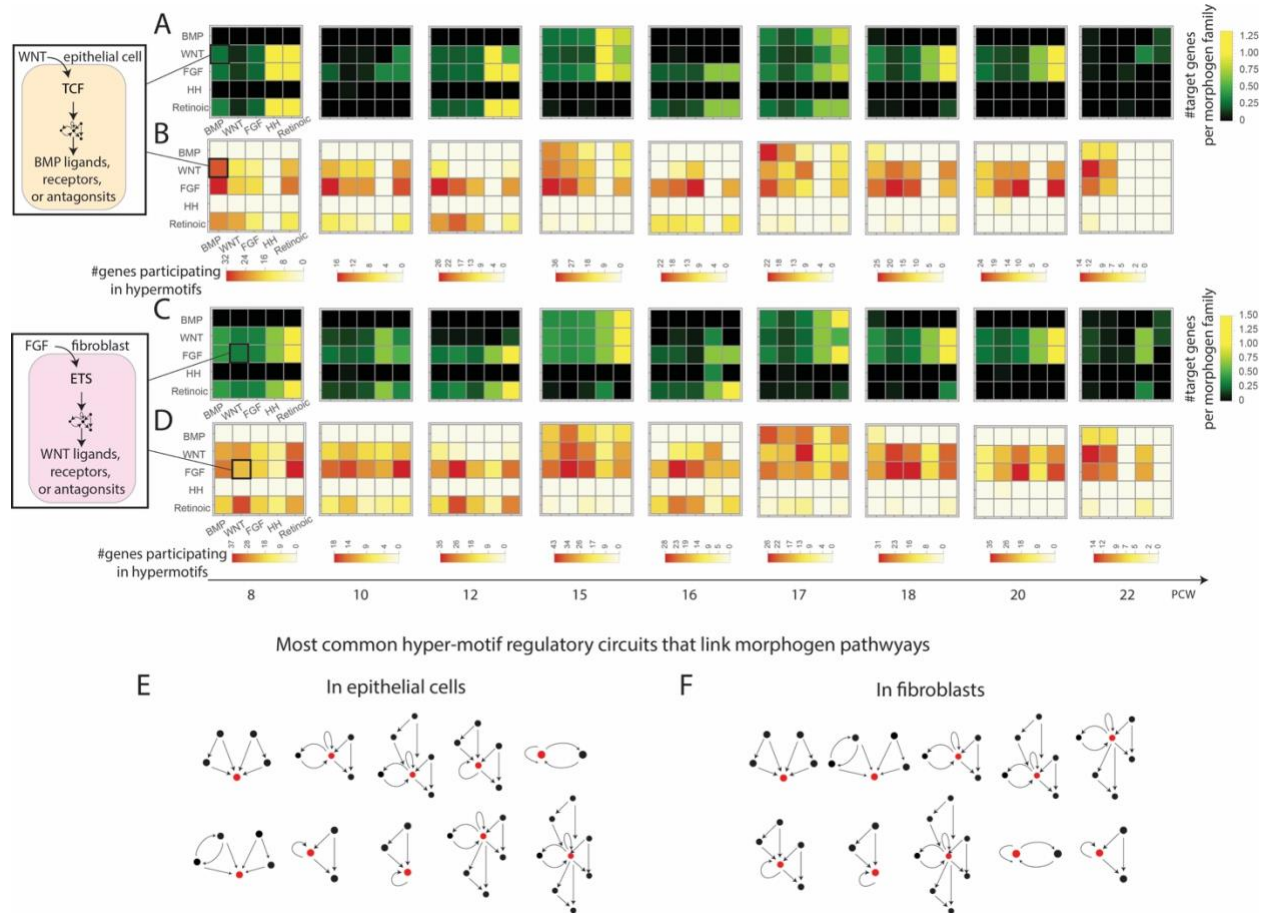


**Figure 7: Dynamical properties of cell-cell hyper-motif circuits.** (A) Illustration of our method to infer the shortest network motif path linking input signal X and output gene Z. Nodes participating in the shortest path are marked in red. Edges and additional nodes participating in the shortest network motif path are marked in orange. (B-J) Examples of three hyper-motif gene regulatory circuit topologies (B, E, H) linking two morphogen signaling pathways between fibroblast and epithelial cells. Dashed arrows represent production of morphogen antagonists. (C-D, F-G, I-J) Simulated dynamical behavior of the circuits in B, E and H with varying parameters or initial conditions (Methods).

### Hyper-motif regulatory circuits frequently regulate morphogen production

Next, we explore the extent to which hyper-motif regulatory circuits regulate morphogen expression directly through regulation of morphogen ligands and indirectly, through regulation of the morphogen receptors and antagonists. For every pair of morphogen families, we examine whether there is a direct path in the gene regulatory network linking one morphogen signaling pathway to expression of another morphogen's ligand, receptor or antagonist. Counting the number of target genes belonging to a certain morphogen family that are regulated by another morphogen signaling show large variation between the pairs of morphogens, as well as variation in the pattern over time (Fig 8A, C). Additionally, we explore the number of genes, within the regulatory paths linking pairs of morphogen signaling, that participate in network hyper-motif circuits. We find that most regulatory paths linking morphogen pairs are enriched with hyper-motif regulatory circuits (Fig 8B, D). Examining the composition of the most common hyper-motif regulatory circuits within these paths, we find that many of the regulatory circuits regulating expression of morphogen signaling correspond to the most statistically enriched hyper-

motifs in the network (Fig 8E-F; Fig 5A), suggesting that these enriched hyper-motif circuits indeed play important roles in regulating morphogen expression patterns.



**Figure 8: Role of hyper-motif circuits in regulating morphogen signaling pathways.** (A, C) Heatmaps of the number of target genes that are regulated downstream to morphogen signaling pathways in epithelial cells (A) and in fibroblasts (C). A square in X row and Y column is colored by the number of ligands, receptors and antagonists belonging to morphogen family Y that are found to be directly regulated in response to morphogen ligands from the X family, divided by the total number of ligands in the Y family. (B, D) Heatmaps of the number of genes in the regulatory circuits described in (A) and (C) that participate in more than one motif role (forming a hyper-motif) in epithelial cells (B) and fibroblasts (D). (E-F) The 10 most common hyper-motifs regulatory circuits linking morphogen pathways in epithelial cells (E) and fibroblasts (F).

## Discussion

Developmental processes are regulated by complex regulatory networks. To disentangle this complexity, we used our recently developed framework exploring how building-block circuits are joined in complex networks to reveal design principles of developmental programs. By applying our approach to intestinal development single-cell data, we found



that developmental GRNs show 5 network motifs that are joined in the network according to specific rules. We find that network motif identity and position of genes within motifs are good predictors for gene expression profiles for certain network motifs within epithelial and fibroblast populations, supporting the validity of the network motif approach to examine gene regulatory networks. Examining major transitions in the network structure, including emergence of enriched network hyper-motifs and appearance of new genes in key positions within the network, highlight major developmental events and their potential regulators. Lastly, we find that morphogen ligands, receptors and co-receptors are regulated directly by combinations of network motifs creating hyper-motif circuits between communicating cells that can generate emergent dynamical properties.

Our approach offers a new way to categorize developmental genes according to their roles in the network motifs. This categorization distinguishes between input genes, effector genes and hub genes that are core regulators combining multiple network motifs. It will be interesting to apply this method to developmental processes of other tissues to explore the robustness of gene categorization based on network motif roles.

In our analysis of over- and under-represented network hyper-motifs, we discovered that positive feedback circuits, including autoregulatory and mutual feedback, are highly enriched throughout the developmental process, whereas cascades of feedforward loops are statistically scarce. This finding is surprising, given that both circuit designs provide a delayed response which is an important property in developmental programs. The potential selection against cascades of FFLs may be due to the fact that FFL cascades provide delay in response time at the cost of noise amplification, whereas positive feedback circuits provide delay and sensitivity to input signals without compromising on the ability to buffer noise propagation<sup>51,52</sup>.

We note that our approach provides theoretical predictions on the mesoscale structure of developmental gene regulatory networks at a systems level. Future work where predictions regarding particular TF and target gene positions and roles within the networks are experimentally explored will be integral to complement this approach.

### **Acknowledgments:**

We would like to thank Eviatar Weizman from the Mantoux Bioinformatics institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, for providing support in the application of the NicheNet approach. We would like to thank Scott Pope for supporting initial TF binding motif analysis. We thank all current and former members of the Medzhitov lab as well as David Fawcner-Corbett and Agne Antanaviciute for discussion on this project.



## **Funding:**

This work was supported by the Howard Hughes Medical Institute, the Tananbaum Center for Theoretical and Analytical Human Biology, the Blavatnik Family Foundation, the Food Allergy Science Initiative, and a grant from NIH to R.M. (AI144152–01). M.A. was supported by the EMBO Long-Term Fellowship (ALTF 304-2019), the Zuckerman STEM Leadership program, and the Israel National Postdoctoral Award Program for Advancing Women in Science.

## References

1. Shahbazi, M. N. Mechanisms of human embryo development: from cell fate to tissue shape and back. *Development* **147**, dev190629 (2020).
2. Gilmour, D., Rembold, M. & Leptin, M. From morphogen to morphogenesis and back. *Nature* **541**, 311–320 (2017).
3. Kondo, S. & Miura, T. Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation. *Science* **329**, 1616–1620 (2010).
4. Cohen, M., Briscoe, J. & Blassberg, R. Morphogen interpretation: the transcriptional logic of neural tube patterning. *Curr. Opin. Genet. Dev.* **23**, 423–428 (2013).
5. Beccari, L., Marco-Ferreres, R. & Bovolenta, P. The logic of gene regulatory networks in early vertebrate forebrain patterning. *Mech. Dev.* **130**, 95–111 (2013).
6. Kicheva, A. & Briscoe, J. Control of Tissue Development by Morphogens. *Annu. Rev. Cell Dev. Biol.* **39**, 91–121 (2023).
7. Bragdon, B. *et al.* Bone Morphogenetic Proteins: A critical review. *Cell. Signal.* **23**, 609–620 (2011).
8. Chen, D., Zhao, M. & Mundy, G. R. Bone Morphogenetic Proteins. *Growth Factors* **22**, 233–241 (2004).
9. Turner, N. & Grose, R. Fibroblast growth factor signalling: from development to cancer. *Nat. Rev. Cancer* **10**, 116–129 (2010).
10. Briscoe, J. & Théron, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14**, 416–429 (2013).
11. Steinhart, Z. & Angers, S. Wnt signaling in development and tissue homeostasis. *Development* **145**, dev146589 (2018).
12. Logan, C. Y. & Nusse, R. THE WNT SIGNALING PATHWAY IN DEVELOPMENT AND DISEASE. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
13. Niederreither, K. & Dollé, P. Retinoic acid in development: towards an integrated view. *Nat. Rev. Genet.* **9**, 541–553 (2008).
14. Carnesecchi, J., Pinto, P. & Lohmann, I. Hox transcription factors: an overview of multi-step regulators of gene expression. *Int. J. Dev. Biol.* **62**, 723–732 (2018).
15. Tremblay, M., Sanchez-Ferras, O. & Bouchard, M. GATA transcription factors in development and disease. *Development* **145**, dev164384 (2018).
16. Wasylyk, B., Hahn, S. L. & Giovane, A. The Ets family of transcription factors. *Eur. J. Biochem.* **211**, 7–18 (1993).
17. She, Z.-Y. & Yang, W.-X. SOX family transcription factors involved in diverse cellular events during development. *Eur. J. Cell Biol.* **94**, 547–563 (2015).
18. Golson, M. L. & Kaestner, K. H. Fox transcription factors: from development to disease. *Development* **143**, 4558–4570 (2016).
19. Sebé-Pedrós, A. & Ruiz-Trillo, I. Chapter One - Evolution and Classification of the T-Box Transcription Factor Family. in *Current Topics in Developmental Biology* (ed. Frasch, M.) vol. 122 1–26 (Academic Press, 2017).

20. Lang, D., Powell, S. K., Plummer, R. S., Young, K. P. & Ruggeri, B. A. PAX genes: Roles in development, pathophysiology, and cancer. *Biochem. Pharmacol.* **73**, 1–14 (2007).
21. Afzal, Z. & Krumlauf, R. Transcriptional Regulation and Implications for Controlling Hox Gene Expression. *J. Dev. Biol.* **10**, 4 (2022).
22. Bergsland, M. *et al.* Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* **25**, 2453–2464 (2011).
23. Davidson, E. H. *et al.* A Genomic Regulatory Network for Development. *Science* **295**, 1669–1678 (2002).
24. Balaskas, N. *et al.* Gene Regulatory Logic for Reading the Sonic Hedgehog Signaling Gradient in the Vertebrate Neural Tube. *Cell* **148**, 273–284 (2012).
25. Barolo, S. & Posakony, J. W. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev.* **16**, 1167–1181 (2002).
26. Cohen, M., Page, K. M., Perez-Carrasco, R., Barnes, C. P. & Briscoe, J. A theoretical framework for the regulation of Shh morphogen-controlled gene expression. *Development* **141**, 3868–3878 (2014).
27. Davidson, E. H. Emerging properties of animal gene regulatory networks. *Nature* **468**, 911–920 (2010).
28. Exelby, K. *et al.* Precision of tissue patterning is controlled by dynamical properties of gene regulatory networks. *Development* **148**, dev197566 (2021).
29. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci.* **102**, 4936–4942 (2005).
30. Green, J. B. A. & Sharpe, J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* **142**, 1203–1211 (2015).
31. Wolpert, L. Chapter 6 Positional Information and Pattern Formation. in *Current Topics in Developmental Biology* (eds. Moscona, A. A. & Monroy, A.) vol. 6 183–224 (Academic Press, 1971).
32. Turing, A. M. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London Series B*, 237(641), 37-72. (1952).
33. Heller, E. & Fuchs, E. Tissue patterning and cellular mechanics. *J. Cell Biol.* **211**, 219–231 (2015).
34. Briscoe, J. & Small, S. Morphogen rules: design principles of gradient-mediated embryo patterning. *Development* **142**, 3996–4009 (2015).
35. Čapek, D. & Müller, P. Positional information and tissue scaling during development and regeneration. *Development* **146**, dev177709 (2019).
36. Dahmann, C., Oates, A. C. & Brand, M. Boundary formation and maintenance in tissue development. *Nat. Rev. Genet.* **12**, 43–55 (2011).
37. Milo, R. *et al.* Network Motifs: Simple Building Blocks of Complex Networks. *Science* **298**, 824–827 (2002).

38. Kadelka, C. *et al.* A meta-analysis of Boolean network models reveals design principles of gene regulatory networks. *Sci. Adv.* **10**, eadj0822 (2024).
39. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci.* **100**, 11980–11985 (2003).
40. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. (CRC Press, 2019).
41. Cotterell, J. & Sharpe, J. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Mol. Syst. Biol.* **6**, 425 (2010).
42. Adler, M. & Medzhitov, R. Emergence of dynamic properties in network hypermotifs. *Proc. Natl. Acad. Sci.* **119**, e2204967119 (2022).
43. Jiménez, A., Cotterell, J., Munteanu, A. & Sharpe, J. A spectrum of modularity in multi-functional gene circuits. *Mol. Syst. Biol.* **13**, 925 (2017).
44. Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810-826.e23 (2021).
45. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
46. Guantes, R. & Poyatos, J. F. Multistable Decision Switches for Flexible Control of Epigenetic Differentiation. *PLOS Comput. Biol.* **4**, e1000235 (2008).
47. Peter, I. S. Chapter Nine - The function of architecture and logic in developmental gene regulatory networks. in *Current Topics in Developmental Biology* (ed. Peter, I. S.) vol. 139 267–295 (Academic Press, 2020).
48. Friedlander, T., Mayo, A. E., Tlustý, T. & Alon, U. Evolution of Bow-Tie Architectures in Biology. *PLOS Comput. Biol.* **11**, e1004055 (2015).
49. Maimets, M. *et al.* Mesenchymal-epithelial crosstalk shapes intestinal regionalisation via Wnt and Shh signalling. *Nat. Commun.* **13**, 715 (2022).
50. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
51. Hornung, G. & Barkai, N. Noise Propagation and Signaling Sensitivity in Biological Networks: A Role for Positive Feedback. *PLOS Comput. Biol.* **4**, e8 (2008).
52. Munteanu, A., Cotterell, J., Solé, R. V. & Sharpe, J. Design principles of stripe-forming motifs: the role of positive feedback. *Sci. Rep.* **4**, 5003 (2014).
53. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
54. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).
55. Kashtan, N., Itzkovitz, S., Milo, R. & Alon, U. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20**, 1746–1758 (2004).
56. Adler, M., Mayo, A. & Alon, U. Logarithmic and Power Law Input-Output Relations in

Sensory Systems with Fold-Change Detection. *PLOS Comput. Biol.* **10**, e1003781 (2014).

57. Adler, M., Szekely, P., Mayo, A. & Alon, U. Optimal Regulatory Circuit Topologies for Fold-Change Detection. *Cell Syst.* **4**, 171-181.e8 (2017).
58. Adler, M. & Alon, U. Fold-change detection in biological systems. *Curr. Opin. Syst. Biol.* **8**, 81–89 (2018).



## Methods

### Inference of gene regulatory networks from single-cell data

We analyzed single-cell RNA sequencing data from Fawkner-Corbett et al.<sup>44</sup> using the SCENIC package in python (pySCENIC)<sup>45,53</sup>. The pySCENIC pipeline consists of: (1) Obtaining expression data per each time point in the data using the Seurat pipeline<sup>54</sup>; (2) Inferring an adjacency matrix describing co-expression relationships between transcription factors (TFs) and target genes; and (3) Constructing regulons - modules containing TF and target genes by integrating motif information. The activity of a TF in a cell is calculated by integrating the TF target genes' expression. We repeated the pySCENIC pipeline 6 times and kept the overlapping regulons to reduce noise. To focus on TFs with variable activity in the tissue, we filtered out low-variance regulons by removing those with activity variance less than 0.1. To build the gene regulatory networks, we considered pairs of TFs and target genes that have a predicted activity that is higher than the median across all pairs, considering genes that belong to a manually curated list of developmental-related genes (5143 genes).

### Network motif enrichment analysis

To identify network motifs - statistically enriched patterns in the developmental regulatory networks, we used the MFinder package<sup>55</sup> to identify network motifs of up to 3 nodes. We verified these network motifs by detecting enriched patterns using the IGraphM package in Mathematica. To randomize the networks, we used the IGRewire function and to find motifs and their frequencies we used the IGMotifs function.

### Motif role transitions

To explore transitions of genes between network motif roles, we first categorized the genes in the regulatory networks according to their role (or roles) in the motifs. We denote  $N_{m,t}$  as the set of nodes in a GRN that participate in network motifs in time  $t$ . Next, for each time point  $t$ , we categorized nodes in  $N_{m,t}$  to 7 groups according to their roles in the network motifs:  $\{n_{1,t}, n_{2,t}, \dots, n_{7,t}\}$ , for the 7 unique motif roles in our case (Fig 3A). Note that certain nodes may belong to more than one group (which is the case where network motifs are combined). Next, for each one of the 7 motif roles, we consider all successive pairs of node groups within each role, such as:  $\{n_{1,t}, n_{1,t+1}\}, \{n_{1,t+1}, n_{1,t+2}\}$ , etc. For each pair of successive groups, we compute the Jaccard index:  $J(n_{i,t}, n_{i,t+1}) = |n_{i,t} \cap n_{i,t+1}| / |n_{i,t} \cup n_{i,t+1}|$ , which is the size of the intersection of the two groups divided by the size of their union. If all genes in a certain motif role remain in the same motif role in the next time point:  $J = 1$ , and if none of the genes remain in the same motif role:  $J = 0$ .

To build the network of motif role transitions, we defined motif role states for every gene ( $g$ ) in the networks in each time point ( $t$ ):  $x_{g,t}$ , which is a 7-dimensional vector that represents whether gene  $g$  participates in each one of the 7 motif roles or not (for example:  $(1,0,0,0,0,0,0)$ ,  $(1,1,0,0,0,0,0)$ ,  $(1,0,1,0,0,0,0)$ , etc.). Overall, there are  $2^7 = 128$  different possible motif role states (although not all of them are implemented in the developmental GRNs). We then build a network where the nodes represent the different possible motif role states. We then consider for every gene, all pairs of motif role states from successive time points:  $\{x_{g,t} \rightarrow x_{g,t+1}\}$ . For each direct transition between motif role states, we draw an edge connecting the two motif role states (or a self-loop arrow in case the genes retain the same motif role state). We assign weights to the edges to account for the observed frequencies of each transition.

### **Correlations between network motif identities and gene expression profiles**

We consider pairs of specific cell types (clusters) within the epithelial cells (and separately also within the fibroblast populations) across successive time points in the data. For each such pair of cell types and for every pair of successive time points, we compute the Jaccard index of genes participating in a certain network motif or are regulated by a certain network motif. For example, considering the feedforward loop (FFL), first we create a list of all the triplets of genes participating in FFLs in each cell type. We then compute the Jaccard index of the two lists of triplet genes. We compare the Jaccard index with the Euclidean distance between the gene expression profiles of the two cell types in question. We then used the `PearsonCorrelationTest` function in Mathematica to estimate the Pearson correlation for every network motif and its statistical significance.

### **Detection of over- and under-represented network hyper-motifs**

To detect over- and under-represented network hyper-motifs, we detect enrichment of combinations of network motifs. To do so, we follow the following steps (presented in detail in <sup>42</sup>):

1. Compute the Jaccard index for all  $k(k-1)/2$  pairs of motif roles (here  $k = 7$ ) within each time point:  $J(n_{i,t}, n_{j,t}) = |n_{i,t} \cap n_{j,t}| / |n_{i,t} \cup n_{j,t}|$ .
2. For nodes that appear more than once in the same role within the same network, we compute the Jaccard index by the ratio of the number of nodes that appear in a network motif role more than once to the total number of nodes that participate in that motif role.
3. We use the MFinder package<sup>55</sup> to create 100 random networks (for each time point in the data) that have the same number of nodes and edges, the same number of

incoming and outgoing edges per node, and the same frequency of all subgraphs up to three-node subgraphs, as in the real network.

4. After categorizing the nodes according to their motif roles for each of the 100 random networks (per time point), we repeat steps 1 through 2 for all the random networks and compute  $J_{rand}(n_{i,t}, n_{j,t})$  for all pairs of motif roles in the random networks.
5. For every  $i, j$  such that  $i, j \in \{1, 2, \dots, k\}$ , we compute the Z-score of  $J(n_{i,t}, n_{j,t})$  from the distribution of  $\{J_{rand}(n_{i,t}, n_{j,t})\}$ :  $Z_{ij} = (J(n_{i,t}, n_{j,t}) - \text{mean}(\{J_{rand}(n_{i,t}, n_{j,t})\})) / \text{std}(\{J_{rand}(n_{i,t}, n_{j,t})\})$ , and compute the  $P$  value by estimating the cumulative density of  $Z_{ij}$  for a normal distribution with zero mean and unit variance. We use the function `NormalPValue` in the package `HypothesisTesting` in `Mathematica` to compute the  $P$  value. We then use the Benjamini–Hochberg procedure to correct for multiple hypothesis testing, which provides us a corrected q-value for each pair of motif roles.
6. We consider over- and under-represented motif combinations if their q-value is smaller than 0.05, where  $Z_{ij} > 0$  for over-represented combinations and  $Z_{ij} < 0$  for under-represented combinations.

## Mathematical models for statistically enriched network hyper-motif regulatory circuits

We consider the following differential equations to model the dynamical behavior of the hyper-motif circuit that combines two FFLs and a mutual feedback loop (Fig 5B):

$$\begin{aligned} \frac{dx}{dt} &= 1 - x \\ \frac{dy}{dt} &= \frac{y^{n_{yy}}}{y^{n_{yy}} + k_{yy}^{n_{yy}}} \frac{x^{n_{xy}}}{x^{n_{xy}} + k_{xy}^{n_{xy}}} \frac{w^{n_{wy}}}{w^{n_{wy}} + k_{wy}^{n_{wy}}} - y \\ \frac{dw}{dt} &= \frac{w^{n_{ww}}}{w^{n_{ww}} + k_{ww}^{n_{ww}}} \frac{x^{n_{xw}}}{x^{n_{xw}} + k_{xw}^{n_{xw}}} \frac{y^{n_{yw}}}{y^{n_{yw}} + k_{yw}^{n_{yw}}} - w \\ \frac{dz}{dt} &= \frac{x^{n_{xz}}}{x^{n_{xz}} + k_{xz}^{n_{xz}}} \frac{y^{n_{yz}}}{y^{n_{yz}} + k_{yz}^{n_{yz}}} \frac{w^{n_{wz}}}{w^{n_{wz}} + k_{wz}^{n_{wz}}} - z \end{aligned}$$

We chose parameter values that provide bistability to the mutual feedback circuit. We therefore used the following parameter values:

$$\begin{aligned} k_{ww} = k_{wy} = k_{yw} = k_{yy} = 0.3, k_{wz} = k_{xw} = k_{xy} = k_{xz} = k_{yz} = 0.01 \\ n_{ww} = n_{wy} = n_{yw} = n_{yy} = 3, n_{wz} = n_{xw} = n_{xy} = n_{xz} = n_{yz} = 1 \end{aligned}$$

We used the following initial conditions for Fig 5D:

$y_0 = 0.25, w_0 = 0.35, z_0 = 0.01$ , and  $x_0$  as denoted in Fig 5D.

and the following initial conditions for Fig 5E:

$x_0 = 0, w_0 = 0.3, z_0 = 0.01$ , and  $y_0$  as denoted in Fig 5E.

We consider the following differential equations to model the dynamical behavior of the hyper-motif circuit that combines two FFLs forming a multi-input FFL (Fig 5F):

$$\frac{dy}{dt} = \frac{x_1}{x_2} - y$$
$$\frac{dz}{dt} = \frac{x_1 x_2}{y} - z$$

Here we used log linear terms (assuming that the variables are far from saturation) since this is the range of parameters that provide well-known dynamical properties of the coherent and incoherent FFLs such as the fold-change detection property<sup>39,56–58</sup>.

### **Inference of cell communication circuits through morphogen ligand-receptor interactions**

To infer interactions between epithelial and fibroblast populations, we applied NicheNet, an algorithm that infers ligand-receptor interactions within a transcriptomic dataset, using prior knowledge on ligand-receptor interactions and regulatory networks<sup>50</sup>. We applied NicheNet to the intestinal development scRNAseq dataset considering interactions between the specific cell type populations of fibroblasts and epithelial cells across the developmental process. The pairwise interaction between cell types (mature enterocytes, fibroblast progenitors, absorptive cells, etc.) was calculated independently for each time point in the data (9 time points spanning 8-22 PCWs). For each cell type and for each time point, genes of interest (GOIs) were defined as genes that were upregulated in a particular cell type compared to all other cell types in that time point and expressed in at least 25% of the cells within the cell type population. Interaction score was calculated as the Pearson coefficient calculated by 'predict\_ligand\_activities' function, using the GOIs of that time point and the genes that were expressed in the 'receiver' clusters as targets. For each pair of sender-receiver cells, we chose the top 10% quantile ligands. This analysis allowed us to infer interactions between every pair of specific cell type populations among the different fibroblast and epithelial clusters. In Fig 6A we consider the possible interactions through the different morphogen ligand-receptor pairs where we group together the various fibroblast and epithelial clusters.

## Analysis of shortest network-motif paths

To infer regulatory interactions that link a signal (S) perceived by a cell to an expression of a ligand encoded by a target gene (Z), we consider the GRN of that cell type. We then use the function 'FindShortestPath' in Mathematica to compute the shortest path linking a transcription factor X, which is downstream to the S signaling pathway and the target gene Z. Let us denote the set of genes that are in this shortest path  $\{G_S\}$ . We now build a new set of genes that will make up the shortest network-motif path:  $\{G_M\}$ . For each gene  $g_i$  in  $\{G_S\}$ , we perform the following steps:

1. Check if  $g_i$  participates in network motifs in the same time point and in the same cell type.
2. If the answer to #1 is yes - add  $g_i$  and the genes interacting with  $g_i$  in the network motif to  $\{G_M\}$ . \*Note that in case  $g_i$  participates in multiple network motifs of the same topology and in the same motif role, we consider only one instance of this motif in  $\{G_M\}$ . This is similar to having multiple shortest paths in a network, and arbitrarily choosing one of them.
3. If the answer to #1 is no - add only  $g_i$  to  $\{G_M\}$ .

## Mathematical models for regulatory circuits between communicating cells

We model the dynamical behavior of two morphogen ligands that regulate each other's expression through regulatory circuits between a pair of communicating cells (Fig 7). For the cell circuit depicted in Fig 7B we use the following set of differential equations where x represents the FOS/SMAD1/NR2F2/FOXP2 module, y represents HNF4A, Z is PPARG, u is for IHH, and v represents BMP2/4/5/7:

$$\begin{aligned}\frac{dx}{dt} &= \frac{v^{n_{vx}}}{v^{n_{vx}} + k_{vx}^{n_{vx}}} \frac{x^{n_{xx}}}{x^{n_{xx}} + k_{xx}^{n_{xx}}} - x \\ \frac{dy}{dt} &= \frac{k^{n_{xy}}}{x^{n_{xy}} + k_{xy}^{n_{xy}}} \frac{y^{n_{yy}}}{y^{n_{yy}} + k_{yy}^{n_{yy}}} - y \\ \frac{dz}{dt} &= \frac{y^{n_{yz}}}{y^{n_{yz}} + k_{yz}^{n_{yz}}} - z \\ \frac{du}{dt} &= \frac{y^{n_{yu}}}{y^{n_{yu}} + k_{yu}^{n_{yu}}} \frac{z^{n_{zu}}}{z^{n_{zu}} + k_{zu}^{n_{zu}}} - u \\ \frac{dv}{dt} &= \frac{u^{n_{uv}}}{u^{n_{uv}} + k_{uv}^{n_{uv}}} - v\end{aligned}$$

We used the following parameter values for Fig 7C:

$$k_{uv} = k_{vx} = k_{xx} = k_{xy} = k_{yu} = k_{yy} = k_{yz} = k_{zu} = 0.01$$



$$n_{uv} = n_{vx} = n_{xx} = n_{xy} = n_{yu} = n_{yz} = n_{zu} = 1, n_{yy} = 3$$

with the following initial conditions:

$$x_0 = z_0 = u_0 = v_0 = 0.01, y_0 = 1$$

And the following parameter values for Fig 7D:

$$k_{uv} = k_{xx} = k_{xy} = k_{yu} = k_{yz} = k_{zu} = 0.01, k_{vx} = 4.3, k_{yy} = 0.3$$

$$n_{uv} = n_{vx} = n_{xx} = n_{xy} = n_{yu} = n_{yz} = n_{zu} = n_{yy} = 1$$

with the following initial conditions:

$$x_0 = y_0 = z_0 = u_0 = v_0 = 0.01$$

For the cell circuit depicted in Fig 7E we use the following set of differential equations where x is IRF1, y is TCF12/4, Z is SOX6, u is for BMP4, and v represents WNT4. To account for effect of expression of BMP and WNT antagonists, we added direct inhibition terms in the equations for the morphogen ligands:

$$\begin{aligned} \frac{dx}{dt} &= 1 - \alpha x \\ \frac{dy}{dt} &= \frac{v^{n_{vy}}}{v^{n_{vy}} + k_{vy}^{n_{vy}}} \frac{x^{n_{xy}}}{x^{n_{xy}} + k_{xy}^{n_{xy}}} \frac{k^{n_{yy}}}{y^{n_{yy}} + k_{yy}^{n_{yy}}} - y \\ \frac{dz}{dt} &= \frac{k^{n_{xz}}}{x^{n_{xz}} + k_{xz}^{n_{xz}}} \frac{k^{n_{yz}}}{y^{n_{yz}} + k_{yz}^{n_{yz}}} \frac{k^{n_{zz}}}{z^{n_{zz}} + k_{zz}^{n_{zz}}} - z \\ \frac{du}{dt} &= \frac{u^{n_{uu}}}{u^{n_{uu}} + k_{uu}^{n_{uu}}} \frac{k^{n_{vu}}}{v^{n_{vu}} + k_{vu}^{n_{vu}}} \frac{x^{n_{xu}}}{x^{n_{xu}} + k_{xu}^{n_{xu}}} \frac{k^{n_{zu}}}{z^{n_{zu}} + k_{zu}^{n_{zu}}} - u \\ \frac{dv}{dt} &= \frac{k^{n_{uv}}}{u^{n_{uv}} + k_{uv}^{n_{uv}}} \frac{k^{n_{vv}}}{v^{n_{vv}} + k_{vv}^{n_{vv}}} - v \end{aligned}$$

We used the following parameter values for Fig 7F:

$$k_{uu} = k_{vu} = 10.2, k_{uv} = k_{vy} = k_{xy} = k_{xz} = k_{yy} = k_{yz} = k_{zz} = 0.01, k_{vv} = 4.7,$$

$$k_{xu} = 0.3, k_{zu} = 6.3, \alpha = 5.7$$

$$n_{uu} = n_{uv} = n_{vu} = n_{vy} = n_{xy} = n_{xz} = n_{yy} = n_{yz} = n_{zu} = n_{zz} = 1, n_{vv} = n_{xu} = 3$$

with the following initial conditions:

$$x_0 = y_0 = z_0 = u_0 = v_0 = 0.01$$

And the following parameter values for Fig 7G:

$$k_{uu} = k_{uv} = k_{vy} = k_{xy} = k_{xz} = k_{yy} = k_{yz} = k_{zz} = 0.01,$$

$$k_{xu} = k_{zu} = k_{vv} = 0.3, \alpha = 0.01$$

$$n_{vv} = n_{vy} = n_{xy} = n_{xu} = n_{xz} = n_{yy} = n_{yz} = n_{zu} = n_{zz} = 1, n_{uu} = n_{vu} = n_{uv} = 3$$

with the following initial conditions:

$$x_0 = y_0 = z_0 = v_0 = 0.01, u_0 = 0.46$$

For the cell circuit depicted in Fig 7H we use the following set of differential equations where x is KLF5, y is TCF7L2, w is HNF4A, Z is FOXA1, u is for IHH, and v represents WNT4:

$$\begin{aligned} \frac{dx}{dt} &= 1 - \alpha x \\ \frac{dy}{dt} &= \frac{v^{n_{vy}}}{v^{n_{vy}} + k_{vy}^{n_{vy}}} \frac{x^{n_{xy}}}{x^{n_{xy}} + k_{xy}^{n_{xy}}} \frac{y^{n_{yy}}}{y^{n_{yy}} + k_{yy}^{n_{yy}}} \frac{w^{n_{wy}}}{w^{n_{wy}} + k_{wy}^{n_{wy}}} - y \\ \frac{dw}{dt} &= \frac{x^{n_{xw}}}{x^{n_{xw}} + k_{xw}^{n_{xw}}} \frac{w^{n_{ww}}}{w^{n_{ww}} + k_{ww}^{n_{ww}}} \frac{y^{n_{yw}}}{y^{n_{yw}} + k_{yw}^{n_{yw}}} - w \\ \frac{dz}{dt} &= \frac{w^{n_{wz}}}{w^{n_{wz}} + k_{wz}^{n_{wz}}} \frac{k^{n_{zz}}}{z^{n_{zz}} + k_{zz}^{n_{zz}}} - z \\ \frac{du}{dt} &= \frac{z^{n_{zu}}}{z^{n_{zu}} + k_{zu}^{n_{zu}}} \frac{w^{n_{wu}}}{w^{n_{wu}} + k_{wu}^{n_{wu}}} - u \\ \frac{dv}{dt} &= \frac{u^{n_{uv}}}{u^{n_{uv}} + k_{uv}^{n_{uv}}} - v \end{aligned}$$

We used the following parameter values for Fig 7I:

$$k_{ij} = 0.3, n_{ij} = 3 \text{ for every } i, j$$

with the following initial conditions:

$$x_0 = u_0 = v_0 = 0.01, y_0 = 1.1, z_0 = 0.33, w_0 = 0.83$$

And the following parameter values for Fig 7J:

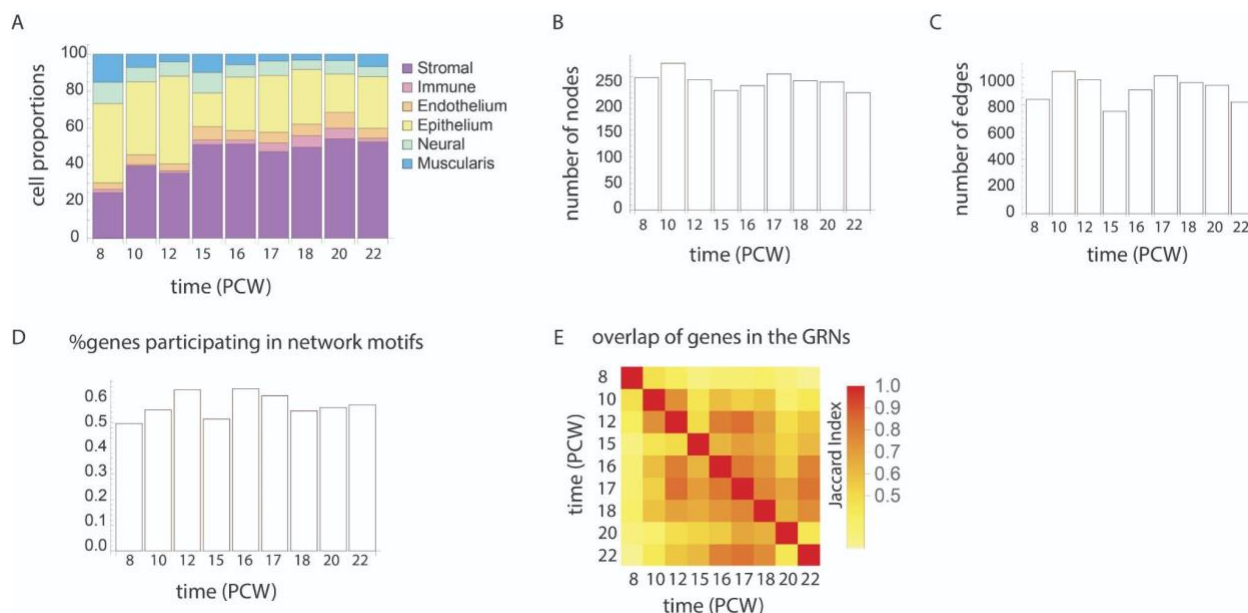
$$n_{ij} = 3 \text{ for every } i, j \text{ except for } n_{vy} = 1$$

$$k_{uv} = k_{wu} = k_{ww} = k_{wy} = k_{wz} = k_{xu} = k_{xw} = k_{yw} = k_{zu} = k_{zz} = 0.01, k_{vy} = k_{xy} = k_{yy} = 0.3$$

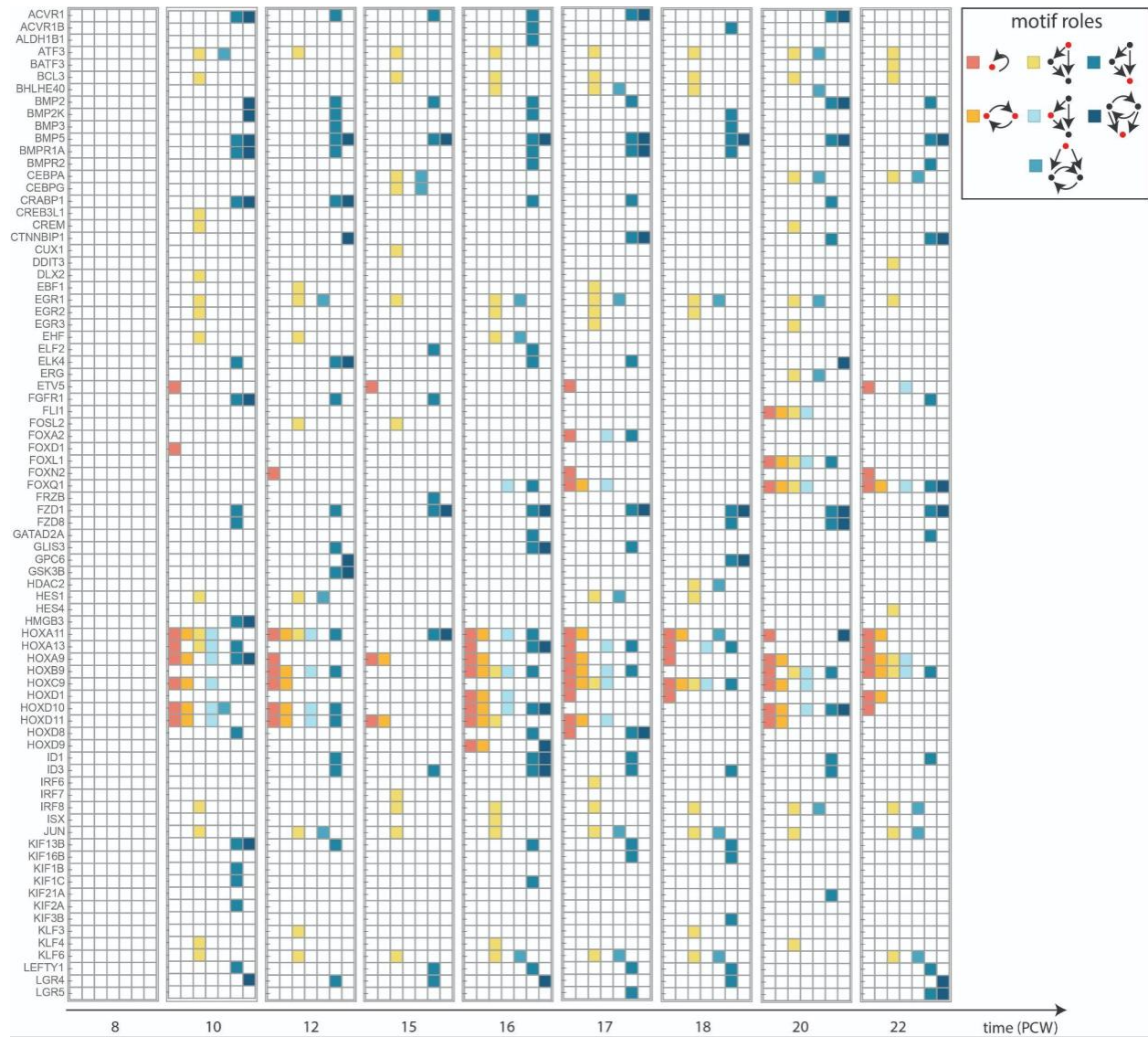
with the following initial conditions:

$$x_0 = y_0 = w_0 = z_0 = v_0 = u_0 = 0.01$$

## SI Figures



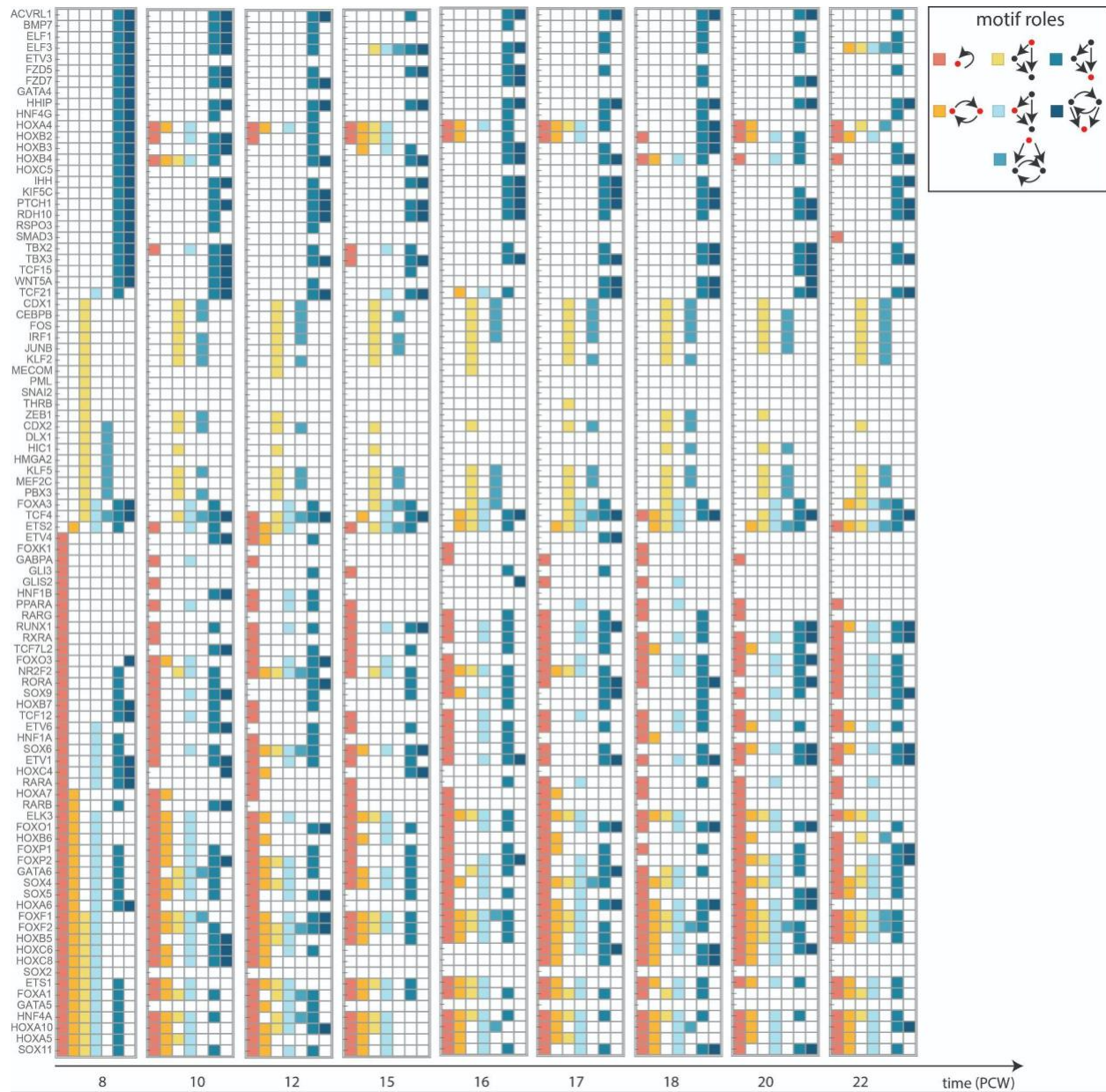
**Figure S1:** (A) Proportions of the 6 major cell type categories in the intestine computed from the single-cell data in Fawkner-Corbett et al.<sup>44</sup> across the 9 time points that follow the developmental process. (B-C) Number of nodes (B) and edges (C) in each network inferred from the single-cell data. (D) Proportion of genes that participate in network motifs out of the total number of genes in each network. (E) Heatmap of the overlap of genes that participate in the different gene regulatory networks (GRNs). Every square is a comparison between two networks from time points in the data. Overlap is computed by calculating the Jaccard index of the two lists of genes participating in the two networks.











**Figure S2:** Table of genes participating in network motifs in at least one of the networks inferred from the single-cell data, and their roles within the network motifs in each time point in the data. Every row is a gene, and it is colored by its role within network motifs in each time point.