Macroscale network feedback structure of transcriptome during cell fate transition

Lingyun Xiong[1, 2, 3, 7], William A. Schoenberg[4, 5], Jeremy D. Swartz[6]

1. EPSRC Systems Biology Programme, Doctoral Training Centre, University of Oxford, UK;
2. Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, UK;
3. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, UK;
4. Department of Geography, University of Bergen, Norway;
5. isee systems inc. Lebanon NH, USA;
6. Media Studies Program, University of Oregon, USA;
7. Correspondence: lingyun.xiong@dtc.ox.ac.uk (L. X.)

Abstract

Organogenesis is a biological process of cell fate transition. Underlying this process are complex changes in the transcriptome. Although these changes have been described extensively in terms of individual genes or pathways, a mcroscale description of gene regulatory network dynamics is lacking. Leveraging time-series data from single-cell transcriptome profiling of developing mouse embryos, we present a novel analytic framework to aggregate dynamic gene expressions across the transcriptome into state changes in 14 core biological processes, and to construct an ordinary differential equation-based mathematical model of the macroscale network capturing its feedback structure. We determine the polarity and magnitude of direct pairwise causal relationships between biological processes and identify higher-order feedback interactions that dominate system behaviour. Despite heterogenous expressions at the gene level, we find that the transcriptome at the macroscale level has a feedback structure that is intrinsically stable and robust. We show the pivotal role of *signaling* in driving systemic changes and uncover the importance of *homeostatic process* and *establishment of localization* in regulating network dynamics. Localized regulatory structures represent domain-specific regulations that are essential to cell fate transition, especially *lipid metabolic process*. Altogether, this study not only provides a holistic picture of macroscale transcriptome dynamics during mouse organogenesis, but also offers insight into key aspects of information flow in the transcriptome that control cell fate transition.

Table of Contents

Macroscale network feedback structure of transcriptome during cell fate transition
- **Abstract**
- **Introduction**
- **Results**
  - A coarse-grain representation of the gene regulatory network
  - A causal model of the macroscale network
  - The pivotal role of signalling in driving systemic changes
  - Secondary roles of homeostatic process and establishment of localization
  - Instances of localized impact structure
- **Discussions**
- **Methods**
  - Data processing
  - Selection of highly variable genes
  - Macroscale modeling of gene regulatory network
  - Data aggregation
  - Network learning scheme
  - Inference on pairwise causal relationship
  - Feedback loop dominance analysis
  - Data visualization
- **References**

Introduction

Transition in cell fate is the central theme of organogenesis: embryonic stem cells diversify into intermediate progenitors that will differentiate into diverse cell lineages of specialized form and function in all organs (Gilbert, 2014). This is a complex yet coordinated process, effected by the spatial and temporal expression of genes (Arnold and Robertson, 2009; Young, 2011; Sladitschek et al., 2020). Cell identity is thus reflected by the collective state of the transcriptome (MacArthur 2009; Cheng et al., 2019; Pijuan-Sala et al., 2019; Cao et al., 2019). Recent years has seen tremendous effort in probing transcriptome-wide gene expression during mouse organogenesis at single-cell resolution. This offers us a unique opportunity to look beyond individual genes and pathways (Tam and Loebel, 2007), to model the dynamics of gene regulatory network as a whole (Davidson, 2002; Lu et al., 2009). In particular, Pijuan-Sala et al. sequntially sampled developing mouse embryos every 6 hours between 6.5 to 8.5 days post-fertilization (inclusive), and performed single-cell RNA sequencing to construct the cell atlas during early mouse organogenesis. Utitlizing this comprehensive dataset, we aim to establish a formal description of the transcriptome dynamics that underlies cell fate transition.

With more than 29,000 genes annotated in the mouse genome, the gene regulatory network is high-dimensional. So far, biological network analyses have mainly focused on subsets of the network, such as transcriptional regulatory networks (Luscombe et al., 2004; Neph et al., 2012), protein-protein interaction networks (Fraser et al., 2002; Han et al.,

2004), and metabolic networks (Jeong et al., 2000; Fondi and Lio, 2015). To reduce the dimensionality of the network while capturing its entirety, we leverage Gene Ontology (GO) annotations. GO is a literature-curated reference database that annotates genes' functions in the hierarchy of cellular subsystems (The Gene Ontology Consortium, 2017; Ma et al., 2018; Bult et al., 2019). Taking this knowledge about genes and cellular subsystems as an anchor point, we are able to model the gene regulatory network at the macroscale level, where the network components represent 14 core biological processes as precedented by the Mouse Genome Institute.

Formalism of modelling network dynamics in biological systems is severely constrained by the presence of nonlinear and feedback-rich interactions amongst network components (Zou et al., 2019). Existing paradigms draw inference either on pre-specified (sub-)network structures (Harush and Barzel, 2017) or using probabilistic techniques without identifying the underlying causal structure (Casadiego et al., 2017; Groß et al., 2019). Here, we take a data-driven approach with a methematical backbone to model the system dynamics in terms of network structure (Tu et al., 2005; Aibar et al., 2017; Peixoto and Rosvall, 2017; Fujii et al., 2020). We agnostically train a causal network model that competently reproduces network dynamics. We then instrument the generated model to infer the nature of interactions from the causal structure of the mathematical model.

In this study, we describe the macroscale dynamics of the transcriptome during mouse organogenesis using a causal model based on ordinary differential equations. We develope an analytic framework for modelling the high-dimensional gene regulatory network at the macroscale, which algorithmically generates a network model to explain system behaviour. We find that the feedback structure of the macroscale network is intrinsically stable and robust. Amongst the 14 core biological processes, *signaling* plays a central part in regulating system dynamics during mouse organogenesis, and *homeostatic process* and *establishment of localization* demonstrate a secondary role. Tertiary influencers include *system development* and *cell differentiation*. We show that *lipid metabolic process* exerts significant impact on the network dynamics by negative feedbacks, which is potentially entangled in a bistable relationship with *signaling* in facilitating cell fate transitation.

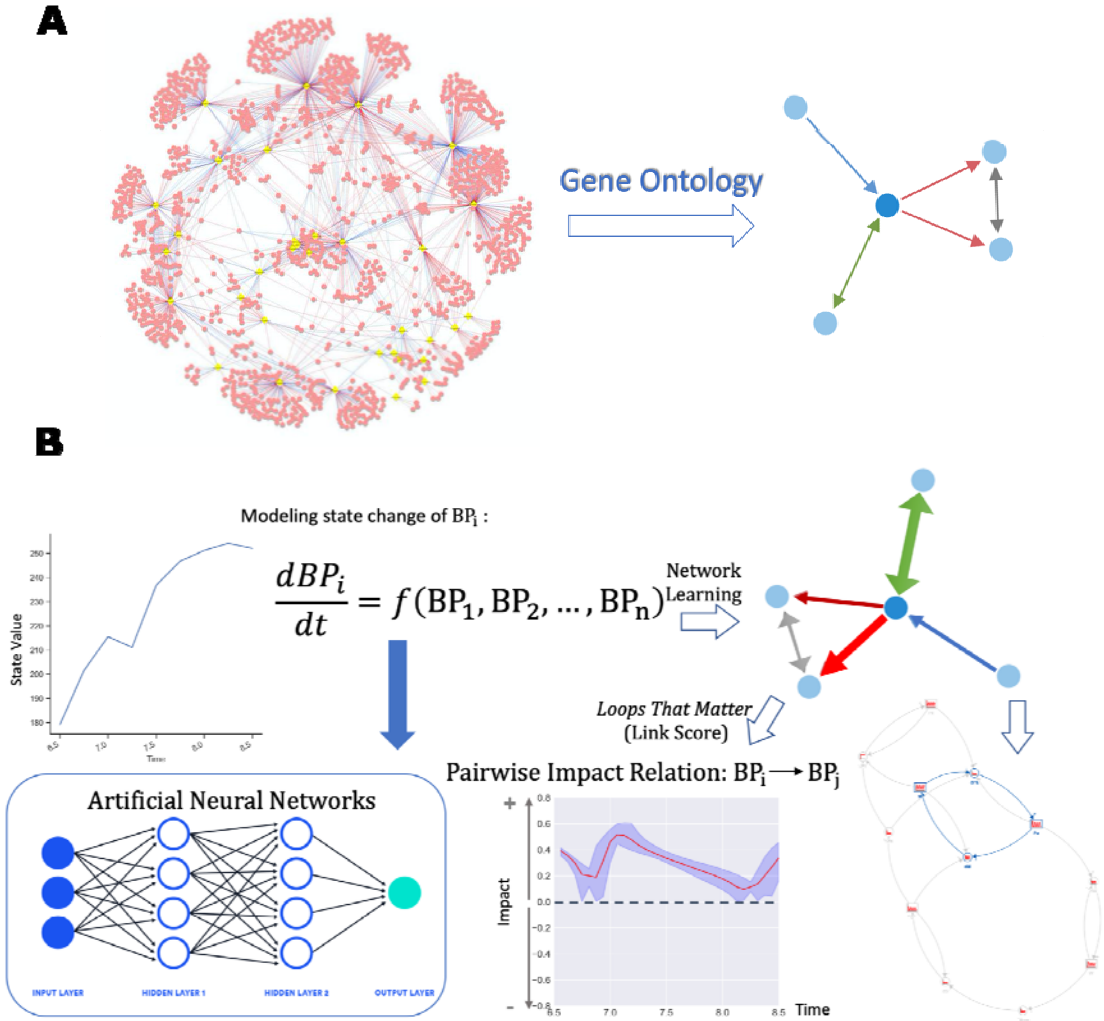

Figure 1. Graphical summary of methodologies. (A) Leveraging Gene Ontology annotations, we condensed the entire mouse gene regulatory network into a 14-node macroscale representation. Each node in the macroscale representation stands for a top-level biological process sustaining cellular function. (B) To agnostically construct a causal model of the macroscale network that recapatulates the observed system dynamics, we used the Feedback System Neural Network (FSNN) method (Schoenberg, 2020) based on the aggregate gene expression levels in each biological process. In brief, state change of each biological process is modelled by a differential equation, whose form is captured by an artificial neural network. The network model was trained and tested. The mathematics of the generated model were then analzyed to infer direct interactions amongst the 14 biological processes, in terms of both pairwise causal relationships and feedback loops that dominate the behaviour of the nonlinear dynamical system.

Results

A coarse-grain representation of the gene regulatory network

In Gene ontology (GO), genes are annotated for its functions in diverse molecular/biochemical processes, such as DNA repair and cell growth (The Gene Ontology Consortium, 2017). Detailed or coarse, the annotations are related in a largely hierarchical directed acyclic graph (DAG). Child-parent relationships in the DAG link lower-level subsystem-oriented functions to those pertaining to upper-level broad categories. We traced the paths in the DAG and collapsed the lower-level terms to the top-level ones, condensing the gene regulatory network into a 14-node representation (i.e., macroscale network; **Figure 1A and 3**; **Methods**). Each node encapsulates genes involved in a distinct functional domain of the biological system (**Table 1**; **Supplementary Table 1 and 2**).

Single-cell RNA sequencing data of devlopment mouse embryos generated by Pijuan-Sala et al. was retrieved and processed according to authors' instructions. This provides a matrix of normalized expression values of genes in respective cells. Cells are marked for specific stage they are sampled (E6.5-8.5, sampled every 6 hours, 9 time points in total). We selected cells annotated as endoderm lineage and we modeled variance of gene expression levels amongst these cells. We identified 982 highly variable genes (HVGs; **Figure 2A**; **Methods**), many of which are master regulators of embryo development (e.g., CDX2, OCT3/4, WNT5/6 and BMP2/4; **Supplementary Table 2**) and biomarkers for gut organogenesis (e.g., Krt8, Krt18, Epcam, and Apela; **Supplementary Table 2**). There are two converging trajectories of differentiation during gut organogenesis (Pijuan-Sala et al., 2019), one from anterior primitive streak to definitive endoderm to gut (T1), and the other from visceral endoderm to gut (T2). We thus split the cells into two, corresponding to the distinct but related cell populations. Next, we conducted a two-stage data aggregation scheme to group gene expression levels to represent the state value (i.e. gross activity) of each biological process. (**Methods**). As a result, we have generated 14 time-series in every set (two trajectories, 4 sets each; **Figure 2C**; **Methods**). These sequential aggregate gene expression levels constitute the empirical basis of the downstream causal modelling.

Table 1: Summary of the 14 core biological processes based on Gene Ontology annotations.

GO ID	Biological Process	Description (Bult et al., 2019)	No. of HVGs§
GO:0016043	**Cellular component organization**	A process that results in the assembly, arrangement of constituent parts, or disassembly of a cellular component.	351
GO:0097659	**Nucleic acid templated transcription**	The cellular synthesis of RNA on a template of nucleic acid (DNA or RNA).	211
GO:0048731	**System development**	The process whose specific outcome is the progression of an organismal system over time, from its formation to the mature structure.	350
GO:0030154	**Cell differentiation**	The process in which relatively unspecialized cells acquire specialized structural and/or functional features that characterize the cells, tissues, or organs of the mature organism.	308
GO:0050896	**Response to stimulus**	Any process that results in a change in state or activity of a cell or an organism as a result of a stimulus.	473
GO:0023052	**Signalling**	The entirety of a process in which information is transmitted within a biological system.	314
GO:0008283	**Cell population proliferation**	The multiplication or reproduction of cells, resulting in the expansion of a cell population.	180
GO:0019538	**Protein metabolic process**	The chemical reactions and pathways involving a protein. Includes protein modification.	232
GO:0051234	**Establishment of localization**	Any process that localizes a substance or cellular component. This may occur via movement, tethering or selective degradation.	264
GO:0006629	**Lipid metabolic process**	The chemical reactions and pathways involving lipids, compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent.	73
GO:0002376	**Immune system process**	Any process involved in the development or functioning of the immune system, an organismal system for calibrated	165

		responses to potential internal or invasive threats.	
GO:0008219	**Cell death**	Any biological process that results in permanent cessation of all vital functions of a cell.	165
GO:1901135	**Carbohydrate derivative metabolic process**	The chemical reactions and pathways involving carbohydrate derivative.	56
GO:0042592	**Homeostatic process**	Any biological process involved in the maintenance of an internal steady state.	134

§ A list of HVGs tagged for each biological process can be found in **Supplementary Table 2**.

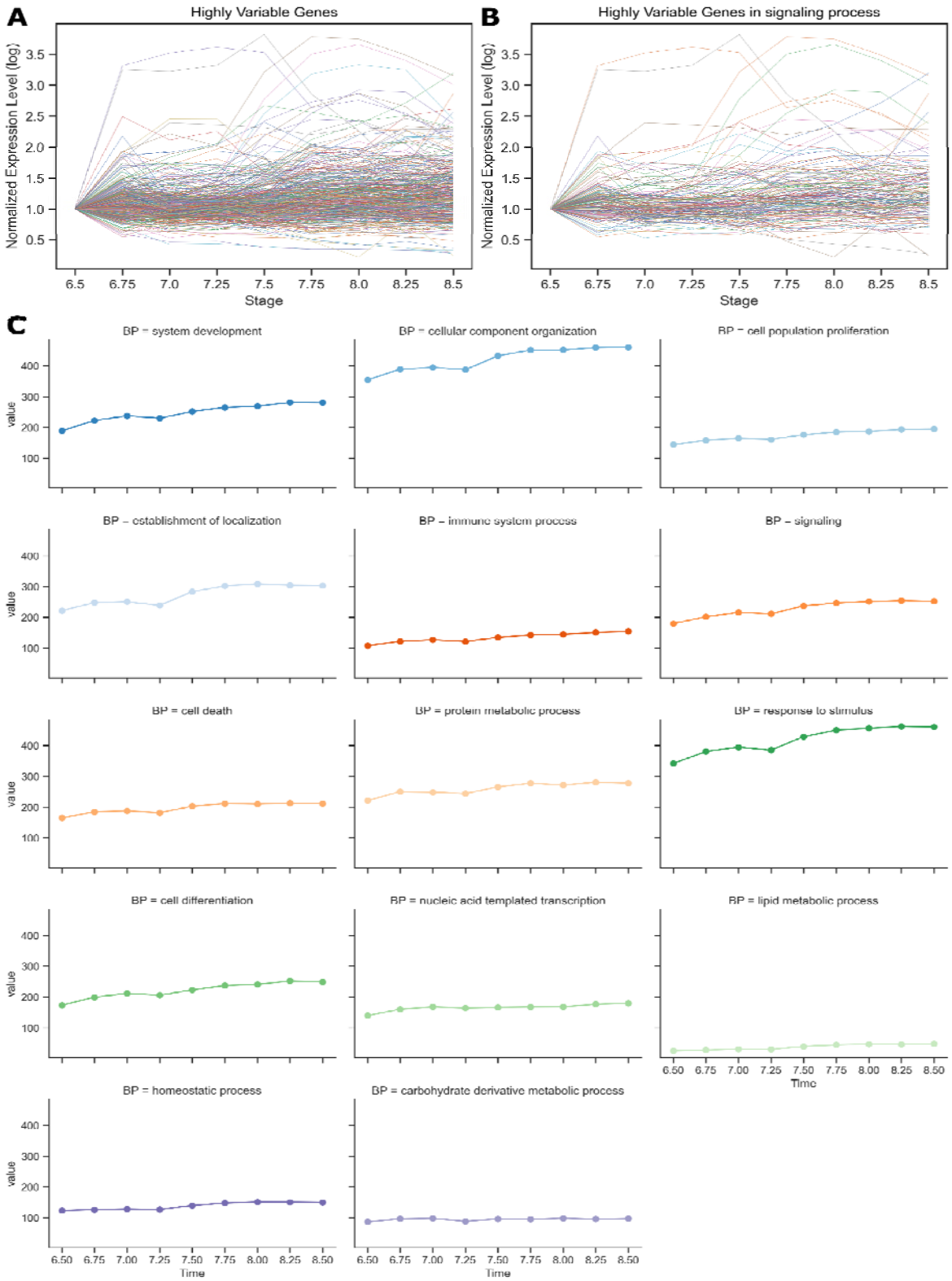

Figure 2. Aggregation of gene expression values to represent the gross activity of the 14 core biological processes (only T1 is shown). (A) Average gene expression levels of

highly variable genes during mouse gut organogenesis (E6.5-8.5, sampled every 6 hours). Expression values were normalized to the initial time point (E6.5) for visualization. (B) Average gene expression levels of highly variable genes tagged for *signaling*. (C) Aggregate gene expression levels (i.e. state values) of the 14 core biological processes.

A causal model of the macroscale network

Based on the state values of the 14 biological processes at the 9 sequential time points, we constructed a causal model of the macroscale network using Feedback System Neural Networks (FSNN; Schoenberg, 2020; **Methods**) (**Figure 1B**). For the two differentiation trajectories mentioned above, we generated two unique models: T1 and T2 (**Figure 3**). The trained models reproduced the test datasets, with an average mean squared error (R2) of 93% for T1 (range: [44.9%, 99.3%]) and 79% for T2 (range: [43.3%, 97.3%]), respectively (**Figure 3A and 3C**; **Supplementary Table 3**).

We analyzed the structure of the ordinary differential equations, to determine direct pairwise causal relationships between network components as well as dominant higher-order interactions (**Methods**). The link score metric was used to quantify the relative contribution of the state change in source node to the state change in target node (**Methods**). The link score measures both the polarity and magnitude of the causal relationship of interest. If the link score is positive (i.e. the signs of the first derivative of the source and target nodes are consistent), the feedback relationship is reinforcing; if negative (i.e. the signs of the first derivative of the source and target nodes are opposite), the feedback relationship is inhibitory. Ranging from 0 to 1, the magnitude of the link score reflects the strength of the causal relationship, which is calculated as the fraction of state change in target node accounted for by the state change in source node. Key higher-order feedback interactions were assessed and ranked via feedback loop dominance analysis (Schoenberg et. al, 2019; **Methods**). Static visualization of the network feedback structure demonstrates the cumulative impact of each node on its target nodes over the entire simulation period (**Figure 3B and 3D**). A dynamic view is offered by stacking frames of the instantaneous impact relations over time (Movies in **Supplementary Materials**).

Biological systems are known to harbour scale-free networks. Indeed, the degree of nodes in our causal models approximately follows a power-law distribution, with an average degree of 3.5 and an average path length of 1.8 (**Supplementary Materials**). Despite the heterogeneity in gene expressions during embryo development (**Figure 2A and 2B**), we observe that the macroscale network is intrinsically stable and robust, as seen in the concordant network structure for both trajectories (**Figure 3B and 3D**; Movies in **Supplementary Materials**). In the following sections, we present three commonalities in the network feedback structure: the primary role of *signalling* in regulating network dynamics during cell fate transition, the secondary roles of *homeostatic process* and *establishment of localization*, and localized impact structure that are essential to cell fate transition.

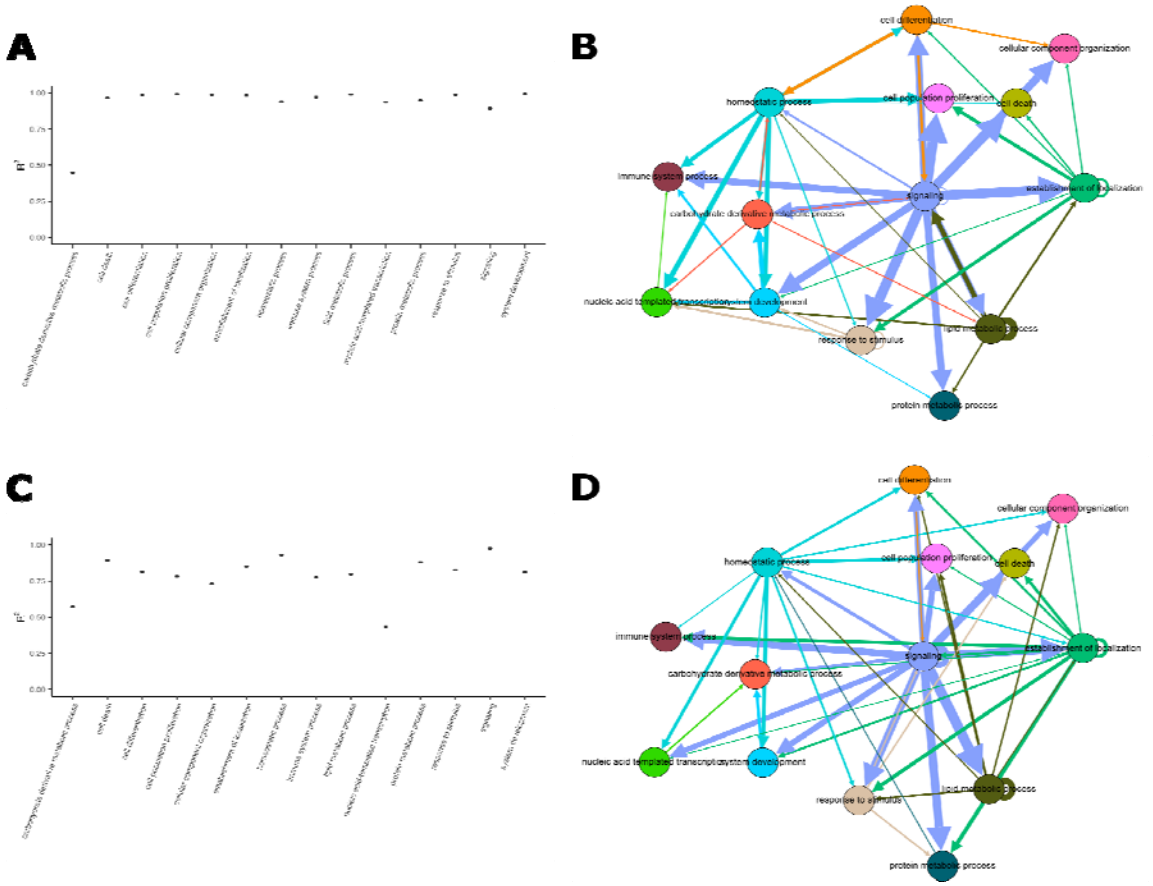

Figure 3. Two causal models of macroscale gene regulatory network during mouse gut organogenesis. (A, B) Trajectory 1 (T1): model is trained and tested on data drawn from cells annotated as anterior primitive streak, definitive endoderm and gut. (C, D) Trajectory 2 (T2): model is trained and tested on data drawn from cells annotated as visceral endoderm and gut. (A, C) Outcome of model parameterisation, shown as mean square error (R2) measured between model prediction and test data for each biological process. (B, D) Static view of the network feedback structures. Nodes represent the 14 core biological processes. Directed edges indicate the presence of causal relationship from the state change in source node to the state change in target node. Edge width is proportional to the cumulative impact over the entire simulation period. Note that the arrow does not indicate the polarity of feedback (positive/negative). Only the strongest 49 edges (the upper quartile of all 196 possible directed connections) are included for visualization.

The pivotal role of signalling in driving systemic changes

Central in the network models is *signalling*, whose state change induces state changes across the entire network (Purple edges; **Figure 3B and 3D**). Embodying "the entirety of a process of information transfer within a biological system" (**Table 1**; **Supplementary Table 1**) (The Gene Ontology Consortium, 2017), *signalling* plays a pivotal role of regulating network dynamics during mouse organogenesis, during which signal transduction has been regarded quintessential (Basson, 2012). Constituent genes in *signalling* mainly encode transcription factors (e.g., OCT3/4, GATA4, Sox family proteins, and T-box family proteins) and key factors in signal transduction pathways (e.g., WNT pathway, TGF-β pathway, FGF pathway, and heat shock response) (**Supplementary Table 2**).

State change in *signalling* induces and reinforces the state changes in all 14 biological processes, including itself (**Figure 4A**; **Supplementary Materials**). Its impact accounts for up to 70% of the state changes in its target nodes, including *system development* and *cell differentiation*, the canonical players in mouse embryo development (**Table 1**; **Supplementary Table 1**) (The Gene Ontology Consortium, 2017). Furthermore, amongst all the higher-order interactions inferred from the causal models, positive feedback of *signalling* on itself (loop R1) is one of the most prominent feedback loops that explain system bahaviour. It accounts for 2.24% and 2.12% of the total of system behaviour in T1 and T2, respectively (**Figure 3 and 5**; **Supplementary Tables 4 and 5**). This positive feedback loop is necessary for the system to be bistable or multistable during cell fate transition (Ferrell, 2012). This feature of network topology reflects fundamental design principles of biological computation in the system.

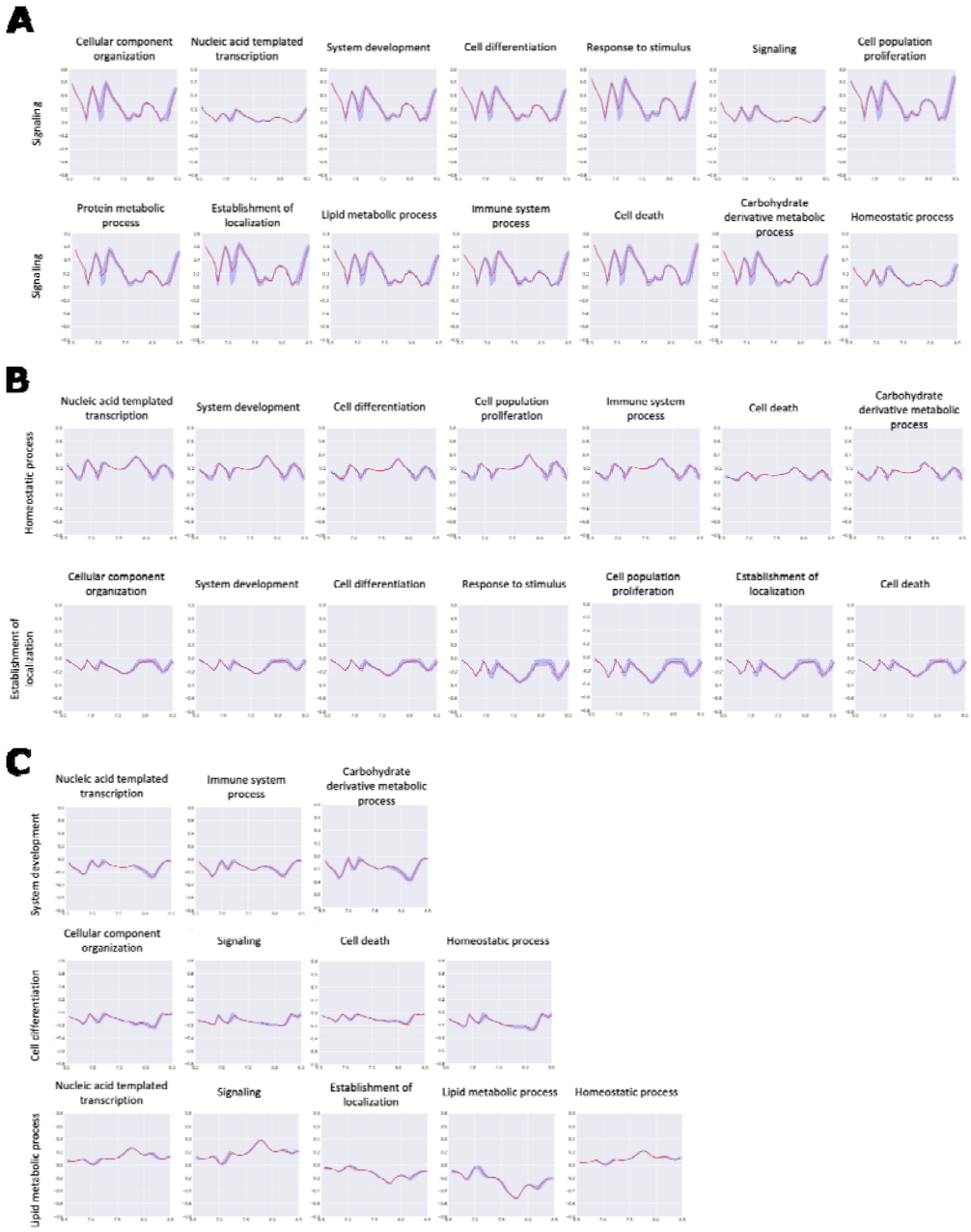

Figure 4. Inferred pairwise causal relationships measured by link score (T1). (A) Impact of state change in *signaling* on the state changes in all 14 nodes of the macroscale network during the entire simulation period is measured using link score (*signaling* target node of interest). (B) Impact of state changes in *homeostatic process* and *establishment of localization* on the state changes in half of the macroscale network. (C) Example of localized pairwise causal relationships directed from *system development*, *cell differentiation* and *lipid metabolic process*. See **Supplementary Materials** for T2.

Secondary roles of homeostatic process and establishment of localization

Homeostatic process and *establishment of localization* regulate dynamic changes in half the network (blue and green edges; **Figure 3B and 3D**). *Homeostatic process* refers to "any biological process involved in the maintenance of an internal steady state" (**Table 1**; **Supplementary Table 1**) (The Gene Ontology Consortium, 2017). Upon internal or external stimulus, products of constituent genes in *homeostatic process* issue instructions to specific cellular subsystems to restore system behaviour, for example, to reorganize cytoskeleton and to switch between anabolic and catabolic processes in metabolism. Akin to *signalling*, it is in essence an ontological regulator process, which is critical for system regulation and control (Ferrell, 2016). Notably, state change in *homeostatic process* reinforces state changes in *system development* and *cell differentiation*, accounting for up to 40% of state changes (**Figure 4B**; **Supplementary Materials**).

On the other hand, *establishment of localization* is an effector process, which executes mobility commands "via movement, tethering and selective degradation of molecules or cellular components" (**Table 1**; **Supplementary Table 1**) (The Gene Ontology Consortium, 2017). Its state change negatively regulates state changes in *system development* and *cell differentiation*, as well as itself (**Figure 4B**; **Supplementary Materials**). The self-inhibitory feedback of *establishment of localization* (loop B2), like the self-reinforcing feedback of *signaling* (loop R1), plays an important role in the overall feedback loop structure: explaining 1.87% and 2.77% of the total system behaviour in T1 and T2, respectively (**Figure 3 and 5**; **Supplementary Tables 4 and 5**). This indicates the important role of effector processes in regulating network dynamics.

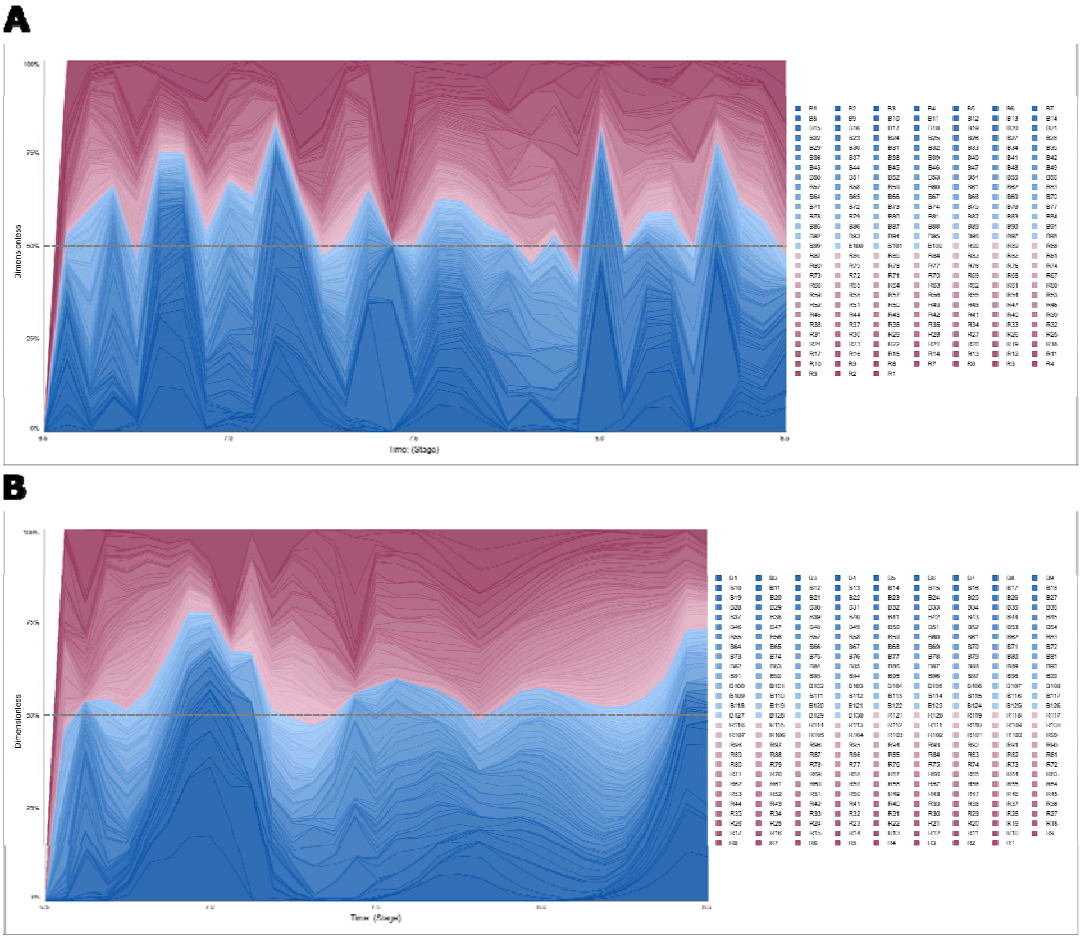

Figure 5. Inferred higher-order interactions in the causal models. Feedback loop dominance analysis of the two causal models (A: T1; B: T2) measures and ranks the impact of indiviual feedback loops on system dynamic behaviour, shown here as the proportion of variance explained by each feedback loops during the simulation period (horizontal axis). Influence measures of individual feedback loops are shaded by two colour gradients (red: positive/reinforcing feedback [even-numbered negative link scores in the loop]; blue: negative/inhibitory feedback [odd-numbered negative link scores in the loop]). The higher-ranked feedback loops are shaded in darker hues and arranged towards the extremes of vertical axis. Examples of the feedback loops presented: B1 -- self-inhibitory feedback of *lipid-metabolic process*; R1-- self-reinforcing feedback of *signalling*; B2 -- self-inhibitory feedback of *establishment of localization*; B4 -- negative feedback loop from *cellular component organization signalling cell differentiation homeostatic process system development carbohydrate derivative metabolic process nucleic acid templated transcription cellular component organization*. (**Supplementary Tables 4 and 5**).

Instances of localized impact structure

Localized interactions are also noted in the two network models, especially the two afrorementioned canonical players in mouse embryo development: *system development* and *cell differentiation* (cyan and orange edges; **Figure 3B and 3D**; **Table 1**; **Supplementary Table 1**) (The Gene Ontology Consortium, 2017). Markedly, state changes in *cell differentiation* negatively regulate state changes in *signalling* and *homeostatic process* (**Figure 4C**; **Supplementary Materials**). *System development* negatively regulates nucleic acid transcription, consistent with previous reports (Pan et al., 2006; Papatsenko et al., 2018). Intriguingly, *system development* also influences on *carbohydrate derivative metabolic process* (with a magnitude up to 40%).

Strikingly, we observed pronounced negative feedback of *lipid metabolic process* on itself (loop B1) in both trajectories (brown edges; **Figure 3B and 3D**). State change in *lipid metabolic process* reinforces state changes in *signalling* while it counteracts state changes in *establishment of localization*, the primary and secondary hubs of the macroscale network (**Figure 4C**; **Supplementary Materials**). Moreover, the self-inhibitory feedback of *lipid metbolic process* is if not more, at least as explanatory of the netword dynamics as the self-reinforcement of *signaling* (loop R1): explaining 3.15% and 3.07% of the total system behaviour in T1 and T2, respectively (**Figure 3 and 5**; **Supplementary Tables 4 and 5**). It is noted that loop B1 persistently precedes loop R1 in time (**Figure 5**). Considering the mutual pairwise causal relationships of *signalling* (**Figure 4A**) and *lipid metabolic process* (**Figure 4C**), the two functional domains are plausibly entangled in a relationship resembling a bistable switch (**Figure 6**). This relationship comprises a recurring scene of information flow during cell fate transition, echoing recent developement on understanding the function of peroxisome and PPAR proteins (Ahmadian et al., 2013; Londi and Semenkovich, 2014; Liu et al., 2018). This represents an axis in regulating network dynamics during mouse organogenesis that warrants further investigation.

In addition, incidental impact of *carbohydrate derivative metabolic process* on *signalling*, *homeostatic process*, *nucleic acid templated transcription*, and *lipid metabolic process* is also noted (**Supplementary Materials**). Metabolic state of the cell reflects the availability of nutrients and cellular capacity to use them effectively. Metabolic activity thus fine-tunes the signaling network in the cell (Wellen and Thompson, 2012; Gomes and Blenis, 2015; Causton, 2018). Together, this reveals the essential role of metabolism, as an effector of mass and energy flow, in the network feedback structure governing cell fate transition. Again, this emphasises the interactions between regulator and effector processes in controlling network dynamics during organogenesis.

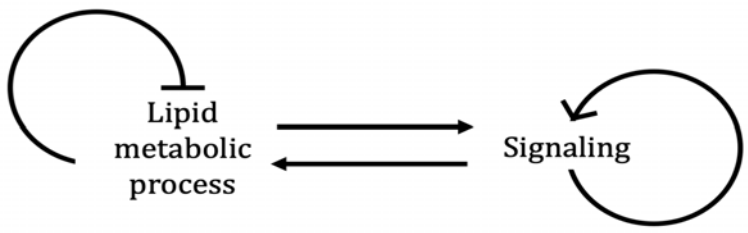

Figure 6. Hypothesized bistable feedback relationship between *lipid metabolic process* and *signaling*.

Discussions

Transcriptome dynamics underlie cell fate transition during embryo development. Besides the actions of master regulators and biomarkers, it is increasingly clear that the collective state of the transcription network reflects and effects cell identity. In this study, we provide a formal description of macroscale transcriptome dynamics during mouse organogenesis by constructing a causal model and delineating the feedback structure embedded in the network topology. Our analysis characterises key aspects of information flow across the network that are known to induce changes and uncovers features that have not been acknowledged in full. We point out new axes of inter-relationships between ontological regulator and effector processes, especially between *signalling* and *lipid metabolic process*. These interactions plausibly control information flow during mouse organogenesis, which warrants further investigation. Encompassing the entire transcriptome, we have thus integrated mass, energy, and information transfer within the same analytical framework.

The causal models we have generated for the macroscale gene regulatory network are time-dependent nonlinear feedback systems that is biologically interpretable, offering a dynamic view of the transcriptome during mouse organogenesis. The bird's-eye view offers a vantage point for us to examine the large-scale organization of the system during a phase of drastic transformation at the organismal level. During this process, we observed that the macorscale network is highly organised and the information flow is intrinsically robust. This regularity prompts us to postulate that the macroscale feedback structure epitomizes the origin of stability of the system during its natural progression. If true, it would be of interest to compare our causal models to those generated from biological systems under pathological conditions, such as cancer (Rockne et al., 2020; Rozenblatt-Rosen et al., 2020). Meanwhile, this macroscale representation inevitably overshadows granularity that is required for experimental/clinical actions. Nevertheless, given the largely hierarchical organization of GO annotations, it is feasible to aggregate at the level of concrete subsystems, such as different pathways invovled in DNA damage response, instead of the top-level biological processes. This allows for a mesoscale representation of the network, which is currently an active area of research (Goodsell et al., 2020).

Due to pervasive nonlinear interactions and combinatorial circuilarity in feedbacks, network dynamics of biological systems remains a challenge to investigate (Kurz et al., 2016). In this article, we have inferred network feedback structure by constructing a causal

model with no prior specification. Building upon network learning algorithm and numerical methods, this analytic framework not only offers means to visualize network dynamics, but also pinpoints essential characteristics of the network topology that underlie systemic dynamics. Many real-world networks share properties with gene regulatory network. It is conceivable that this analytic framework is readily applicable to these fields.

Methods

Data processing

Single-cell RNA-seq data from developing mouse embryos (C57BL/6 genetic background) over a time-course of 48 hours (embryonic day E6.5-E8.5) was generated and curated by Pijuan-Sala et al., (2019). Raw mapped read counts covering 29,452 genes across the genome in 116,312 cells were downloaded from https://github.com/MarioniLab/EmbryoTimecourse2018. We processed the data following authors' instructions, including doublet removal, reads normalization, and batch correction. After initial processing, cells of endoderm lineage were selected by clustering information available in meta-data (cluster 9 [excluding sub-cluster 1 & 3] and cluster 11). Cells were split into two developmental trajectories based on cell type annotations: trajectory 1 includes cells annotated as anterior primitive streak, definitive endoderm, and gut; trajectory 2 includes cells annotated as visceral endoderm and gut. Due to low sampling at early time points in trajectory 1, data imputation was performed to reach a balanced sample sizes in both trajectories, especially at the early time points. In brief, UMAPs for all cell types in stage E6.5 and E6.75 were calculated and force-directed graphs were constructed in line with the original study. Neighbours of cells in trajectory 1 were identified by measuring Euclidean distance to the centroid and cell type annotation (epiblast and primitive steak). With a distance cut-off of 2.6, we imputed 125 'similar' cells at E6.5 and 23 at E6.75. As a result, we have 3,588 cells in trajectory 1 and 3,072 cells in trajectory 2. All gene expression values are normalized to sequencing size factor and log2 transformed. Data handling was performed in R (version 3.6.0).

Selection of highly variable genes

Variance in gene expression was modelled using 'modelGeneVar' function from the scran R package, with loess span of 0.05 and block settings on 3 sequencing batches. Genes with significantly higher variance than the fitted trend (Benjamini–Hochberg-corrected $P < 0.05$) and positive biological component were selected. Genes were excluded if the mean log2 normalized count is below 10−3, on the Y chromosome, or is Xist.

Macroscale modeling of gene regulatory network

Functions of genes and gene products are annotated based on Gene Ontology (GO), a literature-curated reference database (annotation data was downloaded from Mouse Genome Informatics: http://www.informatics.jax.org). In the partially hierarchical structure of the GO directed acyclic graph, child-parent relation links terms of lower-level biological subsystems, such as small complexes or single reactions, to those pertaining to

upper-level biological systems, such as organelles or broad processes (Explorable on Mouse Genome Informatics: http://www.informatics.jax.org/vocab/gene*ontology). To build a macroscale representation of the mouse gene network, we utilized such membership information from the cut-down version of the GO maintained by Mouse Genome Informatics (Mouse GO slim, file name 'goslim*mouse'). This was performed using OWLTools (https://github.com/owlcollab/owltools/wiki/Map2Slim), to aggregate all the lower-level annotations to 14 top-level modules (macroscale functional domains of biological processes). Each annotated gene is tagged with at least one of the 14 biological processes.

Data aggregation

For each trajectory, log2 normalized read counts of all sampled HVGs in each single cell were aggregated at each of the 9 time points (Embryonic day E6.5-E8.5, sampled every 6 hours). For each time points, cells were randomly split into four equal-length sets (four-fold learning scheme). For each set, the average expression level of each gene at a time point was calculated by taking the arithmetic mean of the log2 normalized read counts of the gene in all cells staged at the specific time point. Second, the average expression level of all genes tagged in the same macroscale process were summed at each time point, generating a time-series of state values, representing the gross expression level of each biological process at each time point. As a result, we have 8 sets of data: two trajectories, each consists of four sets of 14 time-series. Each time-series represents the dynamic state values of the macroscale biological process of interest.

Network learning scheme

To construct a causal model of the macroscale network capturing the system dynamics, we built a FSNN (Feedback System Neural Network) (Schoenberg, 2020), which is an ordinary differential equation (ODE)-based machine learning method (Eq. 1).

$$\frac{d\overrightarrow{BP}}{dt} = tanh\left(\sum \frac{\overrightarrow{BP}}{\overrightarrow{mag}} \cdot \overrightarrow{weight} + \overrightarrow{bias}\right) \cdot \overrightarrow{mag} \quad \text{(Eq. 1)}$$

Our FSNN had 14 states ($\overrightarrow{BP}$) and was initialized with all weights and biases set to 0, thus initially representing a fully disconnected network. The initial value of each state was loaded from the data at E6.5. The FSNN for the model of both trajectories used a single perceptron to model the change in each biological process. The input to each perceptron was the current value of all of the 14 biological processes. The FSNN differential equation was simulated using the RK4 algorithm with a Δt of 1/16 (unit: day). The state values were lineally rescaled using a constant magnitude ($\overrightarrow{mag}$) which was calculated as 1.5 times the initial value of the state from the first training set to avoid signal attenuation issues due to the $tanh$ activation function. We trained the FSNN using 3/4 of the data and tested on the remaining 1/4. The FSNN was constructed and trained using Stella Architect 2.0 which uses Powell's BOBYQA optimization algorithm as implemented in DLIB version 19.7 to determine the weights and biases. The two trajectories were trained and tested separately, thus generating two distinct models.

Inference on pairwise causal relationship

The two causal models (FSNNs) of the macroscale network were then analysed to infer direct pairwise impact relations from source process to target process. Here, we utilized the link score measure (Eq.2; Schoenberg et. al, 2019) to determine the contribution of one biological process to the change of another in the generated models.

$$Link\ Score(x \rightarrow y) = \begin{cases} \left| \frac{\Delta_x y}{\Delta y} \right| \cdot\ sign\left(\frac{\Delta_x y}{\Delta x}\right), \\ 0, \Delta y = 0\ or\ \Delta x = 0 \end{cases} \quad \text{(Eq. 2)}$$

The sign of the link score indicates the polarity of the relationship. A positive value means that the change in the source variable produces a change in the same direction in the target variable, independent of state chnges in other variable; while a negative value means the opposite. The magnitude (after normalization across all incoming variables to the target) demonstrates the relative strength of the causal relationship of interest relative to all the other incoming sources to a given target. The mean and variance of link scores were derived from the training and testing sets of time-series data. This analysis was conducted using Stella Architect 2.0.

Feedback loop dominance analysis

To discover the origins of behaviour for each causal model from a feedback loop perspective, feedback loop dominance analysis was done using the Loops that Matter method (Schoenberg et. al, 2019) implemented in Stella Architect 2.0. The feedback loop discovery process was done using the SPA algorithm because it was impossible to enumerate and measure the 14 factorial feedback loops produced by each FSNN (Eberlein & Schoenberg, 2020). The analysis was performed on just the testing set for each model. The feedback loop dominance analysis measures the share of model behaviour which is attributable to each feedback loop in the model at each time point. Those values were averaged over the full simulation period to measure the total contribution of each feedback loop to the behaviour of each model over the full simulation time period (Schoenberg & Eberlein, 2020) to clarify the most impactful feedback processes during organogenesis.

Data visualization

A static view of the feedback structure of each model was produced using the link score data from the testing set of each model. Network visualization were created for each model where each node represented a biological process, and each edge represented the average relative link score measured across the full simulation period. The number of edges was filtered, and we selected the upper quartile of strongest connections. Gephi (v0.91; Bastian et al., 2009) was used to create map of the feedback structure. In addition, using Gephi, overtime videos of the progression of causal influences were created using the link scores measured at each dt of the FSNN, where again, only the upper quartile of edges were plotted.

References

Ahmadian, Maryam, Jae Myoung Suh, Nasun Hah, Christopher Liddle, Annette R Atkins, Michael Downes, and Ronald M Evans. 2013. 'PPARγ Signaling and Metabolism: The Good, the Bad and the Future'. *Nature Medicine* 19 (5): 557–66. https://doi.org/10.1038/nm.3159.

Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, et al. 2017. 'SCENIC: Single-Cell Regulatory Network Inference and Clustering'. *Nature Methods* 14 (11): 1083–86. https://doi.org/10.1038/nmeth.4463.

Arnold, Sebastian J., and Elizabeth J. Robertson. 2009. 'Making a Commitment: Cell Lineage Allocation and Axis Patterning in the Early Mouse Embryo'. *Nature Reviews Molecular Cell Biology* 10 (2): 91–103. https://doi.org/10.1038/nrm2618.

Aulehla, Alexander, and Olivier Pourquié. 2008. 'Oscillating Signaling Pathways during Embryonic Development'. *Current Opinion in Cell Biology* 20 (6): 632–37. https://doi.org/10.1016/j.ceb.2008.09.002.

Balazsi, G., A.-L. Barabasi, and Z. N. Oltvai. 2005. 'Topological Units of Environmental Signal Processing in the Transcriptional Regulatory Network of Escherichia Coli'. *Proceedings of the National Academy of Sciences* 102 (22): 7841–46. https://doi.org/10.1073/pnas.0500365102.

Basson, M. Albert. 2012. 'Signaling in Cell Differentiation and Morphogenesis'. *Cold Spring Harbor Perspectives in Biology* 4 (6). https://doi.org/10.1101/cshperspect.a008151.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. 'Gephi: An Open Source Software for Exploring and Manipulating Networks'. *Proceedings of the Third International ICWSM Conference*. http://w ww.aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

Bult, Carol J, Judith A Blake, Cynthia L Smith, James A Kadin, Joel E Richardson, the Mouse Genome Database Group, A Anagnostopoulos, et al. 2019. 'Mouse Genome Database (MGD) 2019'. *Nucleic Acids Research* 47 (D1): D801–6. https://doi.org/10.1093/nar/gky1056.

Cao, Junyue, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, et al. 2019. 'The Single-Cell Transcriptional Landscape of Mammalian Organogenesis'. *Nature* 566 (7745): 496–502. https://doi.org/10.1038/s41586-019-0969-x.

Casadiego, Jose, Mor Nitzan, Sarah Hallerberg, and Marc Timme. 2017. 'Model-Free Inference of Direct Network Interactions from Nonlinear Collective Dynamics'. *Nature Communications* 8 (1): 2192. https://doi .org/10.1038/s41467-017-02288-4.

Causton, Helen C. 2020. 'Metabolic Rhythms: A Framework for Coordinating Cellular Function'. *European Journal of Neuroscience* 51 (1): 1–12. https://doi.org/10.1111/ejn.14296.

Cheng, Shangli, Yu Pei, Liqun He, Guangdun Peng, Björn Reinius, Patrick P.L. Tam, Naihe Jing, and Qiaolin Deng. 2019. 'Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition and X Chromosome Dynamics during Early Mouse Development'. *Cell Reports* 26 (10): 2593-2607.e3. https://doi.org/10.1016/j.celrep.2019.02.031.

Davidson, E. H. 2002. 'A Genomic Regulatory Network for Development'. *Science* 295 (5560): 1669–78. https ://doi.org/10.1126/science.1069883.

DLIB [Computer software]. (2017). Retrieved from https://github.com/davisking/dlib

Eberlein, R. & Schoenberg, W. (2020). Finding the Loops that Matter.

Ferrell, James E. 2012. 'Bistability, Bifurcations, and Waddington's Epigenetic Landscape'. *Current Biology* 22 (11): R458–66. https://doi.org/10.1016/j.cub.2012.03.045.

Ferrell, James E. 2016. 'Perfect and Near-Perfect Adaptation in Cell Signaling'. *Cell Systems* 2 (2): 62–67. https://doi.org/10.1016/j.cels.2016.02.006.

Fondi, Marco, and Pietro Liò. 2015. 'Genome-Scale Metabolic Network Reconstruction'. In *Bacterial Pangenomics*, edited by Alessio Mengoni, Marco Galardini, and Marco Fondi, 1231:233–56. Methods in Molecular Biology. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-1720-4_15.

Fraser, H. B. 2002. 'Evolutionary Rate in the Protein Interaction Network'. *Science* 296 (5568): 750–52. https://doi.org/10.1126/science.1068696.

Fujii, Keisuke, Naoya Takeishi, Motokazu Hojo, Yuki Inaba, and Yoshinobu Kawahara. 2020. 'Physically- Interpretable Classification of Biological Network Dynamics for Complex Collective Motions'. *Scientific Reports* 10 (1): 3005. https://doi.org/10.1038/s41598-020-58064-w.

Gilbert, Scott F. 2014. *Developmental Biology*. 10. ed. Sunderland, Mass: Sinauer Assoc.

Gomes, Ana P, and John Blenis. 2015. 'A Nexus for Cellular Homeostasis: The Interplay between Metabolic and Signal Transduction Pathways'. *Current Opinion in Biotechnology* 34 (August): 110–17. https://doi.org/10.1016/j.copbio.2014.12.007.

Goodsell, David S., Arthur J. Olson, and Stefano Forli. 2020. 'Art and Science of the Cellular Mesoscale'. *Trends in Biochemical Sciences* 45 (6): 472–83. https://doi.org/10.1016/j.tibs.2020.02.010.

Groß, Alexander, Barbara Kracher, Johann M. Kraus, Silke D. Kühlwein, Astrid S. Pfister, Sebastian Wiese, Katrin Luckert, et al. 2019. 'Representing Dynamic Biological Networks with Multi-Scale Probabilistic Models'. *Communications Biology* 2 (1): 21. https://doi.org/10.1038/s42003-018-0268-3.

Han, Jing-Dong J., Nicolas Bertin, Tong Hao, Debra S. Goldberg, Gabriel F. Berriz, Lan V. Zhang, Denis Dupuy, et al. 2004. 'Evidence for Dynamically Organized Modularity in the Yeast Protein–Protein Interaction Network'. *Nature* 430 (6995): 88–93. https://doi.org/10.1038/nature02555.

Harush, Uzi, and Baruch Barzel. 2017. 'Dynamic Patterns of Information Flow in Complex Networks'. *Nature Communications* 8 (1): 2181. https://doi.org/10.1038/s41467-017-01916-3.

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. 2000. 'The Large-Scale Organization of Metabolic Networks'. *Nature* 407 (6804): 651–54. https://doi.org/10.1038/35036627.

Kurz, Felix T., Jackelyn M. Kembro, Ana G. Flesia, Antonis A. Armoundas, Sonia Cortassa, Miguel A. Aon, and David Lloyd. 2017. 'Network Dynamics: Quantitative Analysis of Complex Behavior in Metabolism, Organelles, and Cells, from Experiments to Models and Back: Network Dynamics'. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 9 (1): e1352. https://doi.org/10.1002/wsbm.1352.

Liu, Yi, Jennifer Colby, Xiangsheng Zuo, Jonathan Jaoude, Daoyan Wei, and Imad Shureiqi. 2018. 'The Role of PPAR-δ in Metabolism, Inflammation, and Cancer: Many Characters of a Critical Transcription Factor'. *International Journal of Molecular Sciences* 19 (11): 3339. https://doi.org/10.3390/ijms19113339.

Lodhi, Irfan J., and Clay F. Semenkovich. 2014. 'Peroxisomes: A Nexus for Lipid Metabolism and Cellular Signaling'. *Cell Metabolism* 19 (3): 380–92. https://doi.org/10.1016/j.cmet.2014.01.002.

Lu, Rong, Florian Markowetz, Richard D. Unwin, Jeffrey T. Leek, Edoardo M. Airoldi, Ben D. MacArthur, Alexander Lachmann, et al. 2009. 'Systems-Level Dynamic Analyses of Fate Change in Murine Embryonic Stem Cells'. *Nature* 462 (7271): 358–62. https://doi.org/10.1038/nature08575.

Luscombe, Nicholas M., M. Madan Babu, Haiyuan Yu, Michael Snyder, Sarah A. Teichmann, and Mark Gerstein. 2004. 'Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes'. *Nature* 431 (7006): 308–12. https://doi.org/10.1038/nature02782.

Ma, Jianzhu, Michael Ku Yu, Samson Fong, Keiichiro Ono, Eric Sage, Barry Demchak, Roded Sharan, and Trey Ideker. 2018. 'Using Deep Learning to Model the Hierarchical Structure and Function of a Cell'. *Nature Methods* 15 (4): 290–98. https://doi.org/10.1038/nmeth.4627.

MacArthur, Ben D., Avi Ma'ayan, and Ihor R. Lemischka. 2009. 'Systems Biology of Stem Cell Fate and Cellular Reprogramming'. *Nature Reviews Molecular Cell Biology* 10 (10): 672–81. https://doi.org/10.1038/nr m2766.

Neph, Shane, Andrew B. Stergachis, Alex Reynolds, Richard Sandstrom, Elhanan Borenstein, and John A. Stamatoyannopoulos. 2012. 'Circuitry and Dynamics of Human Transcription Factor Regulatory Networks'. *Cell* 150 (6): 1274–86. https://doi.org/10.1016/j.cell.2012.04.040.

Pan, Guangjin, Jun Li, Yali Zhou, Hui Zheng, Duanqing Pei, Guangjin Pan, Jun Li, Yali Zhou, Hui Zheng, and Duanqing Pei. 2006. 'A Negative Feedback Loop of Transcription Factors

That Controls Stem Cell Pluripotency and Self-renewal'. *The FASEB Journal* 20 (10): 1730–32. https://doi.org/10.1096/fj.05-5543fje.

Papatsenko, Dmitri, Avinash Waghray, and Ihor R. Lemischka. 2018. 'Feedback Control of Pluripotency in Embryonic Stem Cells: Signaling, Transcription and Epigenetics'. *Stem Cell Research* 29 (May): 180–88. https://doi.org/10.1016/j.scr.2018.02.012.

Peixoto, Tiago P., and Martin Rosvall. 2017. 'Modelling Sequences and Temporal Networks with Dynamic Community Structures'. *Nature Communications* 8 (1): 582. https://doi.org/10.1038/s41467-017-00148-9.

Pijuan-Sala, Blanca, Jonathan A. Griffiths, Carolina Guibentif, Tom W. Hiscock, Wajid Jawaid, Fernando J. Calero-Nieto, Carla Mulas, et al. 2019. 'A Single-Cell Molecular Map of Mouse Gastrulation and Early Organogenesis'. *Nature* 566 (7745): 490–95. https://doi.org/10.1038/s41586-019-0933-9.

Rockne, Russell C, Sergio Branciamore, Jing Qi, David E Frankhouser, Denis O'Meally, Wei-Kai Hua, Guerry Cook, et al. 2020. 'State-Transition Analysis of Time-Sequential Gene Expression Identifies Critical Points That Predict Development of Acute Myeloid Leukemia'. *Cancer Research*, May, canres.0354.2020. https://doi.org/10.1158/0008-5472.CAN-20-0354.

Rozenblatt-Rosen, Orit, Aviv Regev, Philipp Oberdoerffer, Tal Nawy, Anna Hupalowska, Jennifer E. Rood, Orr Ashenberg, et al. 2020. 'The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution'. *Cell* 181 (2): 236–49. https://doi.org/10.1016/j.cell.2020.03.053.

Schoenberg, William, Pål Davidsen, and Robert Eberlein. 2019. 'Understanding Model Behavior Using Loops That Matter'. *ArXiv:1908.11434 [Physics]*, August. http://arxiv.org/abs/1908.11434.

Schoenberg, William. 2020. 'Feedback System Neural Networks for Inferring Causality in Directed Cyclic Graphs'. *ArXiv:1908.10336 [Cs, Stat]*, May. http://arxiv.org/abs/1908.10336.

Schoenberg, W. & Eberlein, R. (2020). Seamlessly Integrating Loops That Matter into Model Development and Analysis

Sladitschek, Hanna L., Ulla-Maj Fiuza, Dinko Pavlinic, Vladimir Benes, Lars Hufnagel, and Pierre A. Neveu. 2020. 'MorphoSeq: Full Single-Cell Transcriptome Dynamics Up to Gastrulation in a Chordate'. *Cell* 181 (4): 922-935.e21. https://doi.org/10.1016/j.cell.2020.03.055.

Stella Architect 2.0 [Computer software]. (2020). Retrieved from https://iseesystems.com/store/products/ste lla-architect.aspx

Tam, Patrick P. L., and David A. F. Loebel. 2007. 'Gene Function in Mouse Embryogenesis: Get Set for Gastrulation'. *Nature Reviews Genetics* 8 (5): 368–81. https://doi.org/10.1038/nrg2084.

The Gene Ontology Consortium. 2017. 'Expansion of the Gene Ontology Knowledgebase and Resources'. *Nucleic Acids Research* 45 (D1): D331–38. https://doi.org/10.1093/nar/gkw1108.

Tu, B. P. 2005. 'Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes'. *Science* 310 (5751): 1152–58. https://doi.org/10.1126/science.1120499.

Wellen, Kathryn E., and Craig B. Thompson. 2012. 'A Two-Way Street: Reciprocal Regulation of Metabolism and Signalling'. *Nature Reviews Molecular Cell Biology* 13 (4): 270–76. https://doi.org/10.1038/nrm3305.

Young, Richard A. 2011. 'Control of the Embryonic Stem Cell State'. *Cell* 144 (6): 940–54. https://doi.org/10.1016/j.cell.2011.01.032.

Zou, Yong, Reik V. Donner, Norbert Marwan, Jonathan F. Donges, and Jürgen Kurths. 2019. 'Complex Network Approaches to Nonlinear Time Series Analysis'. *Physics Reports* 787 (January): 1–97. https://doi.org/10.1016/j.physrep.2018.10.005.