Reconstruction of dynamic regulatory networks reveals signaling-induced topology changes associated with germ layer specification

Emily Y. Su[1,2*], Abby Spangler[1*], Qin Bian[1,2], Jessica Y. Kasamoto[2], Patrick Cahan[1,2,3]

[1]Institute for Cell Engineering, Johns Hopkins School of Medicine, Baltimore MD 21205 USA

[2]Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore MD 21205 USA

[3]Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore MD 21205 USA

* These authors made equal contributions.

Correspondence to: patrick.cahan@jhmi.edu

Declarations of interest: none

Keywords: Single cell RNA-seq, Gene regulatory network, Network inference, Directed differentiation, Gastrulation, Embryoid body, Peg3, Wnt

Abstract

Elucidating regulatory relationships between transcription factors (TFs) and target genes is fundamental to understanding how cells control their identity and behavior. Computational gene regulatory network (GRN) reconstruction methods aim to map this control by inferring relationships from transcriptomic data. Unfortunately, existing methods are imprecise, may be computationally burdensome, and do not illustrate how networks transition from one topology to another. Here we present Epoch, a computational network reconstruction tool that leverages single cell transcriptomics to infer dynamic network structures. Epoch performs favorably when benchmarked using data derived from *in vivo*, *in vitro,* and *in silico* sources. To illustrate the usefulness of Epoch, we applied it to identify the dynamic networks underpinning directed differentiation of mouse embryonic stem cells (ESC) guided by multiple primitive streak induction treatments. Our analysis demonstrates that modulating signaling pathways drives topological network changes that shape cell fate potential. We also find that Peg3 is a central contributor to the rewiring of the pluripotency network to favor mesoderm specification. By integrating signaling pathways with GRN structures, we traced how Wnt activation and PI3K suppression govern mesoderm and endoderm specification, respectively. The methods presented here are available in the R package Epoch, and provide a foundation for future work in understanding the biological implications of dynamic regulatory structures.

Introduction

Gene regulatory networks (GRNs) model the regulatory relationship between a set of regulators, or transcription factors (TFs), and their target genes. The topology of these networks, defined by edges that map regulatory interactions between TFs and targets, offers a molecular-level view of a controlled system in which genes work together as part of a framework to accomplish specific cell functions (Karlebach and Shamir, 2008; Le Novère, 2015). Uncovering the topology of these GRNs is fundamental in answering a number of questions including: understanding how cellular identity is maintained and established (Davidson and Erwin, 2006), elucidating mechanisms of disease caused by dysfunctional gene regulation (Morgan et al., 2020; Qin et al., 2019), and finding novel drug targets among others (Carro et al., 2010). In the context of cell fate engineering, accurate mapping of such GRNs would thus enable the identification of TFs required to activate expression of target genes of interest such that we could more accurately control cell fate transitions or cell behavior (Cahan et al., 2021; Rackham et al., 2016). Unfortunately, how best to map these relationships remains both an experimental and computational challenge.

Experimental approaches, including ChIP-ChIP and ChIP-seq, have provided valuable insights into identifying regulatory targets and TF binding site motifs of specific TFs in certain cell lines and cell types. Similarly, chromatin accessibility assays that detect TFBS footprints at regulatory genomic regions, such as ATAC-Seq (Buenrostro et al., 2013), enable the inference of regulatory networks. However, these types of experimental approaches are generally limited in scope (e.g. ATAC-defined networks are limited to TFs with known motifs) and scalability (e.g. it is infeasible to perform ChIP-seq for all TFs in all cell types). For these reasons, computational methods to efficiently infer GRN structures are needed. Existing methods have leveraged advances in data collection and machine learning approaches that have allowed for the statistical inference of GRNs via gene expression data. These include tools that rely on information theoretic measures (Faith et al., 2007; Margolin et al., 2006; Meyer et al., 2007), ensemble methods (Huynh-Thu et al., 2010), Bayesian approaches (Yu et al., 2004; Hartemink, 2005), or differential equation-based (ODE) approaches (di Bernardo et al., 2005). Unfortunately, computational methods developed to date suffer from several drawbacks including low precision and low sensitivity. Some of the leading contributors to the low performance include the difficulty in distinguishing direct from indirect interactions (Marbach et al., 2010), the confounding effects of Simpson's Paradox (Trapnell, 2015), and the fact that bulk derived data does not offer perturbation sufficient to detect true regulatory relationships (Stark et

al., 2003). Attempts to ameliorate these by aggregation across methods have met with modest success (Marbach et al., 2012).

With the advent of droplet-based systems for single-cell RNA-sequencing (Klein et al., 2015; Macosko et al., 2015; Zheng et al., 2017), and the corresponding rapidly growing body of single-cell transcriptomic data, computational analysis techniques have seen widespread advancement in harnessing the single-cell resolution to identify cell types and states (Villani et al., 2017) and to trace developmental trajectories (Treutlein et al., 2014). Of particular interest, trajectory inference (TI) methods, or pseudotemporal ordering methods, such as the transcriptomic-based methods Monocle (Trapnell et al., 2014), Slingshot (Street et al., 2018), and DPT (Haghverdi et al., 2016), as well as cytometry-based methods such as SPADE (Qiu et al., 2011), have allowed for computational modelling of dynamic processes like cell differentiation. Such methods order cells in an unbiased fashion along a trajectory following a linear, bifurcating, or more complex graph structure. In doing so, TI methods have proven valuable in the identification of novel cell subpopulations (Chen et al., 2016) and the elucidation of differentiation trees (Velten et al., 2017).

Unsurprisingly, several recently published tools have aimed to take advantage of the resolution offered by single-cell transcriptomics to infer GRN structures (Aibar et al., 2017; Matsumoto et al., 2017; Qiu et al., 2020). Just like their bulk counterparts, these methods can be categorized into information theoretic, ensemble, Bayesian, and ODE approaches. Unfortunately, they also suffer from similar limitations, and aside from low precision and sensitivity (Chen and Mar, 2018), many of these methods are computationally burdensome, limiting users' ability to hone analysis through iterative application (Bonnaffoux et al., 2019). Additionally, though many of these methods employ pseudotemporal analysis to aid in reconstruction, they are limited in their ability to explain how network topology evolves over time. Importantly, a dynamic topology implies that regulatory relationships between TFs and targets change over time and thereby increases the number of reachable cell states (**Figure 1**). More specifically, the dynamic and noncommutative nature of regulatory networks permits independent control of targets or genetic programs that would otherwise be simultaneously activated with non-sequential, or combinatorial, logic (Letsou and Cai, 2016). Thus, for GRNs to more accurately model the emergence of distinct cell fates in development, they must encapsulate this dynamic behavior changing GRN topology. Moreover, because GRN topology dictates how a cell responds to perturbation, uncovering the dynamic GRN aids in understanding how the landscape of reachable cell states changes over time, which has significant implications in the quest to engineer cell fate.

For those reasons, we developed a computational GRN reconstruction tool called Epoch, which leverages single-cell transcriptomics to efficiently reconstruct dynamic network structures. Three key strategies exist at the center of Epoch: first, reconstruction is limited to dynamically expressed genes. Second, Epoch employs an optional “cross-weighting” strategy to reduce false positive interactions. Third, Epoch divides pseudotime into “epochs”, or time periods, and extracts a dynamic network and predicted top regulators influential in driving topology changes. Finally, Epoch includes a number of functionalities to aid in network analysis and comparison, including integration of the reconstructed GRN with major signaling pathways and the subsequent tracing through the GRN of shortest paths from signaling effectors toward activation or repression of target genes.

To demonstrate the utility of our tool we benchmarked Epoch’s performance against commonly used computational GRN reconstruction tools on synthetically generated data, *in vivo* mouse muscle development data, and *in vitro* directed differentiation data. We demonstrate that Epoch performs well in precision-recall analysis and in runtime comparison. We further highlight how Epoch’s stepwise-structured pipeline allows for user flexibility in adapting it to other GRN reconstruction methods.

Finally, we applied Epoch to answer open questions in gene regulation and directed differentiation. In particular, as pluripotent cells differentiate and undergo gastrulation, their transcriptomic profiles diverge, ultimately specifying and committing to distinct germ layers. Pluripotent cells undergoing *in vitro* differentiation often do so via embryoid body (EB) formation, which mimics *in vivo* gastrulation events such as primitive streak formation. Importantly, cell migration and specification *in vivo* are guided by a tightly-orchestrated combination of signals. Notably, Wnt3a, Bmp4, and Nodal concentration gradients (generated by Bmp and Wnt antagonist secretion on the anterior side of the embryo) are established early on, and are responsible for initiating primitive streak (Brennan et al., 2001; Yamamoto et al., 2004). For example, Wnt3a activates Brachyury in posterior epiblast cells, generating mesodermal cells (Perea-Gomez et al., 2002). Additionally, a number of FGFs have roles in coordinating events in early embryo development including the exit from pluripotency. For example, primitive streak cells both synthesize and respond to FGFs, including Fgf4 and Fgf8, the absence of which results in failure of cells to migrate away from the primitive streak, resulting in the lack of mesodermal and endodermal tissue development and the disrupted patterning of the neuroectoderm (Sun et al., 1999). The orchestration of development by signaling pathways is mediated by the direct regulation of distinct transcriptional targets by pathway effectors (**Figure 1**). However, signaling cascades also have non-transcriptional consequences, including the

alteration of epigenomic state and thus GRN topology (Fagnocchi et al., 2016; Mohammad and Baylin, 2010). The extent to which signaling pathways influence cell fate by altering GRN topology is unknown. To address this question, we performed scRNA-seq on day 0 through day 4 *in vitro* mouse embryonic stem cell (mESC) directed differentiation guided by four separate primitive streak induction treatments. We used Epoch to reconstruct GRNs associated with each treatment and compared the resulting topologies.

Specifically, we used Epoch to explore why cells treated with different induction treatments exhibited differing mesodermal fate potential, the lineage that was overall least likely chosen by differentiating cells. Our analysis revealed a number of topological features affecting mesodermal fate potential, including a central role for Peg3 in regulating the exit of pluripotency and transition to mesodermal fate. Additionally, by integrating signal transduction pathways with Epoch-reconstructed networks, we found that primitive streak induction via GSK inhibition produced a network topology that allows for more effective mesodermal specification. This was in direct contrast to other treatments in which network topologies more easily specified endodermal fate through suppression of PI3K signaling and subsequent activation of Foxo1. Indeed, there were remarkably large differences between target genes of the same TF across treatments, supporting a model where activation of signaling pathways directly alters network topology to influence cell fate potential. We believe that these results and similar analyses can be used to hone more efficient differentiation of mesodermal-fated cells. Further, we expect that analogous analyses can be performed to explain fate decisions in other dynamic biological processes.

Results

Epoch workflow relies on single-cell analysis techniques to infer dynamic network structures

Epoch takes as minimum input processed single-cell transcriptomic data and pseudotime or equivalent annotation from any trajectory inference method a user chooses to utilize (**Figure 2**). From here, Epoch employs its first step: limiting reconstruction to dynamically expressed genes in an effort to both focus on genes playing an active role in cell state changes, and to ensure that identified interactions are ones that can be inferred from perturbations in transcriptomic data. To accomplish this, Epoch employs a method reminiscent of Slingshot (Street et al., 2018), in which individual genes are modeled across pseudotime using a generalized additive model. Upon filtering the data for dynamically expressed genes, Epoch

reconstructs an initial static network via a CLR-like (Context Likelihood of Relatedness) method, employing either Pearson correlation or mutual information to infer interactions (Faith et al., 2007).

In the second step, which we have termed “cross-weighting”, Epoch employs an optional network refinement step that relies on a cross-correlation-based weighting scheme. This helps to reduce false positives that may result from indirect interactions or that may represent non-logical interactions, ultimately improving precision (**Supplementary Figure 1a**). In this step, the expression profiles over pseudotime for each TF-target pair is aligned and progressively shifted to determine an average estimated offset value at which maximum correlation is achieved between the two profiles. A graded-decline weighting factor is computed based on this offset, and used to negatively weight interactions that are less likely to be true positives.

Once a static network has been inferred, Epoch extracts a dynamic network from the static network, and further identifies the top regulators driving changes in network topology and thus cell state. Specifically, this begins with Epoch breaking down pseudotime into “epochs”, or discrete time periods, based on pseudotime, cell ordering, k-means clustering, or user-defined assignment. From here, genes are assigned to epochs based on their activity along pseudotime (a given gene may be assigned to multiple epochs). Epoch then fractures the static network into a dynamic one, composed of “epoch networks”, representing active interactions within a particular epoch, and possible “transition networks”, describing how an epoch network transitions into a subsequent epoch network.

As Epoch was initially built with the improvement of cell fate engineering protocols in mind, we included in the Epoch framework a number of network analysis and network comparison capabilities. For example, Epoch will identify influential transcription factors within a given static or dynamic network by extracting top regulators that appear to have the most influence on network state. Specifically, Epoch will compute and rank TFs according to their PageRank (Brin and Page, 1998) in the context of their epoch networks, and will further look for TFs that simultaneously score high in betweenness and degree centralities as compared to all other TFs. Additionally, Epoch can integrate major signal transduction pathways with reconstructed networks. This can be used to determine the shortest paths through the networks from pathway effectors to functional groups of specified target genes. This allows users to identify and trace topologies capable of activating or repressing specific groups of genes.

Finally, Epoch is modularly designed such that it can be broken into individual steps and flexibly merged and used with any trajectory inference, network reconstruction, and even network refinement tools that currently exist or that may be developed in the future.

Epoch performs favorably on synthetically generated data

To assess the performance of Epoch in reconstructing static networks, we benchmarked it on synthetically generated data. We generated fifteen synthetic datasets using the Dyngen R package (Cannoodt et al., 2020). Networks were designed to encompass varying regulatory motifs and spanned 100-420 genes including 20-70 transcription factors. Synthetic single-cell datasets generated from these networks varied in size, from as low as 20 to as high as 1000 cells. We compared Epoch to the original versions of GENIE3 (Huynh-Thu et al., 2010), which is the GRN reconstruction engine of SCENIC (Aibar et al., 2017), and CLR (Faith et al., 2007) as well as variations on these methods that incorporated them within the Epoch framework. Thus, we assessed a total of 11 methods: CLR using mutual information (original CLR), CLR using Pearson correlation, CLR using mutual information with cross-weighting, CLR using Pearson correlation with cross-weighting, GENIE3 (original GENIE3), GENIE3 limited to dynamically expressed genes, GENIE3 limited to dynamically expressed genes with cross-weighting, Epoch with Pearson correlation (with and without cross-weighting), and Epoch with mutual information (with and without cross-weighting).

We reconstructed individual networks with each method for each dataset, using normalized simulation time as the pseudotemporal annotation for Epoch. To assess performance, we compared each resulting network to a set of five random networks, unique to each reconstructed network, that were generated by permuting edge weights of targets for each transcription factor in the reconstructed network (**Figure 3a**). As a gold standard, all networks were compared against the original network used to generate the synthetic single-cell data. Based on fold improvement in area under the precision-recall curve over the random networks, all variations of Epoch outperformed original versions of CLR and GENIE3 (**Figure 3b, Supplementary Figure 1b**). Additionally, we found that Epoch using mutual information with or without cross-weighting resulted in higher fold improvement over random compared to any other method. In all cases, it was apparent that key steps in the Epoch framework improved overall static reconstruction. Specifically, limiting reconstruction to dynamically expressed genes and applying the cross-weighting scheme both led to increased fold improvement in AUPR over random, demonstrating the utility of Epoch's framework.

Epoch performs favorably on real world data

While the results from assessment of networks reconstructed from synthetic data demonstrated Epoch's favorable performance as a network inference tool, we acknowledge that there are caveats associated with synthetic data, including limited stochasticity and reliance on simulations that may not encompass the full extent of potential cell state responses seen in real biological systems. Thus, we next turned our attention to benchmarking on real world data. To assess Epoch's performance, we began by assessing the networks reconstructed by the same 11 methods listed above from data of E12.5 mouse muscle development (**Figure 3c**). For this dataset, we used Diffusion based pseudotime (DPT) to compute pseudotime for Epoch reconstruction (Haghverdi et al., 2016). We used Chip-X data from the Enrichr database, which includes ChIP-Seq, ChIP-ChIP, ChIP-PET, DamID used to profile transcription factor binding to DNA, covering 220 TFs and >35,000 genes, as a gold standard network (Lachmann et al., 2010). Based on precision recall, we found that all variations of Epoch once again outperformed original versions of CLR and GENIE3. Interestingly, we found that the two variations of GENIE3 embedded within the Epoch framework performed best overall in this assessment, followed by the two variations of Epoch using mutual information (with and without cross-weighting). It was again apparent that limiting reconstruction to dynamically expressed genes and cross-weighting significantly improved overall reconstruction (**Figure 3d, Supplementary Figure 1c**).

To determine if these results were consistent across other real world datasets, we turned to single-cell transcriptomic data collected from day 4 in vitro mESC directed differentiation (Spangler et al., 2018). We limited our reconstruction to cells in earlier stages of differentiation that had not yet transitioned into any committed germ layer-like state, as determined by pseudotemporal analysis, RNA velocity analysis, expression of Zfp42, marking naive pluripotent cells, and expression of Fgf5, marking cells in a primed state (**Supplementary Figure 1d**). The filtering ensured the trajectory remained linear, which, while absolutely not a requirement for Epoch reconstruction, was employed to minimize performance variability of trajectory inference methods in which branched structures may be more difficult to order. This allowed us to narrow our focus on reconstruction performance while minimizing impact from potentially inaccurate pseudotemporal ordering. Based on precision-recall, we found that Epoch variations using cross-weighting performed best, followed by GENIE3 embedded within the Epoch framework (**Supplementary Figure 1e**). Overall, these benchmarking results demonstrate Epoch's utility as a reconstruction tool, and further demonstrate the usefulness of Epoch as a flexible framework that can improve reconstruction of other methods.

Single cell RNA-sequencing of early in vitro mouse ESC directed differentiation

With the goal of exploring the dynamic GRN topology underlying lineage specification in gastrulation, we collected scRNA-seq data from day 0 through day 4 *in vitro* mESC directed differentiation guided by four separate treatments encouraging primitive streak formation (**Figure 4a**). Briefly, cells were allowed to differentiate in serum-free differentiation media before being treated with one of four treatments on day 2: Wnt3a and Activin A alone (WA), or with one of Bmp4 (WAB), GSK inhibitor (WAG), or Noggin (WAN). Multiple rounds of differentiation were staggered such that we could harvest samples representative of d0-d4 at one time for sequencing using the MULTI-seq protocol (McGinnis et al., 2019). After barcode classification, samples were preprocessed using the SCANPY pipeline (Wolf et al., 2018), and RNA velocity analysis was performed (Bergen et al., 2019; La Manno et al., 2018) (**Figure 4b,c**).

We found the presence of three distinct lineages, which we confirmed to be representative of neuroectoderm (lineage expressing Sox1), mesoderm (lineage expressing Mesp1), and endoderm (lineage expressing Foxa2) based on gene expression and further verified by singleCellNet (SCN) analysis (Tan and Cahan, 2019) (**Figure 4d, Supplementary Figure 2, 3a**). RNA velocity additionally supported our cluster annotations. We found that the majority of the differentiating cells transitioned toward the neuroectoderm (clusters labeled 0,2, 0,3, 4, 2, 5, 9) fate with smaller populations transitioning toward mesoderm (cluster labeled 8) and endoderm fate (clusters labeled 7,1 and 7,0).

Undifferentiated cells populated cluster 1, which was positive for Zfp42 expression, a pluripotency marker (Rogers et al., 1991). This population additionally classified as inner cell mass, further supporting their pluripotent annotation. RNA velocity indicated these cells transitioned into cluster 3, which marked the beginning of Fgf5 expression, an epiblast maker (Haub and Goldfarb, 1991; Pelton et al., 2002). SCN revealed early ectoderm and epiblast signature for cluster 3, indicating to us that it was representative of a primed cell state. Cells in this cluster transitioned to cluster 0 and 0,4. Cells in cluster 0 were largely representative of primitive streak, as determined by Brachyury expression, with some early ectoderm and residual epiblast signature, as determined by SCN. Interestingly, cluster 0,4, which expressed a lower level of Brachyury and exhibited some primitive streak and early ectoderm SCN signature, was comprised of day 0 cells. It is possible that these cells, which had yet to undergo the differentiation protocol, had escaped LIF-controlled naïve pluripotency and were already in a primed early ectodermal state.

Cluster 0 cells expressing Mixl1, a regulator of mesoderm and endoderm (Hart et al., 2002) further branched into two distinct lineages according to RNA velocity: mesoderm (cluster

333 8) and endoderm (clusters 7,1 and 7,0). Cluster 8 was largely enriched for Mesp1, and some
334 cells at later time (determined by RNA velocity) in this cluster also expressed Tbx6. SCN
335 analysis revealed that this cluster exhibited multiple mesoderm-related signatures including
336 primitive streak, anterior paraxial mesoderm, presomitic mesoderm, and somites, ultimately
337 supporting its mesoderm annotation. Along the endoderm path, cluster 7,1 transitioned from
338 cluster 0 and further transitioned into cluster 7,0. Both 7,1 and 7,0 are enriched for Foxa2 and
339 Sox17 expression. SCN analysis suggested an increasing endodermal signature down this
340 path, with cluster 7,1 classifying as anterior primitive streak and weaker as definitive endoderm
341 and gut endoderm, while cluster 7,0 more strongly classified as definitive and gut endoderm.
342 The neuroectodermal path begins from both cluster 0 and 0,4 according to RNA velocity
343 results. We annotated clusters 0,3 and 4 as neuromesodermal progenitor (NMP) populations.
344 SCN further verified cluster 0,3, which expresses T and Sox2, as early NMP and cluster 4,
345 which expresses Sox2 and exhibits variable T expression, as late NMP. Cluster 7,2, which
346 expresses Chordin, classified strongly as node. Interestingly, RNA velocity suggested the
347 existence of two parallel paths along the neuroectodermal lineage, with cluster 0,2 transitioning
348 into cluster 5, and cluster 4 transitioning into cluster 2. Along the first path, cluster 0,2, classified
349 weakly as a number of ectodermal cell types including early ectoderm. Because RNA velocity
350 indicated this cluster transitioned into the later neuroectoderm cluster 5, we annotated it as early
351 ectoderm. Cluster 5 classified strongly as forebrain/midbrain, with some weaker future spinal
352 cord classification. Given this, and it's Zic1 and En2 expression, we annotated this cluster as
353 future brain. Cluster 9, which expresses Crabp1 and differentially expresses Tubb3 (suggesting
354 a later neuronal identity compared to cluster 2 or 5), seems to transition from this cluster, and is
355 thus annotated as future brain. Along the second neuroectodermal path, cells in cluster 2
356 classified more strongly as future spinal cord with a weaker forebrain/midbrain classification.
357 Importantly this cluster expresses many patterning genes from the Hox family indicative of
358 spinal cord fated cells, including Hoxd9 and Hoxb6 (Philippidou and Dasen, 2013). For this
359 reason, we annotated this cluster as future spinal cord.
360 Interestingly, in annotating the data, it is apparent that there are differences between
361 induction treatments in terms of the distribution of cells amongst the different populations
362 (**Figure 4e,f, Supplementary Figure 3b**). Notably, cells treated with WAG exhibited a stunted
363 trajectory toward neuroectodermal fate but a fuller trajectory toward mesodermal fate. This was
364 in direct contrast to cells treated with the remaining three treatments (WAB, WAN, WA), in which
365 very few cells differentiated toward mesodermal fate, but instead exhibited strong
366 neuroectodermal or endodermal commitment. Further, while WAB and WA cells tended to

transition into cluster 7,0 (furthest along endodermal development), WAN and WAG cells were more likely to stay in cluster 7,1.

Epoch elucidates regulatory networks driving lineage specification in early in vitro mouse ESC differentiation

To begin our network analysis exploring lineage specification, we applied Epoch to the day 0 through day 4 directed differentiation data, using latent time (Bergen et al., 2020) to order cells, and reconstructing the networks underlying each of the three lineages. Latent time was divided into three epochs, and a dynamic network was extracted for each lineage (**Figure 5, Supplementary Figure 4**). In examining the Epoch-predicted top regulators and their top targets, we observed noteworthy regulatory programs emerging over time along the three lineages. For example, a module including the genes Gsc, Lhx1, Sox17, Foxa1, Foxa2, and Gata6 emerges in the third epoch along the endoderm path. Sox17 and Foxa2 serve as endoderm markers, as knockout of Sox17 has been shown to deplete definitive gut endoderm and lack of Foxa2 results in failure to develop fore- and midgut endoderm (Dufort et al., 1998; Kanai-Azuma et al., 2002). Foxa1 has also been implicated in the initiation and maintenance of the endodermal lineage (Ang et al., 1993), while inactivation of Lhx1, which has been shown to bind to the enhancer region of Foxa2, disrupts the development of definitive endoderm (Costello et al., 2015). The regulator Sox11 and a collection of Hox genes emerge in the third epoch of the neuroectoderm lineage, suggesting the activation of programs that pattern brain and spinal cord (Philippidou and Dasen, 2013). Ultimately, these networks served as a foundation for the remainder of the analyses in this study, representing TF-target regulation present along each path of lineage specification.

In addition to directly extracting top regulators via PageRank and Betweenness-Degree from the reconstructed lineage networks, we highlight a more nuanced methodology to resolve influential TFs based on bagging, since the consensus of many reconstructed networks would be less susceptible to noise. Specifically, we limited our scope to early time, corresponding to a naïve to primed pluripotency transition, and bootstrapped the network reconstruction (**Supplementary Figure 5**). For each of these sampled and reconstructed networks, we predicted the top 10 regulators involved in the transition by both PageRank and Betweenness-Degree. We then applied a consensus approach in which regulators were ranked by their frequency as a top 10 regulator (based on PageRank) or average importance (based on Betweenness-Degree) amongst the bootstrapped networks. The top regulators recovered by

401 both PageRank and Betweenness-Degree were largely overlapping, and included TFs such as
402 Zfp42, Klf2, Sox11, Pou3f1, and Klf4.
403 Of the top 10 regulators predicted through this consensus approach, Klf2 and Klf4 are
404 critical in maintaining ground state pluripotency (Yeo et al., 2014). For example, it has been
405 demonstrated that nuclear export of Klf4 causes rapid decline in Klf4 transcription, a trigger that
406 leads to the exit of naïve pluripotency, and the blocking of which prevents ESC differentiation
407 (Dhaliwal et al., 2018). Thus, it is unsurprising that they are predicted to have an influential role
408 in dictating network topology. Interestingly, a number of predicted top regulators, such as
409 Pou3f1 and Sox11, are known to promote neural fate (Barral et al., 2019; Wang et al., 2013;
410 Zhu et al., 2014). Knockdown of another predicted top regulator, Tcea3, has also been shown to
411 bias differentiation toward mesendodermal fates (Park et al., 2013). Taken together, this
412 suggests that the underlying network is primed for the neuroectodermal specification fate
413 choice, aligning with the default model of neural induction in mESCs (Muñoz-Sanjuán and
414 Brivanlou, 2002; Tropepe et al., 2001) and the apparent preference for neuroectoderm fate
415 amongst cells in this dataset.
416
417 *Peg3 is a central regulator in mesodermal WAG networks and may drive differences in gene*
418 *module expression amongst treatments*
419
420 We next focused our attention on mesodermal fate specification in mESC directed
421 differentiation. Specifically, in observing both the scarcity of the mesodermal population (in
422 comparison to neuroectodermal fate) as well as the discrepancy amongst treatments in
423 propensity for reaching mesodermal fate, we asked why WAG treated cells had the greatest
424 potential for reaching this population. We hypothesized that underlying these differences in fate
425 potential amongst treatments were treatment-specific GRNs. To this end, we reconstructed and
426 compared dynamic networks for each treatment along the mesodermal path (which we refer to
427 as the treatment-specific networks: WAG, WAB, WAN, and WA networks), and further
428 compared these against the full reconstructed mesoderm network described in the previous
429 section (which is reconstructed from all cells regardless of treatment, and which we refer to as
430 the mesoderm network). We then aimed to understand why WAG treated cells were able to
431 specify mesodermal fate to a greater extent than WAB, WAN, and WA treated cells.
432 We then sought to understand what regulators could be driving differences in mesoderm
433 reachability between the treatments. We expected treatment-driven differences in fate potential
434 to logically arise after cells undergo treatment on day 2, and would be most apparent toward

435 later latent time. We also reasoned that we could quantify and monitor differences in cell state
436 based on the activity of gene modules, defined as a community or unit of nodes that are highly
437 interconnected, within the network. Specifically, understanding the activity of these modules in
438 each treatment group would elucidate the extent to which the mesoderm state is established
439 and further aid in identifying regulators that fail to activate these regulatory programs. Thus, we
440 performed community detection on epoch 2, epoch 3, and their transition, of the mesoderm
441 network and identified distinct TF communities, or modules, within the topology (**Figure 6a,**
442 **Supplementary Figure 7**). We assessed the activity of each module by looking at the average
443 expression of member genes (excluding genes predicted to be repressed in a given module)
444 across latent time for each treatment (**Figure 6b, Supplementary Figure 8a**). Importantly, we
445 found the existence of certain modules that were strongly activated in the WAG treatment that
446 failed to activate or were weakly activated in the remaining three treatments. Three such
447 modules (one activated in each of the second epoch, third epoch, and transition) were
448 particularly interesting as they exhibited strong activity in the WAG treatment along not only the
449 mesoderm path, but also the full dataset. The TFs in these modules were Peg3, Lhx1, Hoxb2,
450 Foxc1, Tshz1, Meis2, Mesp1, Tbx6, Foxc2, Prrx2, Meox1, and Notch1. Upon extracting the
451 differential network (which we define as the network containing the interactions, or edges, that
452 have large weight differences between two treatments, and thus represent interactions most
453 likely to be unique to a treatment network; see methods) between WAG and WA, we recovered
454 the same modules and further identified the top differential regulators of these TFs themselves
455 (**Figure 6c, Supplementary Figure 8c**). Interestingly, in contrast to the WAG treatment, Peg3
456 was predicted to have a negligible role in the WA treatment.
457 Peg3's central influence in the WAG network, however, implied a pathway linking GSK3
458 inhibition, Peg3 expression, and mesodermal gene expression. Indeed, previous reports have
459 demonstrated a role for GSK3 in maintaining DNA methylation at imprinted loci, including Peg3,
460 in mESCs (Meredith et al., 2015). Additional literature has demonstrated that methylation at
461 imprinted loci may be maintained by Dnmt3a2, which has been suggested to be regulated by
462 the GSK3-dependent transcription factor N-Myc (Popkie et al., 2010). Unfortunately, our data
463 was only able to capture Dnmt3a dynamics (representing composite Dnmt3a and Dnmt3a2
464 expression), and not specifically Dnmt3a2, which is transcribed from an alternative promoter
465 within the Dnmt3a locus. However, since Dnmt3a2 is the predominant isoform in ESCs (Chen et
466 al., 2002), we roughly compared Dnmt3a temporal dynamics and Peg3 expression between the
467 different treatments (**Supplementary Figure 8b**). Notably, Dnmt3a and Peg3 expression were
468 largely mutually exclusive, and there was low-to-no Dnmt3a expression in the WAG treated

469 cells. We then asked if this phenomenon occurs in *in vivo* gastrulation. In the single-cell mouse
470 gastrulation dataset from Grosswendt et al. (Grosswendt et al., 2020), we found relatively high
471 expression of Dnmt3a in epiblast cells which was tapered in primitive streak and mesodermal
472 cells (**Figure 6d**). In contrast, Peg3 expression was low in epiblast and increased in the
473 primitive streak, mirroring our *in vitro* data, and suggesting that Peg3 may have a role in
474 orchestrating the exit of pluripotency and specification of mesodermal fate *in vivo*.
475
476 *Tracing signaling cascades to germ layer transcriptional programs*
477
478 We next sought to trace pathways from signaling effectors to the activation of
479 mesodermal fate. To determine which signal transduction pathways were active at any given
480 point in latent time along the mesodermal lineage, we used ChIP-seq data from ChIP-Atlas (Oki
481 et al., 2018) to compile a list of binding targets for 18 signaling effector TFs (available as part of
482 the Epoch framework). We then computed average expression of these targets across latent
483 time broken down by treatment, and unsurprisingly found differences in the signaling pathways
484 that were activated in each treatment (**Figure 7a**). Of note, Notch signaling was activated in
485 WAB, WAN, and WA treatments, but not in the WAG treatment along the mesodermal lineage
486 (and weakly activated along the neuroectodermal lineage), consistent with literature
487 demonstrating that Notch signaling plays a role in specifying neuroectodermal fate in the
488 neuroectoderm-mesendoderm fate decision (Androutsellis-Theotokis et al., 2006; Aubert et al.,
489 2002).
490 Because Wnt signaling is essential for mesodermal fate, we looked for paths connecting
491 Wnt effectors to the TFs in the WAG-specific modules defined previously. Specifically, we
492 computed the shortest paths from targets of these effectors to the module TFs in each
493 treatment-specific dynamic network (using edge lengths inversely proportional to the cross-
494 weighted score) (**Figure 7b**). Paths were assigned infinite length if such a path did not exist. We
495 normalized these path lengths against the average path length in a given network, and found
496 that for the WAB, WAN, and WA treatments, no paths from Wnt to many of the module TFs
497 existed.
498 Our signaling effector target analysis also revealed increased Foxo1 activity in WAB,
499 WAN, and WA treated cells as compared to WAG treated cells, indicating the suppression of
500 PI3K signaling along the WAB, WAN, and WA trajectories (**Figure 7a**). Importantly, it is known
501 that the establishment of definitive endoderm (DE) requires suppression of PI3K signaling
502 (McLean et al., 2007; Yu et al., 2015). Further, it has been recently shown in hiPSC to DE

503 differentiation that Foxo1 binds to DE-formation-related genes and that inhibition of Foxo1
504 during DE induction impedes DE establishment (Nord et al., 2020). We therefore hypothesized
505 that a second fate choice between mesodermal and endodermal fates further exacerbated the
506 uneven mesodermal fate preference between the treatments. Indeed, our analysis indicates that
507 in the mesodermal-endodermal fate choice, the majority of WAG treated cells differentiate
508 toward mesodermal fate. This is in contrast to the WAN and WA treatment in which cells
509 preferentially differentiate toward definitive endoderm. WAB treated cells exhibited mixed
510 potential, and split roughly equally along the two paths.
511 We sought to elucidate a possible explanation for Foxo1's role in this fate choice by
512 searching for the shortest paths toward a set of endodermal genes. Specifically, similar to the
513 previous mesodermal analysis, we applied community detection to the full endodermal network
514 and identified the cluster containing Sox17 and Foxa2, known master regulators of endodermal
515 fate (**Figure 7c**). This cluster included the transcription factors Sox17, Foxa2, Bmp2, Cited1,
516 Foxa1, Gata6, Hhex, Lhx1, and Tcf7l2. For each treatment-specific network, we computed the
517 shortest paths from Foxo1 targets (as determined in the same manner as the previous Wnt
518 analysis) to these genes (**Figure 7d**). Importantly, we found that many of these regulators,
519 including both Sox17 and Foxa2, were not reachable from Foxo1 in the WAG network,
520 consistent with the observation that WAG treated cells preferentially differentiated toward
521 mesodermal fate over endodermal fate. Interestingly, Foxa2 was not reachable in the WAN
522 network, though Sox17 was. We believe this may explain the differences in distribution of cells
523 between the endodermal clusters 7,1 and 7,0. Namely, of the cells that specify endodermal fate,
524 WA and WAB treated cells more fully transitioned into cluster 7,0. In contrast, WAN treated cells
525 remained mostly in cluster 7,1 and were less likely to differentiate further into 7,0. Previous
526 studies in directed differentiation of human ESCs toward DE have implicated Sox17's regulatory
527 role in establishing DE to be upstream of Foxa2, the loss of which impairs foregut and
528 subsequent hepatic endoderm differentiation (Genga et al., 2019). This is consistent with our
529 cell type annotations of cluster 7,1 (definitive endoderm) and 7,0 (gut endoderm), and offers a
530 possible mechanism of the discrepancy in cell fate preferences between treatments. Taken
531 together, these results demonstrate that WAB, WAN, and WA networks do not allow for
532 activation of mesoderm programs from their cognate effectors, but instead assume topologies
533 that are preferential for endodermal fate, implying profound structural changes in the networks
534 that underlie the treatment-specific differences we see in cell fate. Further, this analysis
535 provides a basis for identifying which signaling pathways must be targeted for directing certain
536 fate transitions, for example through an exhaustive search of effector targets.

537

538 *Target gene differences amongst TFs confirm network topology restructuring dependent on*
539 *treatment*

540

541 Given the apparent restructuring of GRN topology upon modulation of signaling activity,
542 we aimed to understand the extent to which broad, network-wide topology changes were
543 responsible for differences in cell fate potential between treatments. To answer this, we
544 performed two distinct analyses focusing on differences in targets of transcription factors across
545 the four treatment networks.

546 First, we asked if transcription factors actively expressed in all four treatments regulated
547 the same set of target genes. To this end, we focused on the treatment-specific networks along
548 the mesodermal path. We narrowed our analysis to the 72 TFs that were active in all four
549 treatments during day 3 or day 4 along the mesodermal path. We quantified overlap between
550 targets of each TF in a pairwise comparison of the treatments using the Jaccard similarity
551 (**Figure 8a**). As a baseline, we implemented a bootstrapping method in which we reconstructed
552 10 networks for each treatment along the mesodermal path (using 400 sampled cells each); for
553 each TF we calculated the average Jaccard similarity of its targets amongst pairwise
554 comparisons of the reconstructed networks in a treatment. This gave us a baseline, or
555 expected, Jaccard similarity for each TF in each treatment. Overall, we found that
556 overwhelmingly, target differences between treatments were significantly greater than the
557 baselines, indicating vast network topology differences amongst the treatment networks.

558 Next, we asked if these differences in predicted targets had functional consequences in
559 impacting fate potential. To this end, we focused on the 13 TFs amongst the 72 mentioned
560 above that exhibited the largest differences in predicted targets between treatments (**Figure**
561 **8b**). For each treatment, we performed gene set enrichment analysis on the predicted targets of
562 these TFs (**Figure 8c**). Interestingly, we found enrichment for heart development and
563 mesoderm formation in only the WAG treatment, despite the fact that the networks we analyzed
564 were reconstructed along the mesodermal path. Of note, all treatments showed enrichment for
565 embryonic digestive tract morphogenesis. Importantly, prior literature has indeed hinted at
566 environment-based differences in fate potential. For example, in uncovering the regulatory
567 structure underpinning naïve pluripotency in mESCs, the effect of Esrrb/Nanog double
568 knockdown was shown to be dependent solely on culture environment with self-renewal
569 impeded in the presence of 2i alone, but not in the presence of both 2i and LIF (Dunn et al.,

570 2014). Taken together, these results support our hypothesis that manipulating signaling activity
571 results in a topological restructuring of the GRN, ultimately guiding cell fate potential.
572
573
574 **Discussion**
575 Methods of inferring regulatory network structures are essential for understanding cell
576 identity transitions, uncovering mechanisms of disease progression due to dysfunctional gene
577 regulation, and answering other regulatory questions in biology. Unfortunately, experimental
578 methods are unscalable and existing computational methods suffer from poor performance or
579 high computational burden. Here, we present a novel GRN reconstruction tool, Epoch, which
580 leverages single-cell transcriptomics and efficiently infers dynamic network structures. We
581 demonstrate that Epoch outperforms commonly used methods CLR and GENIE3 in both
582 synthetically generated and real-world datasets. Furthermore, Epoch's relative efficiency allows
583 it to be easily applied iteratively for applications requiring fine-tuning of network topology or
584 iterative simulations of changes in network structure.
585 Epoch's utility is additionally enhanced by its flexibility. There are no strict requirements
586 on pseudotemporal input as we demonstrate here using synthetic simulation time, diffusion
587 pseudotime, and RNA velocity-based latent time. We further validate on the *in vivo* muscle
588 development dataset that precision-recall is largely comparable for different pseudotime
589 methods (**Supplementary Figure 1f,g**). Additionally, Epoch's workflow is structured in discrete
590 steps allowing users to pick-and-choose as well as substitute portions of the workflow. We
591 anticipate that this flexibility will also be useful as advances in sequencing technology, such as
592 multi-modal methods, and analysis techniques, such as novel cell ordering or reconstruction
593 methods, can be easily adapted into Epoch's framework. Importantly, because of single-cell
594 transcriptomic data's inherent noisiness, we also asked if smoothing gene expression over time
595 could alleviate noise-related drops in reconstruction performance. We compared performance
596 between unsmoothed and kernel-smoothed data and found that the there was no significant
597 difference (**Supplementary Figure 1h**). However, smoothing functions remain in Epoch's code
598 to aid in cleaner visualizations of gene dynamics.
599 Finally, we applied Epoch to scRNA-seq data from day 0 through day 4 *in vitro* mESC
600 directed differentiation to uncover underlying dynamic networks driving mesodermal lineage
601 specification. Our data consisted of cells transitioning into all three germ layers, and Epoch was
602 able to recover a number of key lineage drivers. Because our data consists of cells undergoing
603 four different treatments (WAG, WAB, WAN, WA), we were additionally able to reconstruct

604 treatment-specific GRNs. Our analysis uncovered significant treatment-induced topological
605 restructuring that altered the lineage specification landscape, resulting in treatment-based
606 differences in neuroectoderm, mesoderm, and endoderm reachability. Specifically, in direct
607 contrast to the cells in the WAB, WAN, and WA treatment groups, the cells in the WAG
608 treatment group had higher mesodermal fate potential, exhibited a relatively stunted
609 neuroectodermal trajectory, and had decreased potential for endodermal fate. Importantly, we
610 uncovered three sets of topological differences that may play a role in specifying mesodermal
611 fate.
612 First, we identified a set of TF modules that were preferentially activated in WAG treated
613 cells compared to cells undergoing the other three treatments along the mesodermal fate.
614 Specifically, these modules encompassed the transcription factors Peg3, Lhx1, Hoxb2, Foxc1,
615 Tshz1, Meis2, Mesp1, Tbx6, Foxc2, Prrx2, Meox1, and Notch1. Upon further analysis, we
616 observed that the WAG network developed a topology that promoted Peg3 activity, which was
617 predicted by Epoch to have a role in regulating a large fraction of these mesodermal TFs. This
618 influence was again specific to WAG treated-cells, and indeed, prior studies have implicated a
619 GSK3-dependent pathway for the epigenetic regulation of imprinted loci including Peg3
620 (Meredith et al., 2015). Peg3 expression has been detected in early somites and other
621 mesodermal tissues (Kuroiwa et al., 1996). Interestingly, it has been separately demonstrated
622 that knockdown of Peg3 in mouse embryonic fibroblasts and neural stem cells increases
623 expression of pluripotency genes and enhances overall reprogramming efficiency (Theka et al.,
624 2017). Taken together these results suggest a role for Peg3 in modulating the exit of
625 pluripotency and mesodermal lineage specification. If so, this finding could paint Peg3 as a
626 highly enticing transcription factor to target with the goal of improving efficiency of directed
627 differentiation protocols toward certain mesodermal fates such as somites.
628 Next, we compared topological differences in the treatment-specific networks by
629 examining the existence of, and computing the length of, shortest paths. Specifically, we sought
630 to trace paths from signaling effectors to the activation of mesodermal genes within these
631 dynamic networks, thereby providing possible mechanisms of how signal transduction pathways
632 not only directly regulate targets corresponding to distinct genetic programs, but also restructure
633 network topology to alter accessibility to targets, which hones the capacity and potential for
634 distinct fates. A number of studies have pointed to the influence of “Wntch” signaling on the
635 neuroectoderm-mesendoderm cell fate decision in gastrulation (Hayward et al., 2008). Our data
636 was consistent with these reports: WAB, WAN, WA treated cells exhibited activated Notch
637 pathways and exhibited full trajectories toward neuroectoderm fate; WAG treated cells exhibited

638 markedly diminished or non-existent Notch signaling and reached mesodermal fates while
639 seemingly displaying a stunted capacity for neuroectodermal fate. Beyond this, we
640 demonstrated that WAG treated cells established a GRN topology that allowed for Wnt-
641 dependent mesoderm specification. Specifically, our analysis indicated the existence of paths
642 from Wnt signaling effectors to the aforementioned mesodermal TFs within the WAG network.
643 Analogous paths were markedly sparser in WAB, WAN, and WA networks, and some
644 mesodermal TFs were entirely unreachable from Wnt effectors. This further supported our
645 hypothesis that signaling-induced topological differences altered the cell fate landscape,
646 resulting in distinct propensities for specific fates, though the nature of the emergence of these
647 topological differences remains unclear. One possibility is that the network structures are
648 dynamically altered by epigenetic events initiated through the various signaling pathways
649 activated in treating the cells. We point to GSK3's reported effect over Peg3 methylation as an
650 example of such a phenomenon. Further studies to elucidate differences in chromatin
651 accessibility between mESCs undergoing directed differentiation by way of distinct treatments,
652 or activation of distinct signaling pathways, will aid in clarifying if such a mechanism is at play.
653 Finally, we focused our analysis on cells specifying mesendoderm fate, and outlined
654 topological differences that could lead to differences in mesoderm vs. endoderm fate potential in
655 cells undergoing different treatments. We applied an analogous shortest path analysis along the
656 endodermal fated path to identify the link between suppression of PI3K signaling, subsequent
657 Foxo1 activation, and endoderm specification. We again found topological differences between
658 WAG and the remaining three treatments, with no paths from Foxo1 to essential endodermal
659 regulators such as Sox17 and Foxa2 in the WAG network, explaining the preference of these
660 cells to transition into mesodermal rather than endodermal fates. Interestingly, we found that
661 Foxa2 was not reachable in the WAN network, which was consistent with our finding that of the
662 WAN treated cells that specified endodermal fate, the majority tended to remain in earlier
663 endodermal stages based on RNA velocity. This was contrast to WAB and WA treated cells,
664 which tended to transition more fully into a later endodermal stage, thereby explaining the
665 observed differences in mesodermal vs. endodermal fate potential between the treatments.
666 Our results ultimately suggest a model of differentiation that is driven by activation of
667 genetic programs, but honed by network topology changes. In other words, it is likely that the
668 signaling-induced restructuring of the GRNs alters the fate potential landscape which allows for
669 independent control of multiple genetic programs, increasing the diversity of reachable cell
670 states. In the case of our mESC directed differentiation, GRN topology clearly restricted the
671 potential of cells toward specific fates, resulting in different propensities for mesoderm,

endoderm, and neuroectoderm depending on treatment. Finally, while it may be the case that both signaling activities described (i.e. the effector regulation of distinct targets and the restructuring of network topology) occur concurrently, it is ultimately difficult to resolve their relative timing.

We believe the single-cell directed differentiation dataset collected in this study will be a useful resource for further analysis of decision-making in early gastrulation. Moreover, it will be illuminating to compare the networks reconstructed from this dataset against networks reconstructed from *in vivo* gastrulation. Furthermore, the methodologies presented here, available in the R package Epoch, are broadly applicable to any biological question that may benefit from uncovering dynamic regulatory network structures and the comparison of such networks. In particular, we can apply Epoch to understand cell fate transitions and decisions at branch points in lineage trajectories, uncover key regulators driving these decisions, and identify paths from the signal transduction level to the transcriptional regulation level. Such an approach provides a powerful strategy to not only elucidate dynamic, multi-scale processes in development, but also to identify signaling pathways to dysregulate for the purposes of directing cell identity transitions *in vitro*.

Acknowledgements

We would like to thank Yuqi Tan, Ray Cheng, and David Johanson for helpful discussions and testing Epoch code. We also thank Chris McGinnis and the Gartner Lab for their help implementing the MULTI-seq protocol and for providing the required reagents. This work was supported by the National Institutes of Health under grant R35GM124725 to PC and by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1746891 to EYS.

Figure legends

Figure 1. Effector activity on static and dynamic networks. Given a static GRN, manipulating signaling activity may result in an effector guiding cell state via two possibilities: directly, the effector may regulate a set of fate-specific genes, or indirectly, the effector may regulate a set of TFs that further regulate the same fate-specific genes. Additionally, given a dynamic network,

706 the state space of possible cell fates increases as the result of effector activity is dependent on
707 a changing network topology, and multiple sets of genetic programs may be regulated through
708 the same signaling mechanism. Further, it poses a third possibility of signaling-guided cell state
709 transitions, in that some GRN topology changes may be induced by the signaling activity itself.
710
711 **Figure 2.** Epoch workflow. Epoch relies on a three-step process to reconstruct dynamic network
712 structures: (1) extraction of dynamic genes and CLR, (2) application of cross-weighting, and (3)
713 extraction of the dynamic network.
714
715 **Figure 3.** Benchmarking Epoch. (A) Synthetic dataset generation, reconstruction, and
716 assessment. (B) Fold improvement in AUPR over random of three versions of Epoch (Pearson,
717 MI, and MI with cross-weighting) against CLR and GENIE3 on GRN reconstruction from
718 synthetic data. (C) Clustering and diffusion pseudotime for E12.5 muscle development data
719 used in benchmarking. (D) Precision-recall curves for three versions of Epoch (Pearson, MI, and
720 MI with cross-weighting), CLR, and GENIE3 based on reconstruction from muscle development
721 dataset. (E) Runtimes for reconstruction. (Runtimes were measured on a MacBook Pro laptop
722 with 2.9 GHz Intel Core i7 processor, 16GB 2133 MHz DDR3 of RAM).
723
724 **Figure 4.** mESC in vitro directed differentiation. (A) Directed differentiation protocol. Cells were
725 harvested from d0 through d4 samples for MULTI-seq. (B) Clustering for MULTI-seq data. (C)
726 RNA velocity for MULTI-seq data. (D) Select marker gene expression. (E) Cell populations by
727 treatment. (F) Quantification of cell fate distribution based on treatment comparing:
728 mesendoderm vs. neuroectoderm, endoderm vs. mesoderm, cluster 7,0 vs. 7,1 along endoderm
729 fate, and cluster distribution along neuroectoderm fate.
730
731 **Figure 5.** MULTI-seq directed differentiation network reconstruction. (A) Latent time annotation
732 for MULTI-seq data. (B) population breakdown by cluster of epoch assignment along each germ
733 layer path. (C) Reconstructed dynamic networks along mesoderm, endoderm, and
734 neuroectoderm paths, with top 5 regulators colored purple and their top targets in gray as
735 determined by PageRank. (D) Top regulators as predicted by Epoch via betweenness and
736 degree for the third epochs along each germ layer path.
737
738 **Figure 6.** Mesodermal network analysis. (A) The epoch 3 subnetwork of the mesodermal
739 dynamic network. TFs are colored by community and faded by betweenness. Blue and red

740 edges represent activating and repressive edges respectively. (B) Average community
741 expression over time by treatment along the mesodermal path. Communities shown (each row)
742 are from the epoch 2 subnetwork (yellow), transition (green), and epoch 3 subnetworks (aqua).
743 (C) The epoch 3 differential network between WAG (on) and WA (off). Interactions in this
744 network represent edges present in the WAG mesodermal network that are not present in the
745 WA mesodermal network. TFs are colored by community and faded by betweenness. Blue and
746 red edges represent activating and repressive edges respectively. (D) Peg3 and Dnmt3a
747 expression in a sampled portion of relevant cell types in gastrulation data from Grosswendt et
748 al.
749
750 **Figure 7.** Integrating signaling effectors and tracing paths to target genes. (A) Average
751 expression of targets of signaling effectors by treatment over time. Pathway and effector are
752 labeled on each row. (B) Shortest path analysis showing length of shortest paths from Wnt
753 effector targets (each row) to mesodermal target genes (each column) within treatment-specific
754 networks. Path lengths are normalized against average path length within a network. Blue and
755 red indicate paths that activate and repress the target gene respectively. If such a path does not
756 exist, it has length of infinity and is therefore white. (C) The epoch 3 subnetwork of the
757 endodermal dynamic network. TFs are colored by community and faded by betweenness. Blue
758 and red edges represent activating and repressive edges respectively. (D) Shortest path
759 analysis showing length of shortest paths from PI3K suppression (targets of Foxo1 activation)
760 (each row) to endodermal genes (each column) within treatment-specific networks. As before,
761 path lengths are normalized against average path length within a network. Blue and red indicate
762 activation and repression of the target gene respectively. If a path does not exist, it has length of
763 infinity and is therefore white.
764
765 **Figure 8.** Similarity of targets of TFs in different treatments. (A) Jaccard similarity of targets of
766 72 TFs. The targets of each TF in a treatment are compared to those in the other three
767 treatments (light blue). As a baseline (turquoise), targets of each TF in bootstrapped
768 reconstructed networks within a treatment are pairwise compared. (B) The 13 TFs selected for
769 GSEA and their average pairwise Jaccard similarity amongst the treatment-specific networks.
770 (C) Summarized results of performing GSEA on targets of TFs in B. Heatmap shows rankings
771 (top 10 shown) based on frequency a term is considered enriched amongst the 13 TFs (1 =
772 most frequent enriched term).
773

774

775 **Methods**

776

777 **mESC maintenance and differentiation**

778

779 GFP-Brychyury mESC cells (Gadue et al., 2006) were maintained and differentiated through
780 day 2 of the directed differentiation protocol according to Spangler et al. (Spangler et al., 2018).
781 At day 2, four different primitive-streak induction treatments were established by adding growth
782 factors Wnt3a (25ng/mL) and Activin A (WA; 9ng/mL) along with one of the following additional
783 growth factors: Noggin (150ng/mL), Gsk inhibitor (10mM), or BMP4 (0.5ng/mL). On the day of
784 sequencing, EBs and monolayers were dissociated to a single cell suspension through the use
785 of TrypLE and 40uM cell strainers.

786

787 **Multi-Seq protocol**

788

789 We utilized the MULTI-seq sample barcoding and library preparation protocol from the McGinnis
790 lab (McGinnis et al., 2019) in order to sequence 27 samples in two 10x capture runs. In short,
791 the protocol involved tagging cell membranes with sample specific barcodes using a lipid-
792 modified oligonucleotide (LMO). The LMOs (reagents obtained from the McGinnis lab) anchor
793 into the cell membrane and allow for the attachment of a normal ssDNA oligonucleotide (MULTI-
794 Seq barcode). A unique MULTI-seq barcode was added to each sample after which, all samples
795 were pooled and sequenced as if they were one sample. An equal number of cells from each
796 sample were combined to make up the pooled samples and ensure equal representation of
797 each sample after down-stream sequencing. MULTI-Seq barcodes were used down-stream to
798 identify which cells were from which samples.

799

800 **Library preparation and sequencing**

801

802 After cells were successfully tagged and pooled, they were submitted to the sequencing core
803 facility for 10x capture and library preparation. A Truseq library preparation was performed with
804 the necessary adjustments made to accommodate the MULTI-Seq platform. Libraries were
805 sequenced on Illumina NovaSeq.

806

807 **Single-cell data processing**

808

809 Sequencing alignment to the mm10 reference genome was performed using CellRanger
810 (version 3.1), and spliced, unspliced, and ambiguous read counts were extracted using
811 Velocyto. Sample barcode classification was performed using the deMULTIplex R package
812 (https://github.com/chris-mcginnis-ucsf/MULTI-seq). Subsequent quality control filtering,
813 normalization, clustering, and differential gene expression analysis was performed using
814 Scanpy (version 1.4.5) (Wolf et al., 2018). Specifically, doublets and negatives from the sample
815 barcode classification were removed, along with 271 cells we identified as residual fibroblasts
816 (Thy1+). Next, genes were excluded if they were detected in less than 5 cells; cells were
817 excluded if their mitochondrial gene content exceeded 10% of their total reads or if they had
818 fewer than 500 unique genes. This left us with 5530 cells and 17528 genes. The data was then
819 normalized and log transformed before highly variable genes were identified (1938 genes). The
820 data was scaled and PCA was performed. Leiden clustering was performed and visualized on a
821 UMAP embedding. We further removed a cluster of possible dying cells expressing relatively
822 high mitochondrial gene content and low total read count, leaving a dataset with 5228 cells and
823 1938 variable genes. Leiden clustering was refined, and RNA velocity was computed using
824 scvelo (version 0.2.2) (Bergen et al., 2020). Latent time was computed with scvelo, using root
825 and end states that were computed through CellRank (version 1.0.0-rc.0). Cell type
826 classification was performed via SingleCellNet (Tan and Cahan, 2019), using complied embryo
827 data for training the classifier. For the E12.5 muscle development data and day 4 early directed
828 differentiation data, diffusion maps and diffusion pseudotime were computed through Scanpy.
829

830 **Synthetic benchmarking**

831

832 15 synthetic datasets were generated using the Dyngen package in R (Cannoodt et al., 2020).
833 Specifically networks were generated with 100-420 genes (including 20-70 TFs), containing
834 varied topological motifs, and cell trajectories were simulated over time. Simulated single-cell
835 sequencing was performed, resulting in datasets that included up to 1000 cells. CLR was
836 implemented with the minet package in R (Meyer et al., 2008), and GENIE3 was implemented
837 with the GENIE3 package in R (Aibar et al., 2017; Huynh-Thu et al., 2010). Variations of these
838 two methods were carried out using the Epoch framework. Performance (AUPR) was compared
839 against sets of 5 random networks generated for each reconstructed network via the random
840 permutation of edge weights across targets.
841

842 **Bootstrapped network reconstruction of early MULTI-seq data**

843

844 We isolated early latent time clusters 1 and 3 and uniformly sampled 400 cells, 10 times to
845 reconstruct 10 networks. For each reconstructed network, we extracted top regulators by
846 PageRank and identified TFs that were most frequently ranked in the top 10 regulators. We
847 further computed the average normalized betweenness and average normalized degree across
848 the 10 reconstructed networks to identify top regulators.

849

850 **Differential network analysis and assessment**

851

852 As one method of network comparison, we used a naïve edge weight-based approach to extract
853 the most differentially weighted edges between networks. Such an approach resulted in
854 "differential networks" which are representative of unique topological features existing in each
855 treatment-specific network. To assess the feasibility of this simple method in extracting real and
856 important topological differences between networks, we first applied the same method to
857 compare the three lineage networks in the dataset. The ground truth on which we assessed
858 these differential networks was a set of lineage-unique "pseudo-ChIP-seq" networks we
859 compiled similar to TF-promoter binding predictions generated by (Lu and Mar, 2020)
860 (**Supplementary Figure 6a**). In short, TF-target interactions were acquired by mapping the
861 PWM of a TF (Weirauch et al., 2014) to the promoter region sequence of potential target genes
862 using the Find Individual Motif Occurrences (FIMO) software (described in more detail below).
863 These ground truth networks were filtered for dynamically expressed genes along given paths
864 (for lineage-specificity), and further filtered for unique edges to mimic extraction of unique
865 topological features (lineage-unique). Using this set of networks, we compared AUPR of our
866 path-differential networks against random networks (generated by shuffling edge weights of
867 targets of each TF in the differential networks). On average the differential networks had a 1.7
868 fold improvement over random AUPR, with many comparisons falling in the 1.8-3.0 range
869 (**Supplementary Figure 6b**).

870

871 **Construction of path-unique TF-target pseudo-gold standard**

872

873 Transcription star site (TSS) annotations were downloaded from the USCS Genome Browser
874 (mouse GRCm38/mm10 genome assembly) (Haeussler et al., 2019). Promoter regions were
875 defined as -750 to +250 around the TSSs. We then acquired position weight matrices (PWMs)

876 of transcription factors from the Catalog of Inferred Sequence Binding Preferences (CIS-BP
877 Database build version 2.00) (Weirauch et al., 2014). We used the Find Individual Motif
878 Occurrences tool (Grant et al., 2011) to obtain potential TF-promoter binding pairs for 648 TFs.
879 We kept all interactions for which p-value < 5 $x10^{-5}$. To generate path-unique gold standards, for
880 each lineage (mesoderm, endoderm, neuroectoderm) we filtered interactions for dynamically
881 expressed genes along each path as computed by Epoch. Further, when assessing the
882 "differential network" pairwise comparisons, we filtered out interactions that were redundant
883 between the gold standards, and only kept interactions unique to each path.
884
885 **Targets of TFs by treatment comparison**
886
887 To quantify the extent of topology differences between treatments along the mesodermal path,
888 we performed pairwise comparison on the four treatment-specific networks and computed
889 Jaccard similarity between predicted targets of TFs. The comparison was carried out on 72 TFs
890 that were selected based on their active expression on day 3 or day 4 (i.e. after induction
891 treatments). We computed baseline similarities for each TF in each treatment by bootstrapping
892 the network reconstruction (10 network reconstructions per treatment, 400 sampled cells per
893 network reconstruction) and averaging their pairwise Jaccard similarities. We selected the 13
894 TFs exhibiting the greatest differences in targets amongst the treatments and performed gene
895 set enrichment analysis using the Enrichr R package (Chen et al., 2013; Kuleshov et al., 2016)
896 and the GO Consortium Biological Processes database (Ashburner et al., 2000; Gene Ontology
897 Consortium, 2021) on the targets of each TF. We further filtered out terms that did not meet the
898 criteria of adjusted p-value<0.05. Finally, we summarized the results between the 13 TFs by
899 counting the frequency (out of 13) that a term was considered enriched and ranked the results
900 accordingly.
901
902 **Comprehensive signaling effector TF targets list**
903
904 We acquired binding score (MACS2) data for 18 signaling effector TFs from the ChIP-Atlas (Oki
905 et al., 2018), and processed each accordingly: target genes were ranked by maximum binding
906 score and the top 2000 targets were retained (or all retained, if less than 2000 targets). To
907 compute average effector activity, target lists were filtered for those appearing in the MULTI-seq
908 directed differentiation data, and not considered in the analysis if the effector had less than 10
909 targets after filtering.

910

911 **Epoch Workflow**

912

913 Epoch takes as input processed (normalized and log-transformed) single-cell transcriptomic
914 data and accompanying pseudotime (or equivalent) annotation. The Epoch workflow is based in
915 three strategies: (1) the extraction of dynamically expressed genes and subsequent static
916 reconstruction with CLR, (2) network refinement using a cross-correlation based strategy
917 termed "cross-weighting," (3) the extraction of a dynamic network and subsequent identification
918 of top regulators. Epoch is available as a package in R, and code and tutorials can be found at
919 https://github.com/pcahan1/epoch.

920

921 **Identification of dynamically expressed genes**

922

923 Limiting network reconstruction to dynamically expressed genes serves two purposes.
924 Importantly, it focuses the network on interactions that are more likely playing a role in the
925 observed biological process. Second, it reduces the instances of false positive interactions and
926 improves precision by limiting possible edges between temporally-variable genes. To select for
927 such genes, Epoch models individual genes across the annotated pseudotime using a
928 generalized additive model. Specifically, Epoch uses the 'gam' package (gaussian family, loess
929 smooth pseudotime) to fit a model for each gene using the backfitting algorithm. Genes are
930 considered dynamic based on significance of the smooth term (pseudotime).

931

932 **CLR**

933

934 The Context Likelihood of Relatedness (CLR) algorithm was first proposed in 2007 by Faith et
935 al. as a method for inferring gene regulatory networks from expression data. In short, CLR
936 computes the mutual information (MI) between expression levels of every potential transcription
937 factor and target gene. From here the method scores the MI values in the context of the network
938 by computing a z-score relying on (1) the distribution of MI values for all potential regulators of
939 the target gene, z_i, and on (2) the distribution of MI values for all potential targets of the
940 transcription factor, z_j:

941 $$z_{ij} = \sqrt{z_i^2 + z_j^2}$$

The result is a weighted adjacency matrix describing the likelihood of TF-target gene interactions that can be thresholded to derive a network structure. For Epoch, a version of CLR is applied to dynamically expressed genes to reconstruct an initial static network via the 'minet' package in R. Users can choose between implementing the MI version or a version using Pearson correlation in place of MI. In either case, Epoch at this step will return a GRN table. Each row outlines an interaction in the initial network identifying the transcription factor, target gene, correlation, and CLR-based z-score.

Cross-weighting

After inferring an initial network structure, Epoch can apply a network refinement step we call cross-weighting. The objective of this step is to negatively weight edges in the initial network that are unlikely to be true interactions, that may, for example, be representative of indirect interactions. To this end, for every TF-target pair in the initial network, Epoch computes cross-correlation across a given lag time (defaults to a fifth of total psuedotime). After ordering the lag times by decreasing correlation, Epoch computes an offset value where maximum correlation is achieved, defined by default as the top third of the ranked lag times. Finally, Epoch scores the offset values of each interactions:

$$weight = \begin{cases} 1 & x \leq w_{min} \\ 0 & x \geq w_{max} \\ -x/(w_{max} - w_{min}) + 1 & else \end{cases}$$

where the offset, x, is computed as described, and maximum and minimum windows can be altered by the user. Z-scores from the initial network are weighted accordingly, ultimately filtering out false positives. Epoch, at this step will return an appended GRN table, including for each interaction, the offset value and new weighted score.

Dynamic Network Extraction

Epoch will extract a dynamic network from the reconstructed static network. Specifically, the process begins by breaking pseudotime into "epochs," or time periods. A number of options are available to users to accomplish this. Briefly, Epoch can define the epochs based on pseudotime, equal cell ordering (resulting in equal number of cells per epoch), k-means clustering, or user-defined manual assignment. The number of epochs is specified by the user,

and can be determined by examining the heatmap of gene expression across pseudotime and estimating the rough number of expression states represented in the data. After epochs are defined, Epoch will assign genes to epochs, based on their activity along pseudotime. In brief, this is either based on activity (i.e. genes active in any epochs will be assigned to those epochs), or based on differential expression (i.e. genes are assigned based on if they are differentially expressed in an epoch). Specifically, if genes are assigned by activity, Epoch will compute, for each gene, a threshold against which average expression of the gene in an epoch is compared. This threshold can be modified by user input. If instead genes are assigned by differential expression, a p-value threshold is used to determine assignments. Finally, “orphan genes,” which we define as dynamically expressed genes that do not get assigned to any epoch, are assigned to the epoch in which their average expression is maximum.

Based on these assignments, Epoch will fracture the static network into a dynamic one composed of “epoch networks” and “transition networks”. Specifically, an edge between regulator and target gene appears in an epoch network if the regulator is assigned to that epoch. Further, an edge will appear in a transition subnetwork under two conditions: (1) For an activating edge, the target is not active in the source epoch, is active in the subsequent epoch, and regulator is active or (2) For a repressive edge, the target is active in the source epoch, is not active in the subsequent epoch, and the regulator is active. Following this step, Epoch will return a dynamic network represented by a list of individual GRN tables. This includes the epoch networks or essentially the dynamic network, as well as transition networks that describe how an epoch network may transition into a subsequent epoch network.

Top Regulator Prediction

Epoch employs two graph theoretic methods to predict “top regulators,” the transcription factors that appear to have the most influence in driving changes in or maintaining topology. First, Epoch will rank regulators by weighted PageRank. In brief, the PageRank centrality measures the importance of nodes in a network based on the number and quality of links of which a node is a part. In essence, the most influential nodes are likely to interact with many influential nodes. Second, Epoch will rank regulators by the product of normalized betweenness and normalized degree. Here, the assumption is that the most influential nodes are likely to be traversed by many shortest paths (and thus have high betweenness) and interact with many other nodes (and thus have high degree). In Epoch, PageRank, betweenness, and degree centralities are implemented through the ‘igraph’ package. By default, cross-weighted edge weights are used,

1009 but can be further specified by the user. In both top regulator prediction methods Epoch will
1010 return ranked lists of nodes, and further specify their corresponding PageRank or normalized
1011 betweenness, normalized degree, and betweenness-degree product.
1012
1013
1014 **Supplementary materials**
1015
1016 **Supplementary Figure 1.** Epoch benchmarking continued. (A) An example from GRN
1017 reconstruction on a synthetic dataset showing the difference in number of false positives based
1018 on whether or not cross-weighting is performed. (B) Full results of benchmarking reconstruction
1019 of synthetic data. Four versions of CLR (Pearson vs. MI, cross-weighting vs. no cross-
1020 weighting), four versions of Epoch (same as CLR), and three versions of GENIE3 (original,
1021 limited to dynamic genes, and with cross-weighting) are compared. (C) Full results of
1022 benchmarking reconstruction of the muscle development data. Methods are the same as in A.
1023 (D) Early mESC directed differentiation data (subsetted from Spangler et al. 2018). Zfp42
1024 expression, Fgf5 expression, and velocity pseudotime are shown. (E) Full results of
1025 benchmarking reconstruction of the early Spangler et al. data. Methods are the same as in A.
1026 (F) Clustering, diffusion pseudotime, velocity pseudotime, and latent time of the E12.5 muscle
1027 development data. (G) Precision recall curves of four variations of Epoch using three different
1028 pseudotime annotations. (H) Effect of smoothing vs. unsmoothing on AUPR of reconstruction
1029 using various methods.
1030
1031 **Supplementary Figure 2.** MULTI-seq directed differentiation cell typing. (A) timepoint,
1032 treatment, total counts, and percent mitochondrial transcripts content of the MULTI-seq data. (B)
1033 Select marker gene expression of the MULTI-seq data. (C) Differential expression analysis of
1034 the MULTI-seq data by cluster.
1035
1036 **Supplementary Figure 3.** MULTI-seq cell type classification and cluster transitions. (A)
1037 SingleCellNet classification results. The classifier was trained from mouse embryo data found in
1038 literature, curated by members in the lab. (B) Sankey diagrams depicting cluster transitions (and
1039 cell fate transition proportions) as broken down by treatment.
1040
1041 **Supplementary Figure 4.** MULTI-seq network reconstruction and top regulators. Heatmaps
1042 depicting TF expression across time and Epoch definitions along the (A) mesoderm path, (B)

endoderm path, and (C) neuroectoderm path. (D) Top 10 regulators in each epoch network along each germ layer path as predicted by Epoch via PageRank.

Supplementary Figure 5. Bootstrapped top regulator prediction for inner cell mass to epiblast-like transition. (A) Workflow for bootstrapping network reconstruction in order to predict top regulators. (B) Only cells from early time were used. Clusters used and corresponding diffusion pseudotime shown here. (C) Top regulators for this early fate transition as predicted using PageRank and Betweenness-Degree.

Supplementary Figure 6. Proof of concept for differential network analysis. (A) Path-unique gold standard creation. (B) Performance (AUPR) over random of differential network in recovering path-unique gold standard edges. Each row is a comparison between two paths. Each column represents unique edges recovered for a given path. As an example, the top left box indicates that in the endoderm vs. neuroectoderm comparison, in the first epoch the endoderm differential network is 1.82 times better than a shuffled network in recovering edges that are specific to endoderm (and not present in neuroectoderm).

Supplementary Figure 7. Mesodermal network and differential networks. The left column ("Full Reconstruction") depicts the dynamic mesodermal network. The middle and right columns depict the WAG vs. WA differential network. Middle is the network unique to WAG, right is the network unique to WA. TFs are colored by community. Blue and red edges represent activating and repressive edges respectively.

Supplementary Figure 8. Mesodermal network analysis continued. (A) Average community expression over time by treatment along the full dataset. Communities shown (each row) are from the epoch 2 subnetwork (yellow), transition (green), and epoch 3 subnetworks (aqua) of the mesodermal network. (B) Peg3 and Dnmt3a expression across time along the mesodermal path of the MULTI-seq data. (C) Top differential regulators of mesodermal genes. Blue and red represent activators and repressors respectively. Y-axis is plotting difference in edge weights between the WAG and WA networks. Thus, TFs appearing above the 0-line are more unique to the WAG network, and those below are more unique to the WA network.

Supplementary Table 1. Cell type annotation by cluster.

1077

1078

Bibliography

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Androutsellis-Theotokis, A., Leker, R.R., Soldner, F., Hoeppner, D.J., Ravin, R., Poser, S.W., Rueger, M.A., Bae, S.-K., Kittappa, R., and McKay, R.D.G. (2006). Notch signalling regulates stem cell numbers in vitro and in vivo. Nature *442*, 823–826.

Ang, S.L., Wierda, A., Wong, D., Stevens, K.A., Cascio, S., Rossant, J., and Zaret, K.S. (1993). The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. Development *119*, 1301–1315.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Aubert, J., Dunstan, H., Chambers, I., and Smith, A. (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. Nat. Biotechnol. *20*, 1240–1245.

Barral, A., Rollan, I., Sanchez-Iranzo, H., Jawaid, W., Badia-Careaga, C., Menchero, S., Gomez, M.J., Torroja, C., Sanchez-Cabo, F., Göttgens, B., et al. (2019). Nanog regulates Pou3f1 expression at the exit from pluripotency during gastrulation. Biol. Open *8*.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2019). Generalizing RNA velocity to transient cell states through dynamical modeling. BioRxiv.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414.

di Bernardo, D., Thompson, M.J., Gardner, T.S., Chobot, S.E., Eastwood, E.L., Wojtovich, A.P., Elliott, S.J., Schaus, S.E., and Collins, J.J. (2005). Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. Nat. Biotechnol. *23*, 377–383.

Bonnaffoux, A., Herbach, U., Richard, A., Guillemin, A., Gonin-Giraud, S., Gros, P.-A., and Gandrillon, O. (2019). WASABI: a dynamic iterative framework for gene regulatory network inference. BMC Bioinformatics *20*, 220.

Brennan, J., Lu, C.C., Norris, D.P., Rodriguez, T.A., Beddington, R.S., and Robertson, E.J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. Nature *411*, 965–969.

Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems *30*, 107–117.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cahan, P., Cacchiarelli, D., Dunn, S.-J., Hemberg, M., de Sousa Lopes, S.M.C., Morris, S.A., Rackham, O.J.L., Del Sol, A., and Wells, C.A. (2021). Computational stem cell biology: open questions and guiding principles. Cell Stem Cell *28*, 20–32.

Cannoodt, R., Saelens, W., Deconinck, L., and Saeys, Y. (2020). dyngen: a multi-modal simulator for spearheading new single-cell omics analyses. BioRxiv.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318–325.

Chen, S., and Mar, J.C. (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. BMC Bioinformatics

19, 232.
Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.
Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F., and Poidinger, M. (2016). Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. Nat. Commun. *7*, 11988.
Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J. Biol. Chem. *277*, 38746–38754.
Costello, I., Nowotschin, S., Sun, X., Mould, A.W., Hadjantonakis, A.-K., Bikoff, E.K., and Robertson, E.J. (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. Genes Dev. *29*, 2108–2122.
Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796–800.
Dhaliwal, N.K., Miri, K., Davidson, S., Tamim El Jarkass, H., and Mitchell, J.A. (2018). KLF4 Nuclear Export Requires ERK Activation and Initiates Exit from Naive Pluripotency. Stem Cell Rep. *10*, 1308–1323.
Dufort, D., Schwartz, L., Harpal, K., and Rossant, J. (1998). The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. Development *125*, 3015–3025.
Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160.
Fagnocchi, L., Mazzoleni, S., and Zippo, A. (2016). Integration of Signaling Pathways with the Epigenetic Machinery in the Maintenance of Stem Cells. Stem Cells Int. *2016*, 8652748.
Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. *5*, e8.
Gadue, P., Huber, T.L., Paddison, P.J., and Keller, G.M. (2006). Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. Proc. Natl. Acad. Sci. USA *103*, 16806–16811.
Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334.
Genga, R.M.J., Kernfeld, E.M., Parsi, K.M., Parsons, T.J., Ziller, M.J., and Maehr, R. (2019). Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early Human Endoderm Development. Cell Rep. *27*, 708–718.e10.
Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.
Grosswendt, S., Kretzmer, H., Smith, Z.D., Kumar, A.S., Hetzel, S., Wittler, L., Klages, S., Timmermann, B., Mukherji, S., and Meissner, A. (2020). Epigenetic regulator function through mouse gastrulation. Nature *584*, 102–108.
Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. *47*, D853–D858.
Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nat. Methods *13*, 845–848.

1172 Hart, A.H., Hartley, L., Sourris, K., Stadler, E.S., Li, R., Stanley, E.G., Tam, P.P.L., Elefanty,
1173 A.G., and Robb, L. (2002). *Mixl1* is required for axial mesendoderm morphogenesis and
1174 patterning in the murine embryo. Development *129*, 3597–3608.
1175 Hartemink, A.J. (2005). Reverse engineering gene regulatory networks. Nat. Biotechnol. *23*,
1176 554–555.
1177 Haub, O., and Goldfarb, M. (1991). Expression of the fibroblast growth factor-5 gene in the
1178 mouse embryo. Development *112*, 397–406.
1179 Hayward, P., Kalmar, T., and Arias, A.M. (2008). Wnt/Notch signalling and information
1180 processing during development. Development *135*, 411–424.
1181 Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory
1182 networks from expression data using tree-based methods. PLoS One *5*.
1183 Kanai-Azuma, M., Kanai, Y., Gad, J.M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y.,
1184 Yonekawa, H., Yazaki, K., Tam, P.P.L., et al. (2002). Depletion of definitive gut endoderm in
1185 Sox17-null mutant mice. Development *129*, 2367–2379.
1186 Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. Nat.
1187 Rev. Mol. Cell Biol. *9*, 770–780.
1188 Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz,
1189 D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to
1190 embryonic stem cells. Cell *161*, 1187–1201.
1191 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S.,
1192 Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set
1193 enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90-7.
1194 Kuroiwa, Y., Kaneko-Ishino, T., Kagitani, F., Kohda, T., Li, L.L., Tada, M., Suzuki, R.,
1195 Yokoyama, M., Shiroishi, T., Wakana, S., et al. (1996). Peg3 imprinted gene on proximal
1196 chromosome 7 encodes for a zinc finger protein. Nat. Genet. *12*, 186–190.
1197 La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber,
1198 K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature
1199 *560*, 494–498.
1200 Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010).
1201 ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X
1202 experiments. Bioinformatics *26*, 2438–2444.
1203 Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. Nat.
1204 Rev. Genet. *16*, 146–158.
1205 Letsou, W., and Cai, L. (2016). Noncommutative biology: sequential regulation of complex
1206 networks. PLoS Comput. Biol. *12*, e1005089.
1207 Lu, T., and Mar, J.C. (2020). Investigating transcriptome-wide sex dimorphism by multi-level
1208 analysis of single-cell RNA sequencing data in ten mouse cell types. Biol. Sex Differ. *11*, 61.
1209 Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas,
1210 A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression
1211 Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.
1212 Marbach, D., Prill, R.J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010).
1213 Revealing strengths and weaknesses of methods for gene network inference. Proc. Natl. Acad.
1214 Sci. USA *107*, 6286–6291.
1215 Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R.,
1216 DREAM5 Consortium, Kellis, M., Collins, J.J., et al. (2012). Wisdom of crowds for robust gene
1217 network inference. Nat. Methods *9*, 796–804.

1218 Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and
1219 Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks
1220 in a mammalian cellular context. BMC Bioinformatics *7 Suppl 1*, S7.
1221 Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N., Hayashi, T., and
1222 Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell
1223 RNA-Seq during differentiation. Bioinformatics *33*, 2314–2321.
1224 McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu,
1225 J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019). MULTI-seq: sample multiplexing
1226 for single-cell RNA sequencing using lipid-tagged indices. Nat. Methods *16*, 619–626.
1227 McLean, A.B., D'Amour, K.A., Jones, K.L., Krishnamoorthy, M., Kulik, M.J., Reynolds, D.M.,
1228 Sheppard, A.M., Liu, H., Xu, Y., Baetge, E.E., et al. (2007). Activin a efficiently specifies
1229 definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase
1230 signaling is suppressed. Stem Cells *25*, 29–38.
1231 Meredith, G.D., D'Ippolito, A., Dudas, M., Zeidner, L.C., Hostetter, L., Faulds, K., Arnold, T.H.,
1232 Popkie, A.P., Doble, B.W., Marnellos, G., et al. (2015). Glycogen synthase kinase-3 (Gsk-3)
1233 plays a fundamental role in maintaining DNA methylation at imprinted loci in mouse embryonic
1234 stem cells. Mol. Biol. Cell *26*, 2139–2150.
1235 Meyer, P.E., Kontos, K., Lafitte, F., and Bontempi, G. (2007). Information-theoretic inference of
1236 large transcriptional regulatory networks. EURASIP J. Bioinform. Syst. Biol. 79879.
1237 Meyer, P.E., Lafitte, F., and Bontempi, G. (2008). minet: A R/Bioconductor package for
1238 inferring large transcriptional networks using mutual information. BMC Bioinformatics *9*, 461.
1239 Mohammad, H.P., and Baylin, S.B. (2010). Linking cell signaling and the epigenetic machinery.
1240 Nat. Biotechnol. *28*, 1033–1038.
1241 Morgan, D., Studham, M., Tjärnberg, A., Weishaupt, H., Swartling, F.J., Nordling, T.E.M., and
1242 Sonnhammer, E.L.L. (2020). Perturbation-based gene regulatory network inference to unravel
1243 oncogenic mechanisms. Sci. Rep. *10*, 14149.
1244 Muñoz-Sanjuán, I., and Brivanlou, A.H. (2002). Neural induction, the default model and
1245 embryonic stem cells. Nat. Rev. Neurosci. *3*, 271–280.
1246 Nord, J., Schill, D., Pulakanti, K., Rao, S., and Cirillo, L.A. (2020). The transcription factor
1247 FoxO1 is required for the establishment of the human definitive endoderm. BioRxiv.
1248 Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R.,
1249 Sese, J., and Meno, C. (2018). ChIP-Atlas: a data-mining suite powered by full integration of
1250 public ChIP-seq data. EMBO Rep. *19*.
1251 Park, K.-S., Cha, Y., Kim, C.-H., Ahn, H.-J., Kim, D., Ko, S., Kim, K.-H., Chang, M.-Y., Ko, J.-
1252 H., Noh, Y.-S., et al. (2013). Transcription elongation factor Tcea3 regulates the pluripotent
1253 differentiation potential of mouse embryonic stem cells via the Lefty1-Nodal-Smad2 pathway.
1254 Stem Cells *31*, 282–292.
1255 Pelton, T.A., Sharma, S., Schulz, T.C., Rathjen, J., and Rathjen, P.D. (2002). Transient
1256 pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in
1257 vitro pluripotent cell development. J. Cell Sci. *115*, 329–339.
1258 Perea-Gomez, A., Vella, F.D.J., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C.,
1259 Pfister, V., Chen, L., Robertson, E., Hamada, H., et al. (2002). Nodal antagonists in the anterior
1260 visceral endoderm prevent the formation of multiple primitive streaks. Dev. Cell *3*, 745–756.
1261 Philippidou, P., and Dasen, J.S. (2013). Hox genes: choreographers in neural development,
1262 architects of circuit organization. Neuron *80*, 12–34.
1263 Popkie, A.P., Zeidner, L.C., Albrecht, A.M., D'Ippolito, A., Eckardt, S., Newsom, D.E., Groden,

1264 J., Doble, B.W., Aronow, B., McLaughlin, K.J., et al. (2010). Phosphatidylinositol 3-kinase
1265 (PI3K) signaling via glycogen synthase kinase-3 (Gsk-3) regulates DNA methylation of
1266 imprinted loci. J. Biol. Chem. *285*, 41337–41347.
1267 Qin, G., Yang, L., Ma, Y., Liu, J., and Huo, Q. (2019). The exploration of disease-specific gene
1268 regulatory networks in esophageal carcinoma and stomach adenocarcinoma. BMC
1269 Bioinformatics *20*, 717.
1270 Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Bruggner, R.V., Linderman, M.D., Sachs, K.,
1271 Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional
1272 cytometry data with SPADE. Nat. Biotechnol. *29*, 886–891.
1273 Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T., McFaline-Figueroa, J.L.,
1274 Saunders, L., Trapnell, C., and Kannan, S. (2020). Inferring Causal Gene Regulatory Networks
1275 from Coupled Single-Cell Expression Dynamics Using Scribe. Cell Syst. *10*, 265–274.e11.
1276 Rackham, O.J.L., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., FANTOM
1277 Consortium, Suzuki, H., Nefzger, C.M., Daub, C.O., et al. (2016). A predictive computational
1278 framework for direct reprogramming between human cell types. Nat. Genet. *48*, 331–335.
1279 Rogers, M.B., Hosler, B.A., and Gudas, L.J. (1991). Specific expression of a retinoic acid-
1280 regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes.
1281 Development *113*, 815–824.
1282 Spangler, A., Su, E.Y., Craft, A.M., and Cahan, P. (2018). A single cell transcriptional portrait of
1283 embryoid body differentiation and comparison to progenitors of the developing embryo. Stem
1284 Cell Res. *31*, 201–215.
1285 Stark, J., Brewer, D., Barenco, M., Tomescu, D., Callard, R., and Hubank, M. (2003).
1286 Reconstructing gene networks: what are the limits? Biochem. Soc. Trans. *31*, 1519–1525.
1287 Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S.
1288 (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC
1289 Genomics *19*, 477.
1290 Sun, X., Meyers, E.N., Lewandoski, M., and Martin, G.R. (1999). Targeted disruption of Fgf8
1291 causes failure of cell migration in the gastrulating mouse embryo. Genes Dev. *13*, 1834–1846.
1292 Tan, Y., and Cahan, P. (2019). SingleCellNet: A Computational Tool to Classify Single Cell
1293 RNA-Seq Data Across Platforms and Across Species. Cell Syst. *9*, 207–213.e2.
1294 Theka, I., Sottile, F., Aulicino, F., Garcia, A.C., and Cosma, M.P. (2017). Reduced expression of
1295 Paternally Expressed Gene-3 enhances somatic cell reprogramming through mitochondrial
1296 activity perturbation. Sci. Rep. *7*, 9705.
1297 Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*,
1298 1491–1498.
1299 Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak,
1300 K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions
1301 are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.
1302 Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai,
1303 T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal
1304 lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.
1305 Tropepe, V., Hitoshi, S., Sirard, C., Mak, T.W., Rossant, J., and van der Kooy, D. (2001). Direct
1306 neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell
1307 stage acquired through a default mechanism. Neuron *30*, 65–78.
1308 Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C.,
1309 Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a

continuous process. Nat. Cell Biol. *19*, 271–281.
Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*.
Wang, Y., Lin, L., Lai, H., Parada, L.F., and Lei, L. (2013). Transcription factor Sox11 is essential for both embryonic and adult neurogenesis. Dev. Dyn. *242*, 638–653.
Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.
Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.
Yamamoto, M., Saijoh, Y., Perea-Gomez, A., Shawlot, W., Behringer, R.R., Ang, S.-L., Hamada, H., and Meno, C. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. Nature *428*, 387–392.
Yeo, J.-C., Jiang, J., Tan, Z.-Y., Yim, G.-R., Ng, J.-H., Göke, J., Kraus, P., Liang, H., Gonzales, K.A.U., Chong, H.-C., et al. (2014). Klf2 is an essential factor that sustains ground state pluripotency. Cell Stem Cell *14*, 864–872.
Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., and Jarvis, E.D. (2004). Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics *20*, 3594–3603.
Yu, J.S.L., Ramasamy, T.S., Murphy, N., Holt, M.K., Czapiewski, R., Wei, S.-K., and Cui, W. (2015). PI3K/mTORC2 regulates TGF-β/Activin signalling by modulating Smad2/3 activity via linker phosphorylation. Nat. Commun. *6*, 7212.
Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.
Zhu, Q., Song, L., Peng, G., Sun, N., Chen, J., Zhang, T., Sheng, N., Tang, W., Qian, C., Qiao, Y., et al. (2014). The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways. Elife *3*.

Figure 1

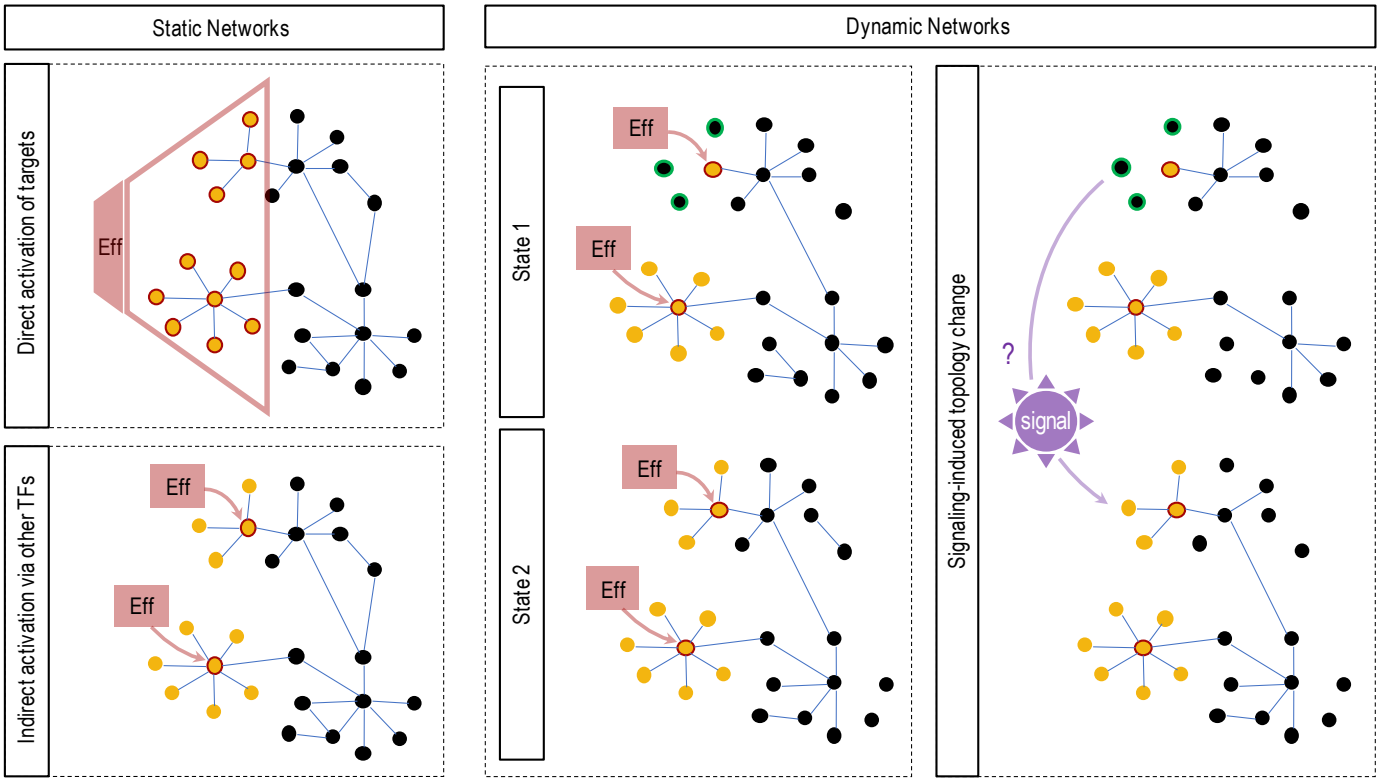

Static Networks
Dynamic Networks
Direct activation of targets
Indirect activation via other TFs
State 1
State 2
Signaling-induced topology change
Eff
signal
?

Figure 2

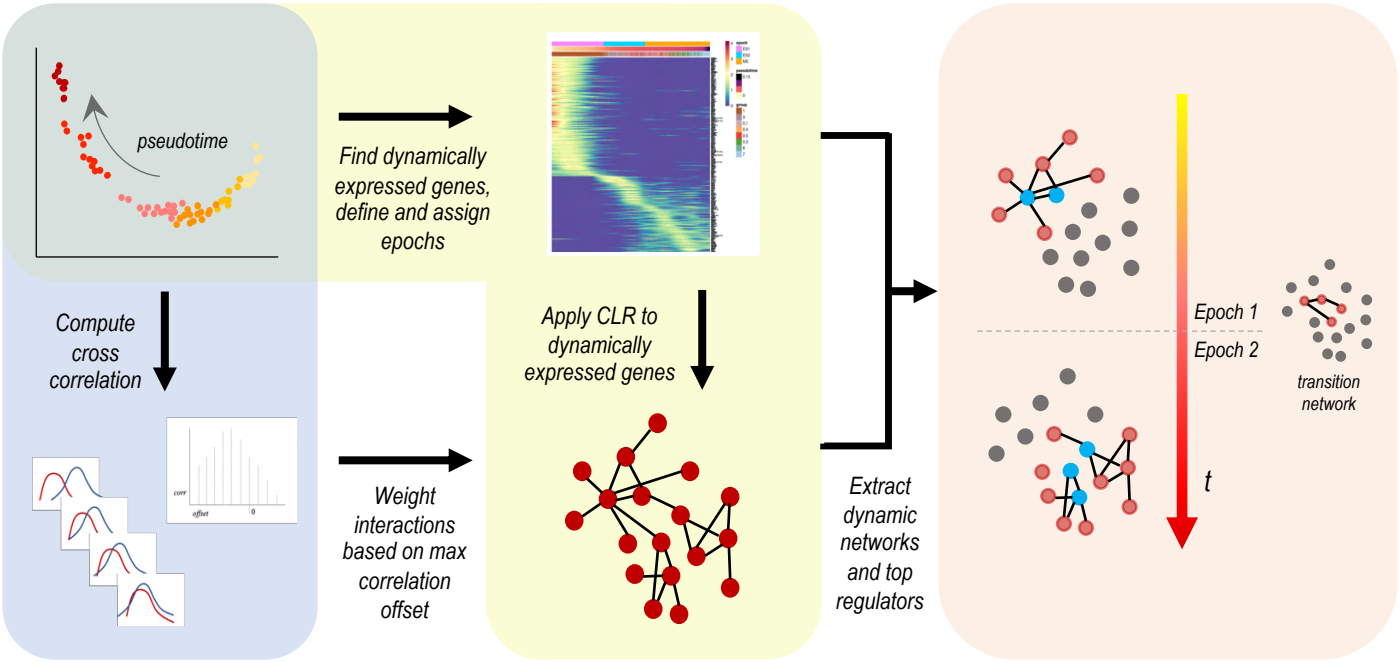
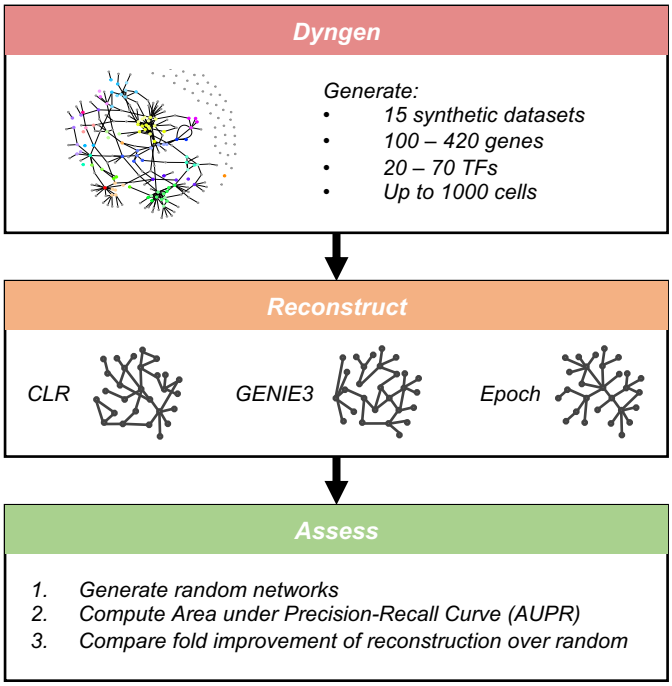
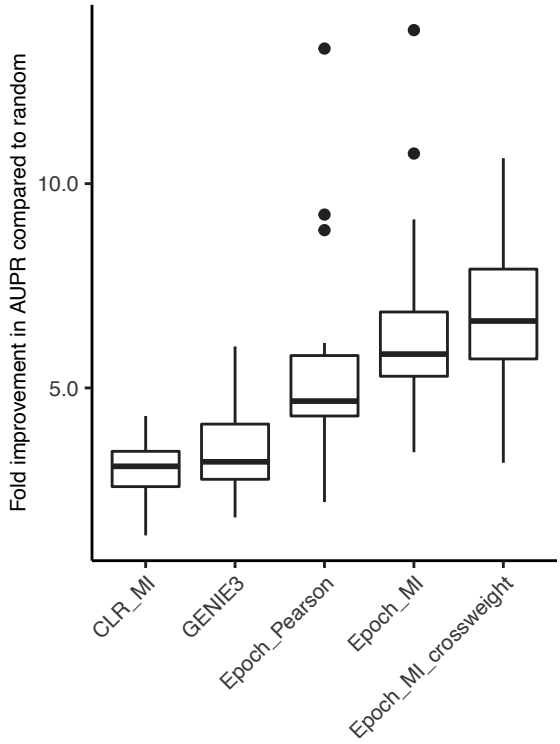

pseudotime
Find dynamically expressed genes, define and assign epochs
Compute cross correlation
Apply CLR to dynamically expressed genes
Weight interactions based on max correlation offset
Extract dynamic networks and top regulators
Epoch 1
Epoch 2
transition network
t

Figure 3

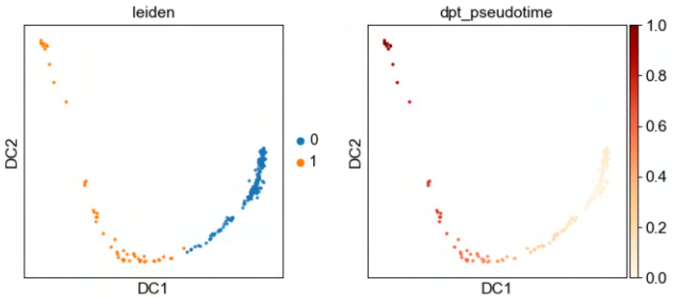
A



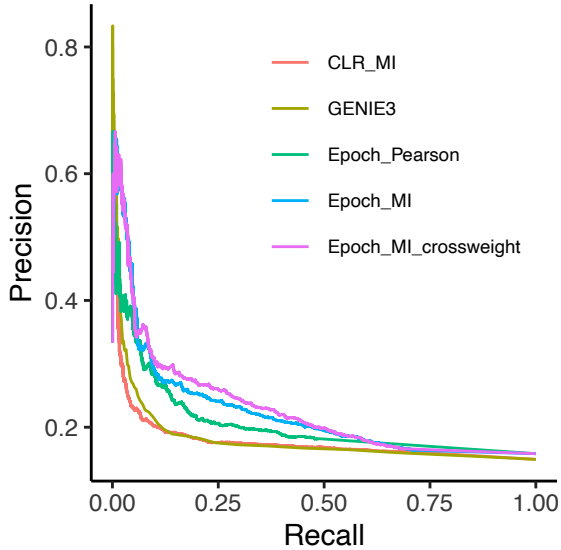
B



C



D

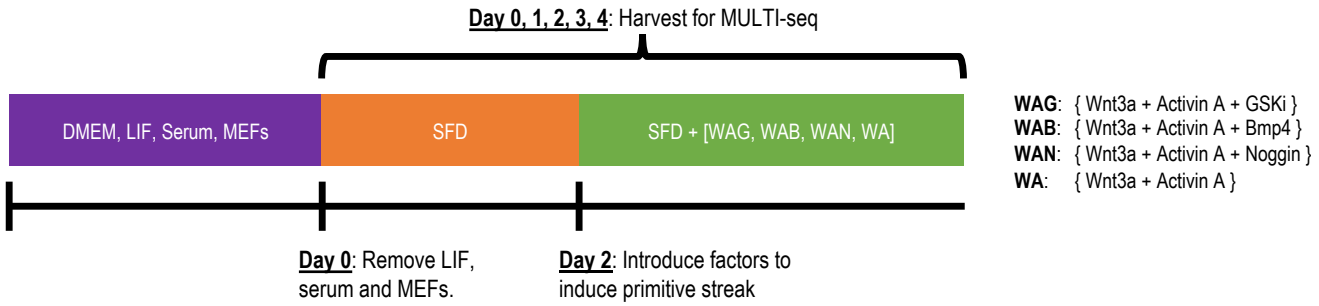


E

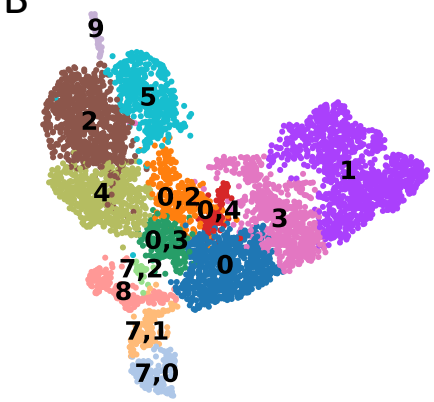
Method	Time
GENIE3	21.4 min
Epoch--Pearson	16.8 sec
Epoch--MI	1.4 min
Epoch--MI + cross-weighting	2.5 min

Figure 4

A

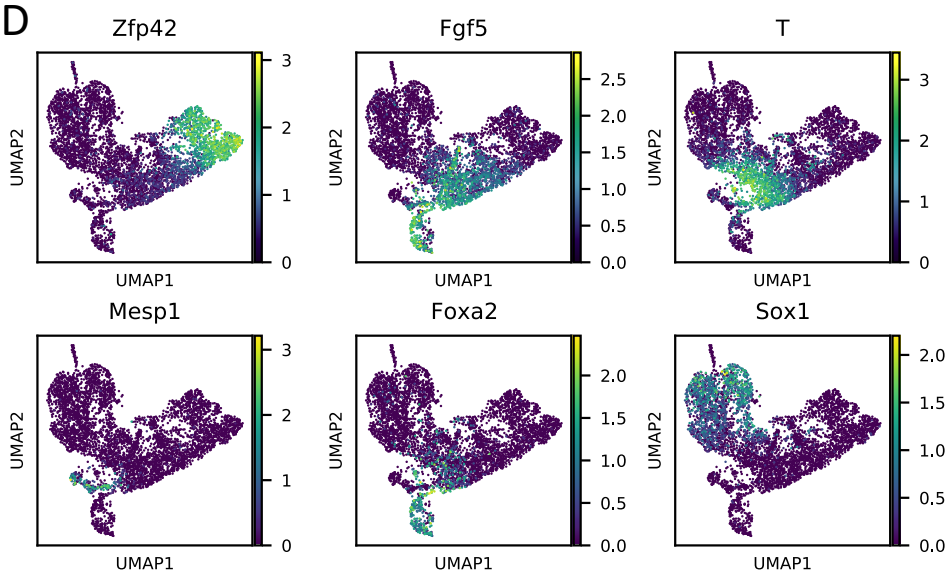


B

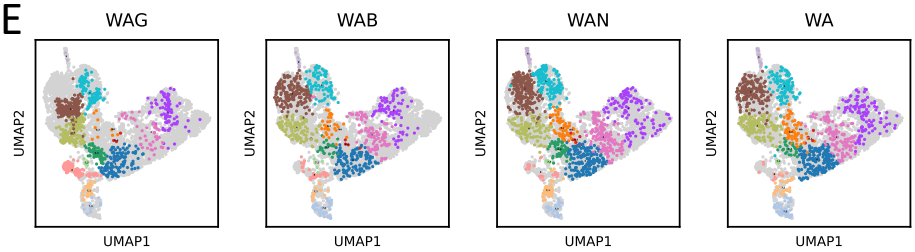


C

D



E



F

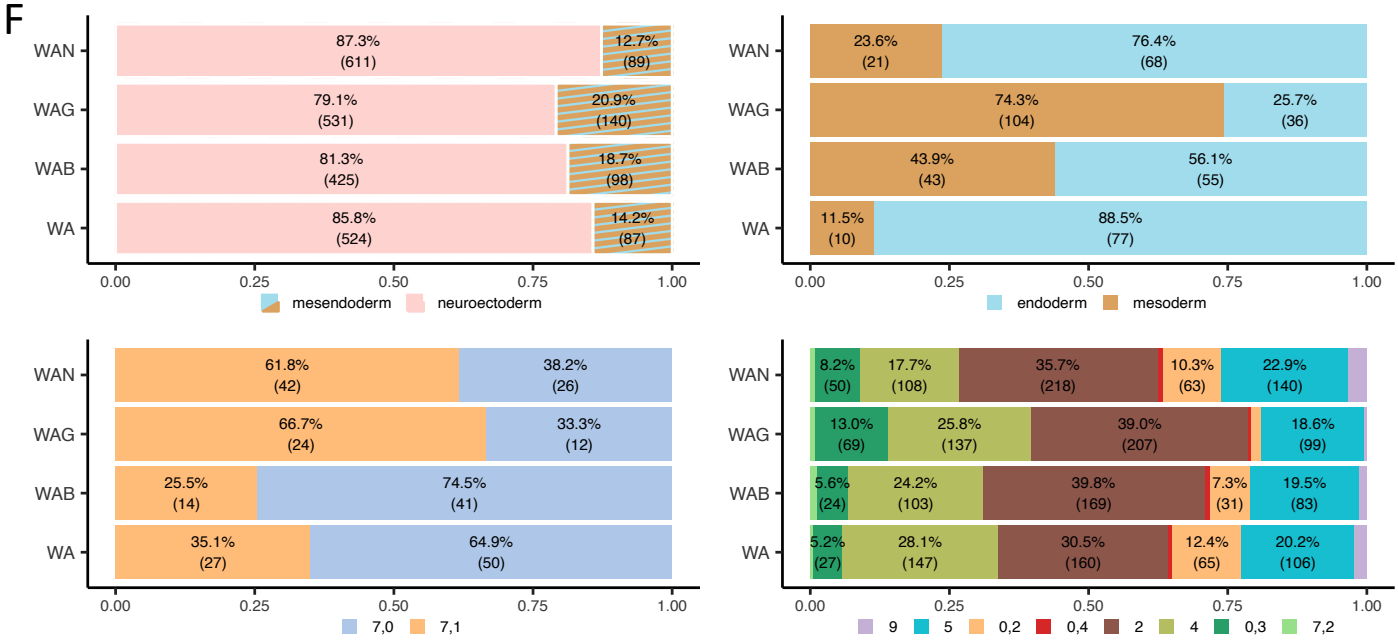


Figure 5

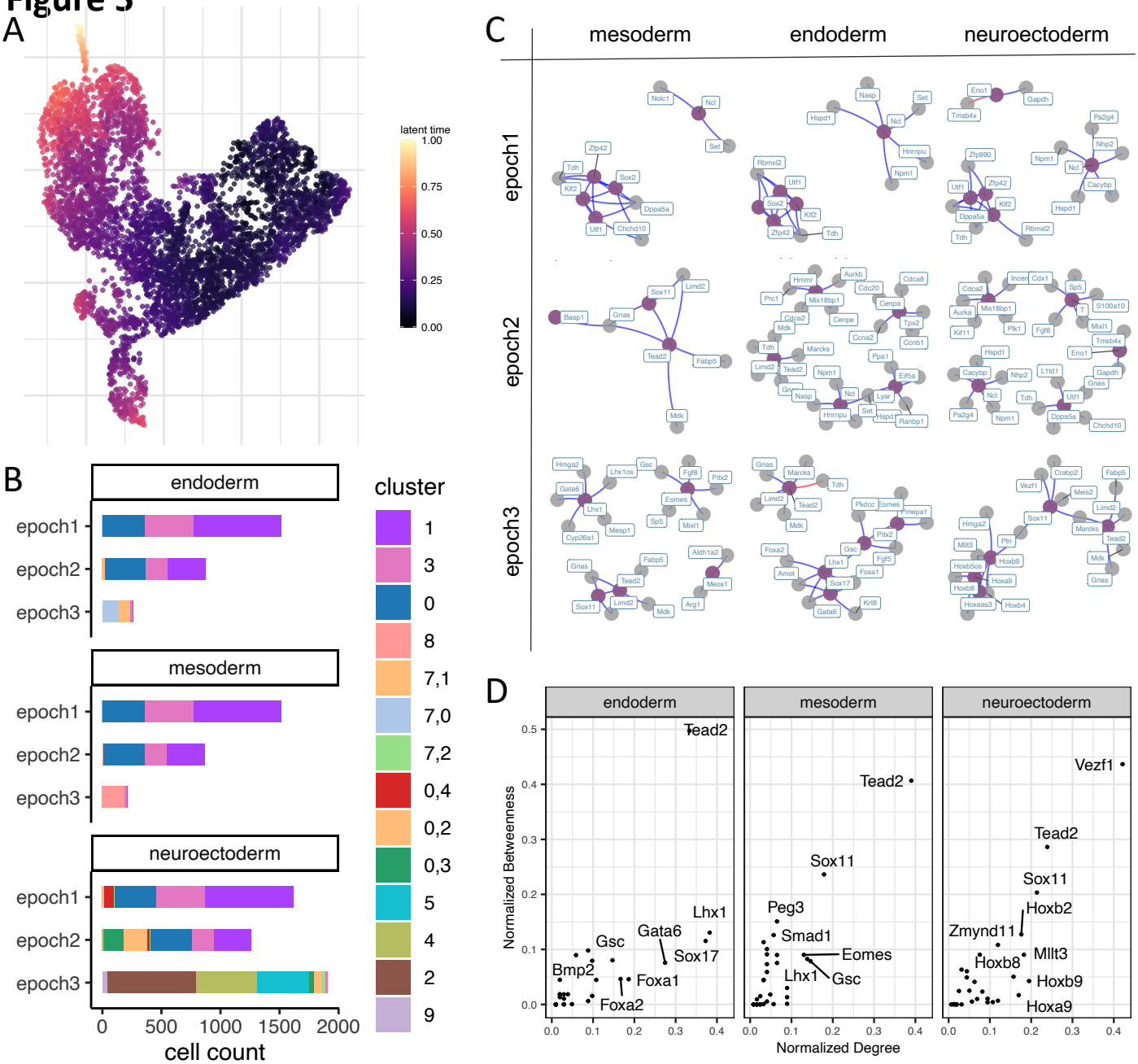
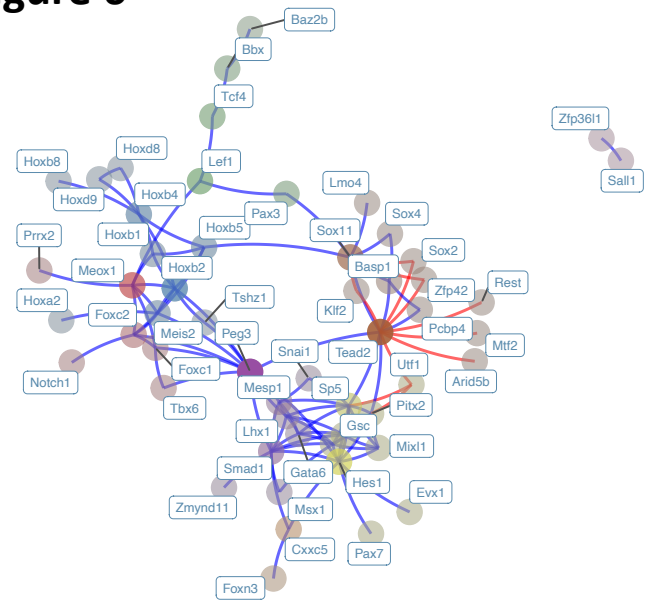
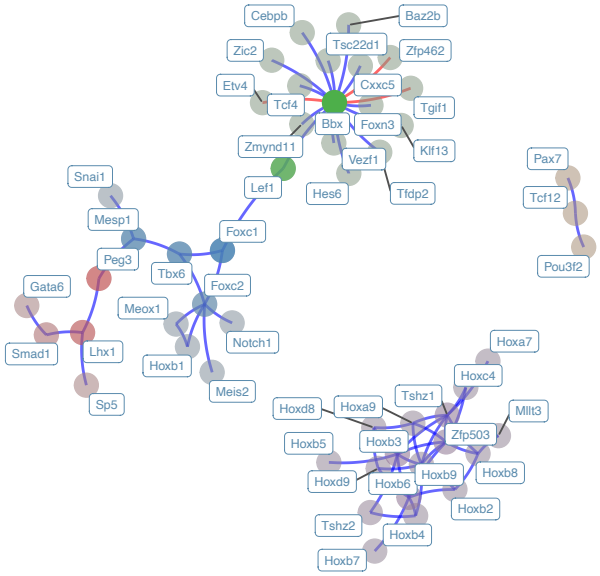

A
latent time
1.00
0.75
0.50
0.25
0.00
B
endoderm
mesoderm
neuroectoderm
epoch1
epoch2
epoch3
cluster
1
3
0
8
7,1
7,0
7,2
0,4
0,2
0,3
5
4
2
9
0
500
1000
1500
2000
cell count
C
mesoderm
endoderm
neuroectoderm
epoch1
epoch2
epoch3
D
endoderm
mesoderm
neuroectoderm
Normalized Betweenness
Normalized Degree
Tead2
Lhx1
Gata6
Gsc
Sox17
Bmp2
Foxa1
Foxa2
Sox11
Peg3
Smad1
Eomes
Vezf1
Zmynd11
Hoxb2
Mllt3
Hoxb8
Hoxb9
Hoxa9

Figure 6

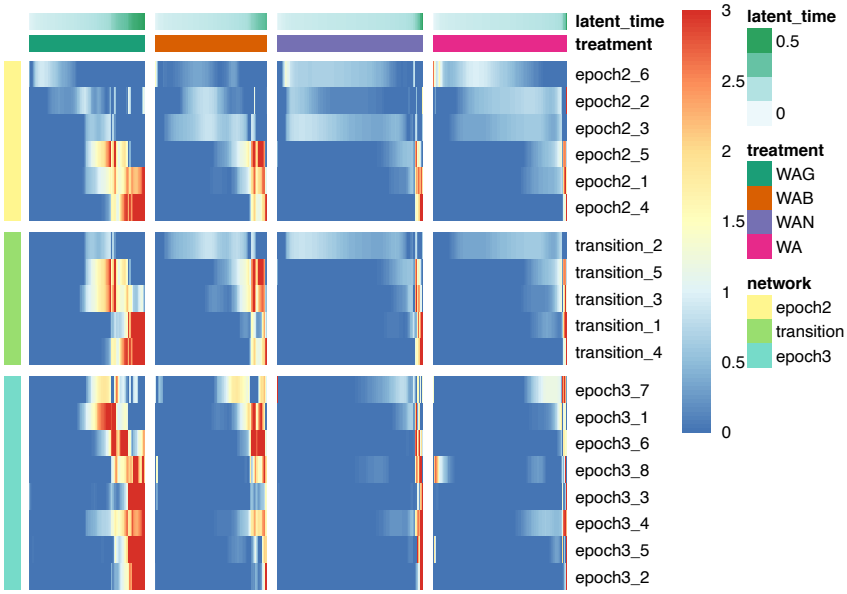
A



C



B



D

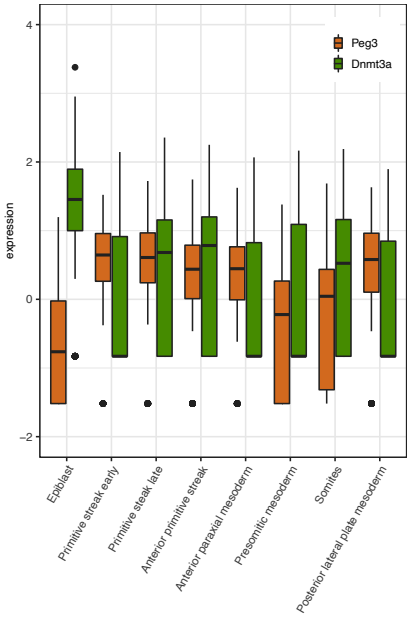
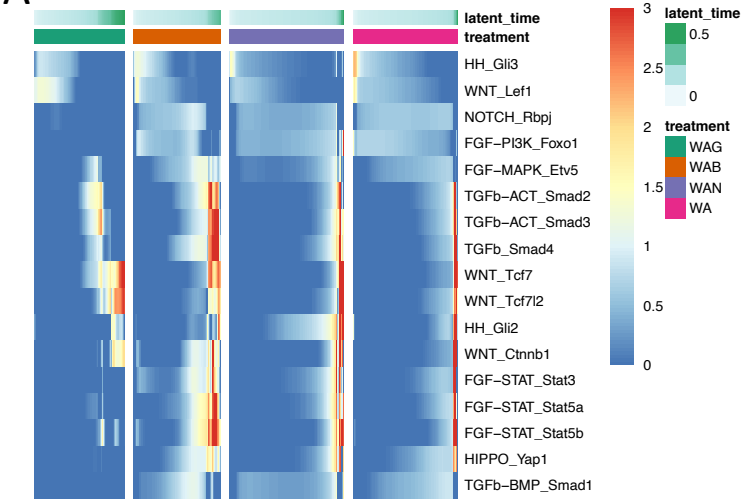
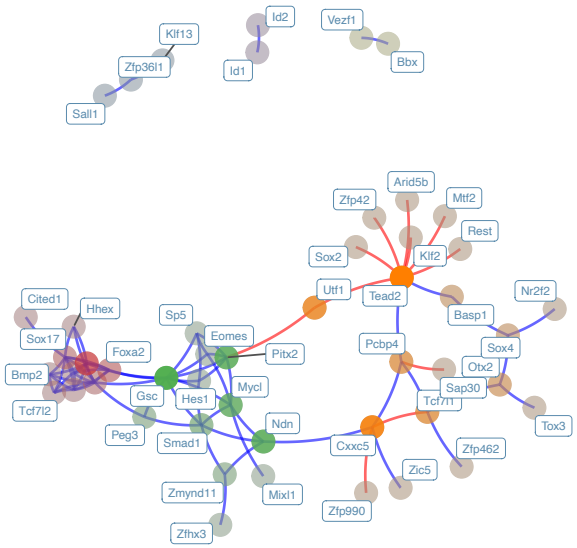


Figure 7

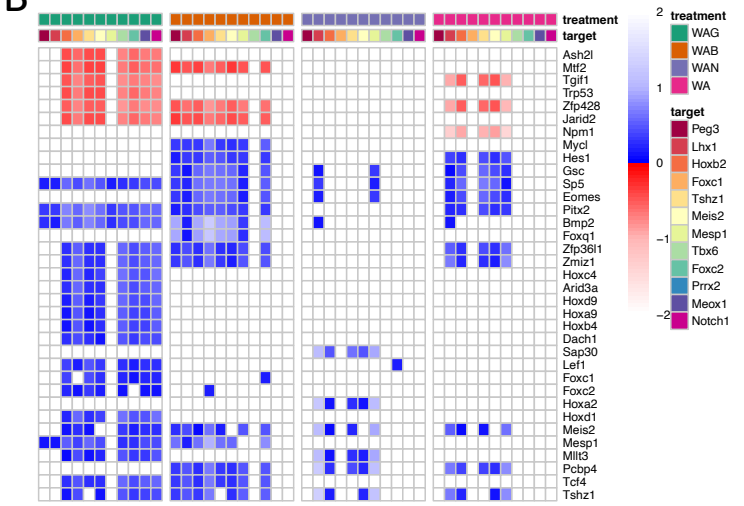
A



C



B



D

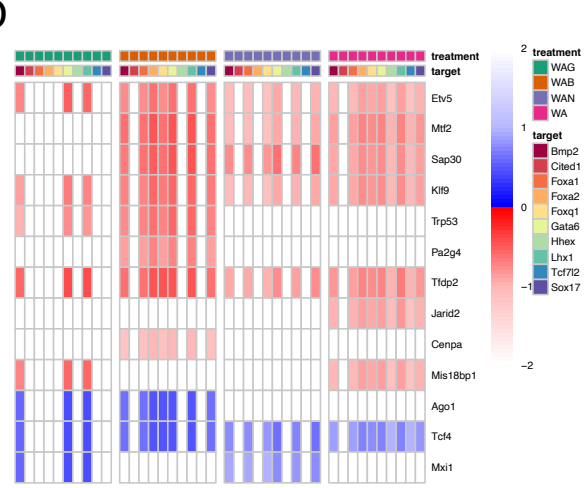
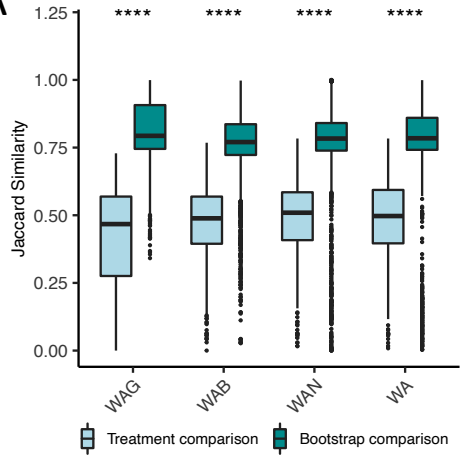


Figure 8

A

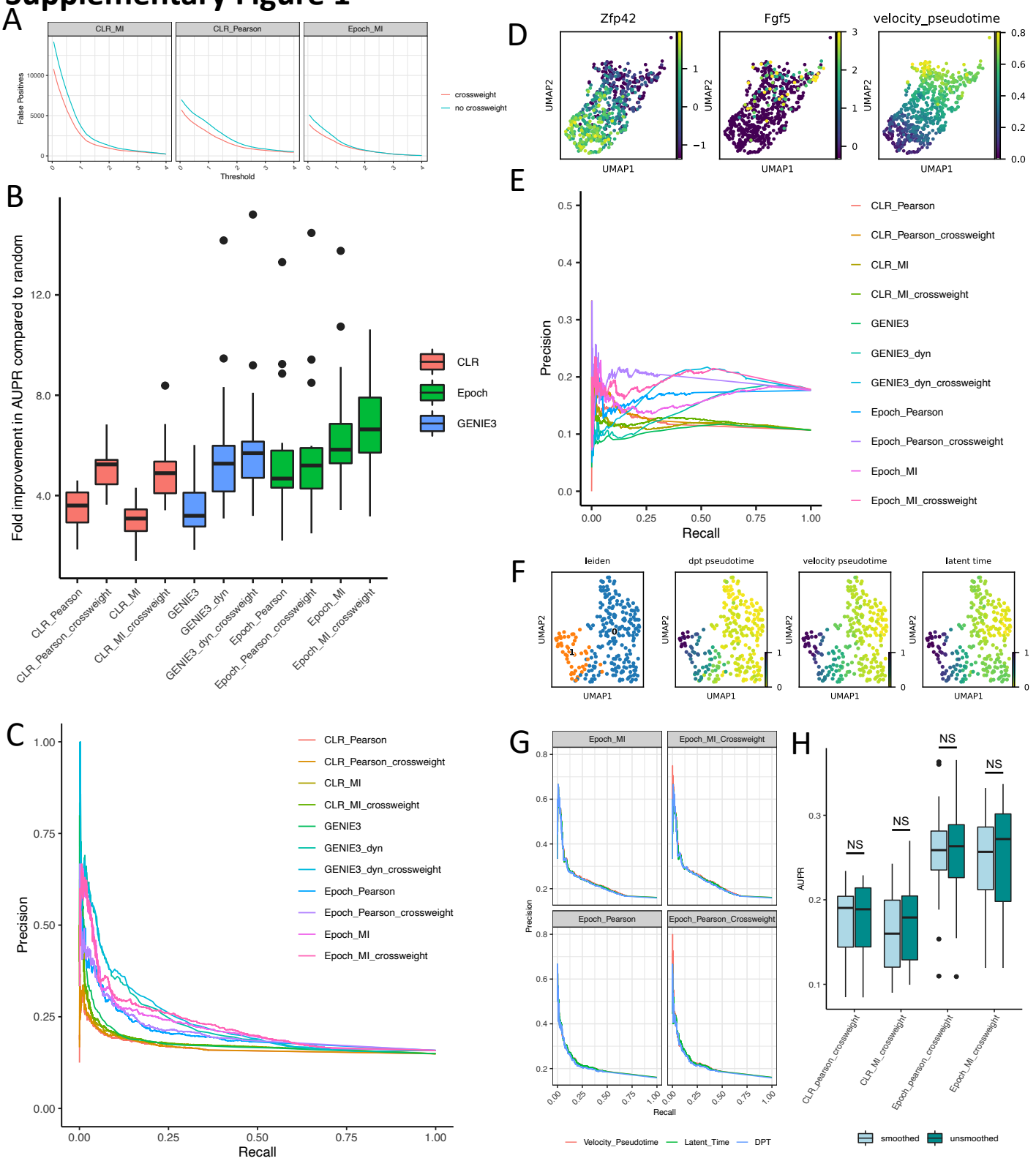


B

TF	avg(JS)	Hes7	0.124	Lmo2	0.260
Prrx2	0.061	Cited1	0.130	Mixl1	0.280
Tbx6	0.068	Gata6	0.192	Lhx1	0.283
Mesp1	0.099	Isl1	0.203	Ets1	0.297
Gata4	0.114	Hey1	0.213		

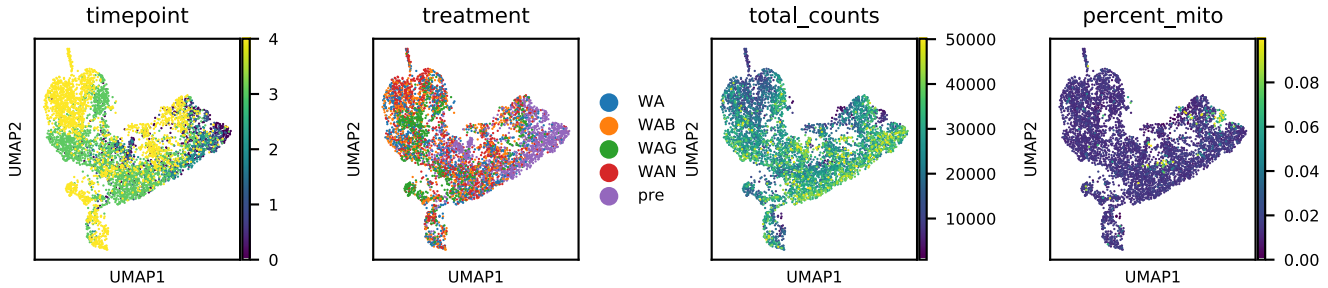
C

Supplementary Figure 1

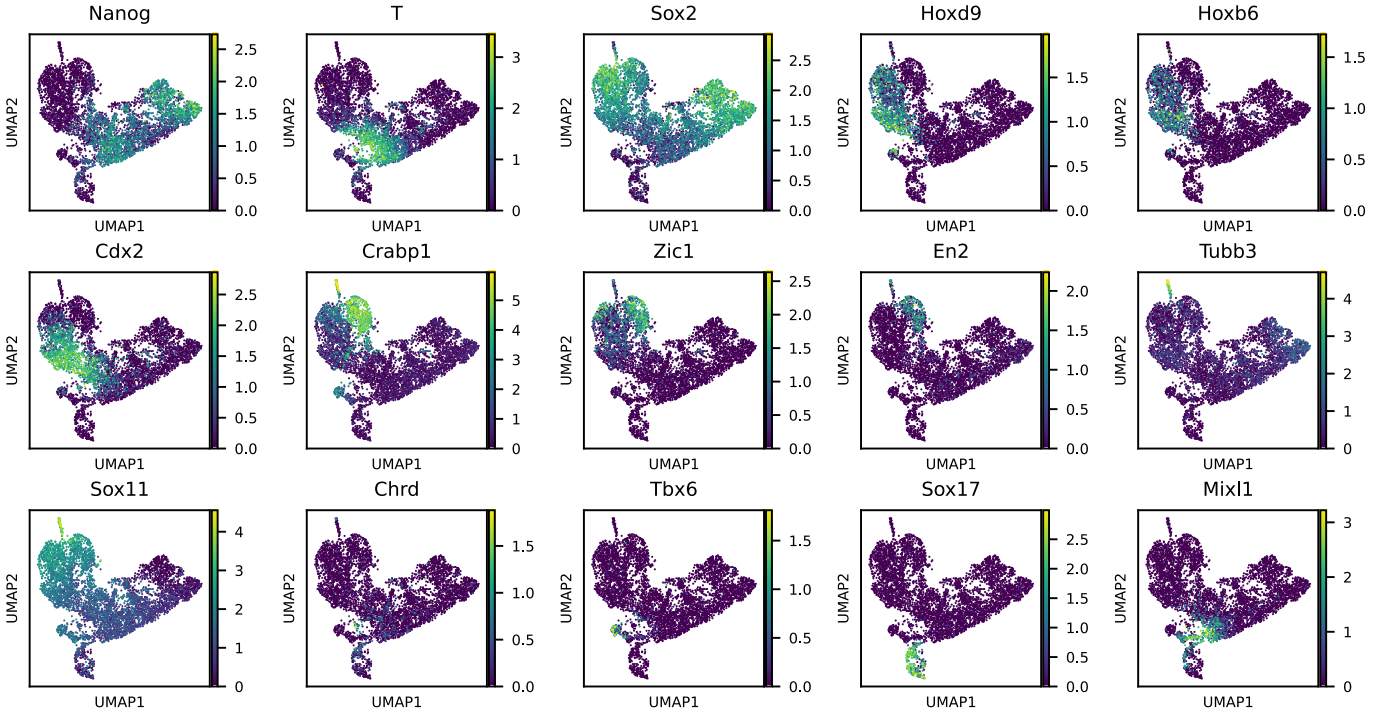


Supplementary Figure 2

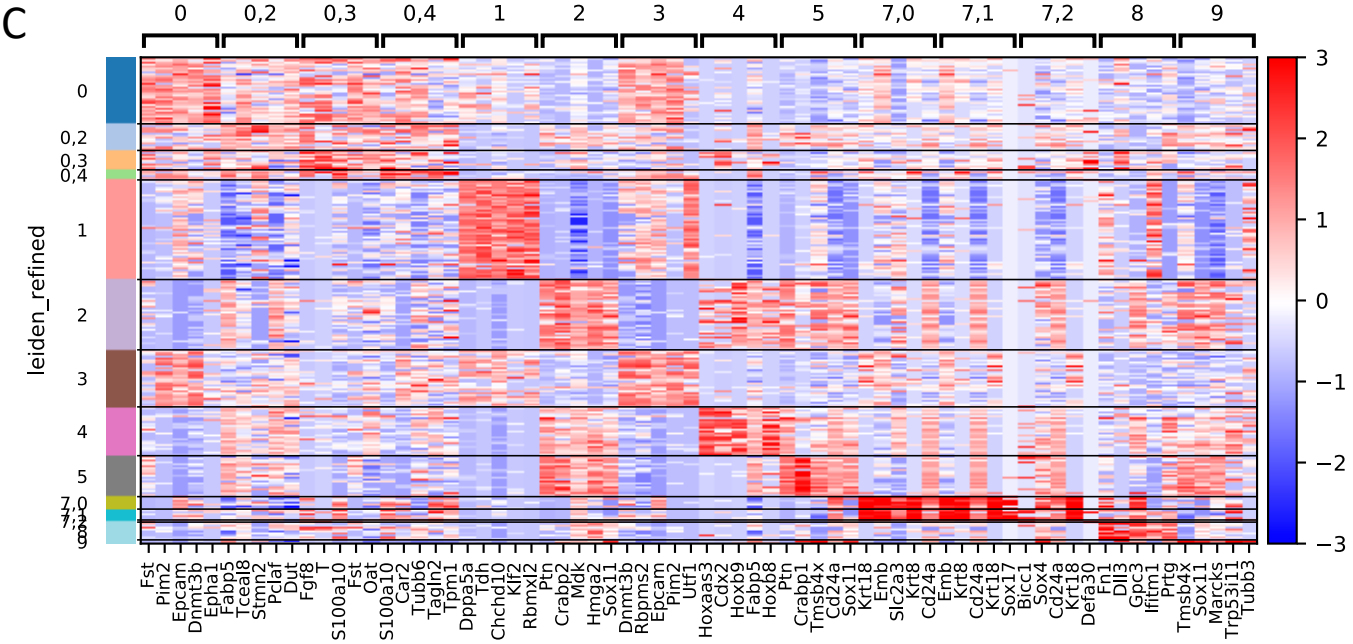
A



B

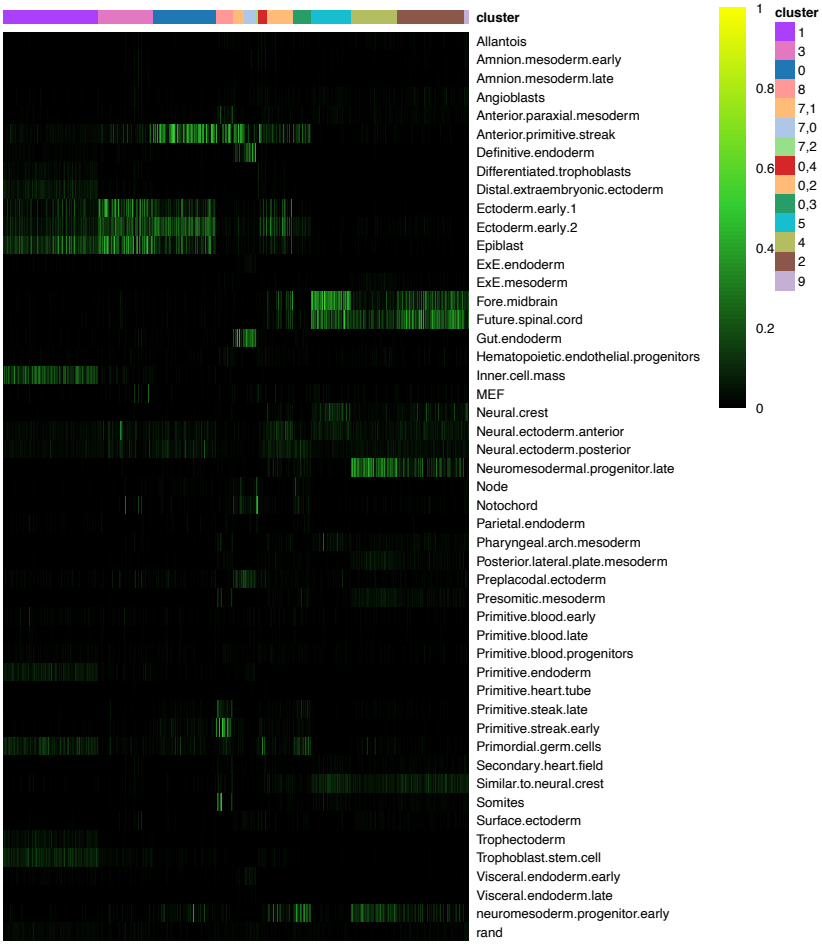


C

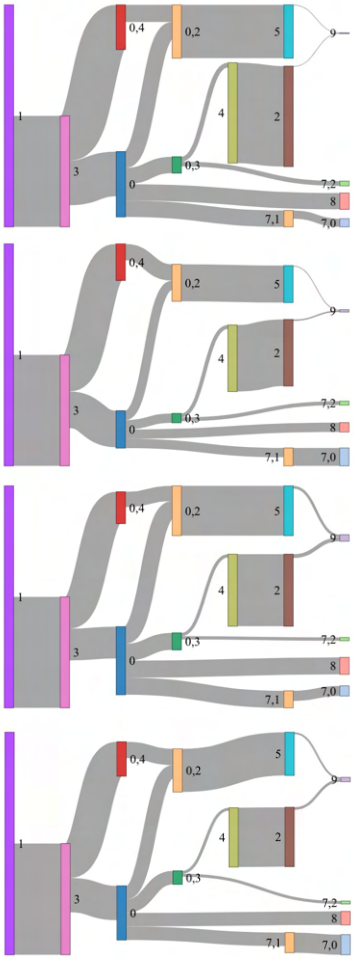


Supplementary Figure 3

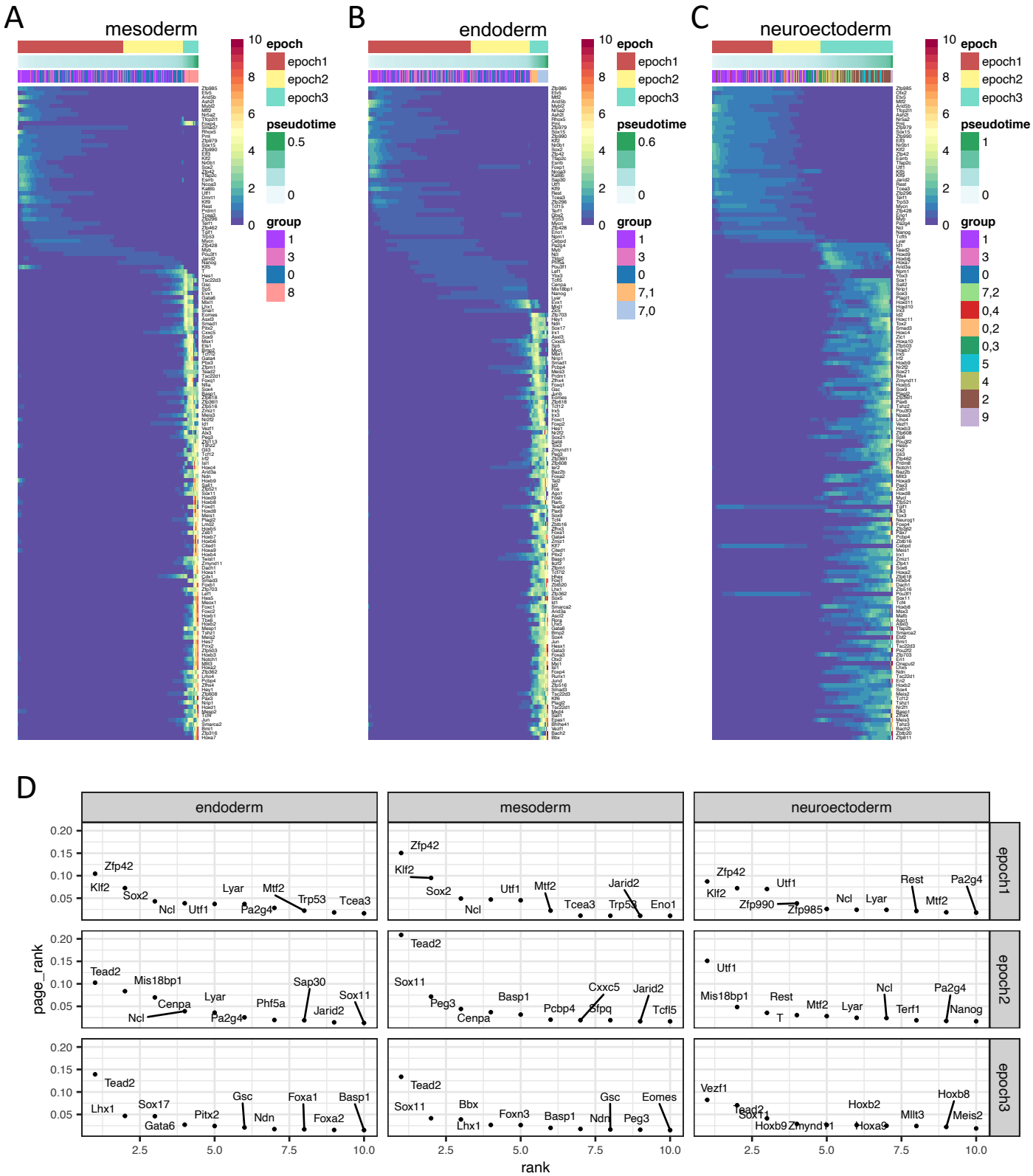
A



B

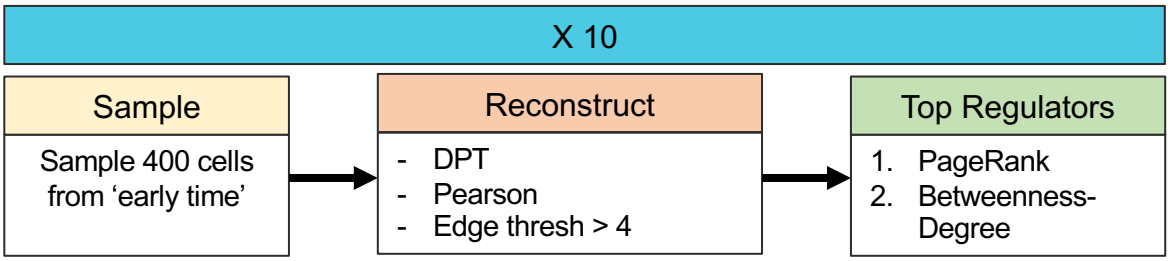


Supplementary Figure 4

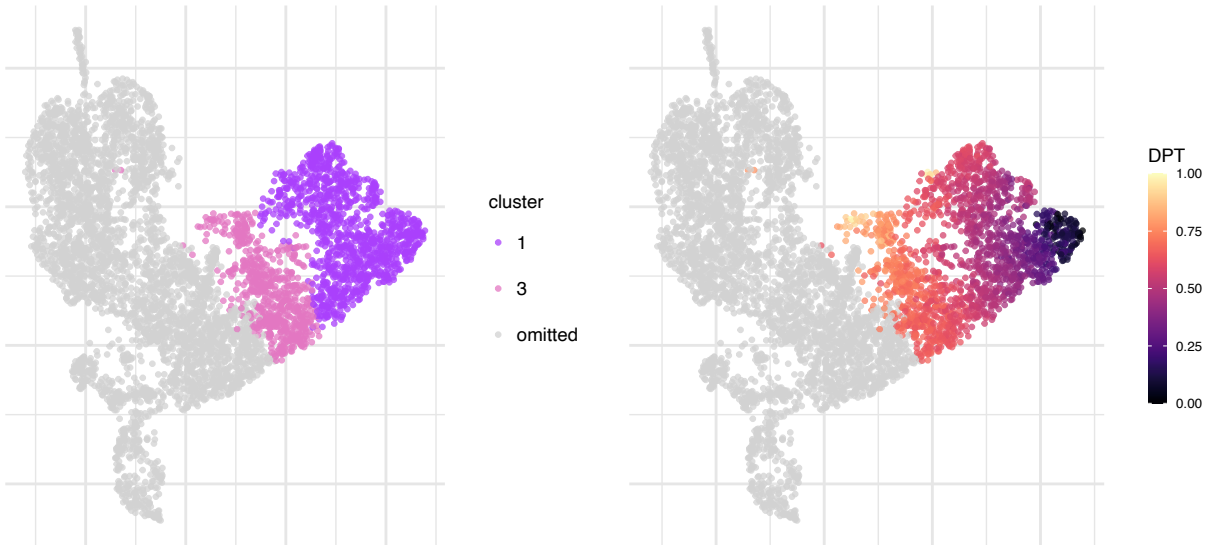


Supplementary Figure 5

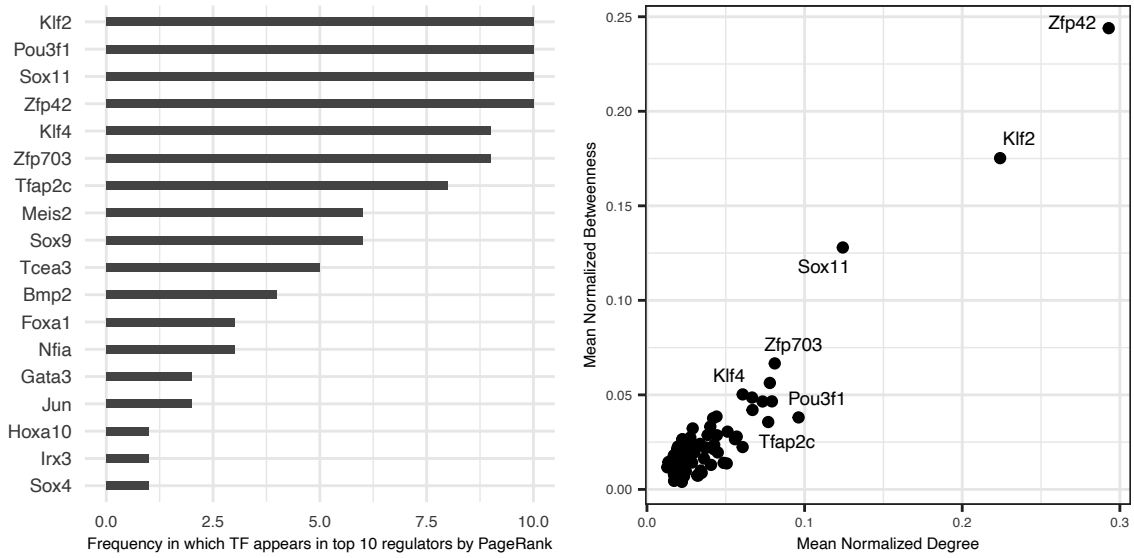
A



B

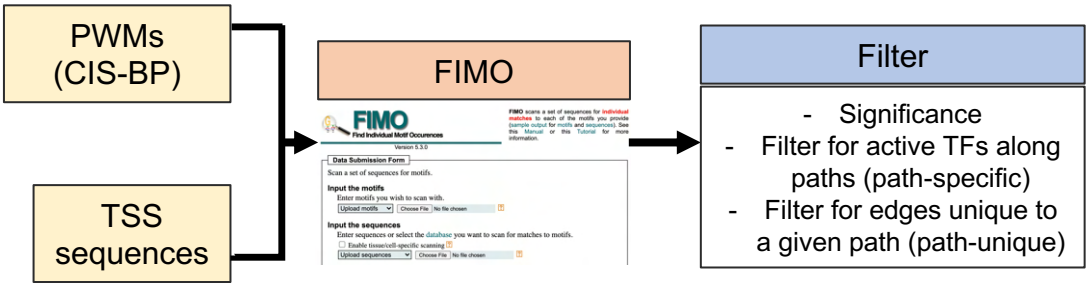


C

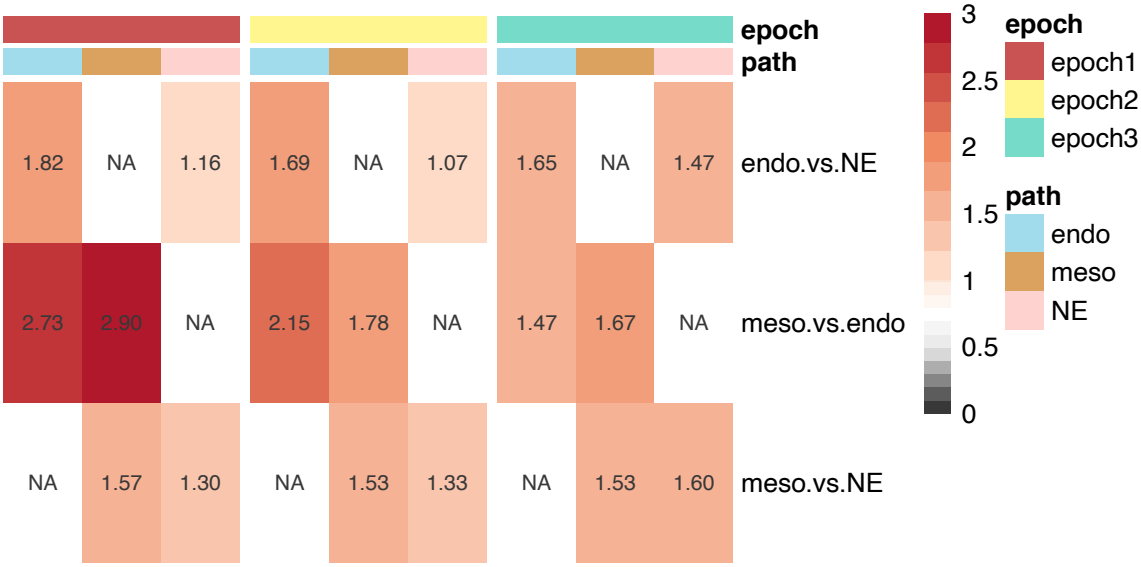


Supplementary Figure 6

A



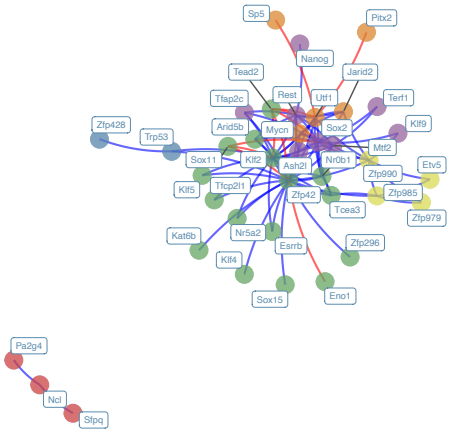
B



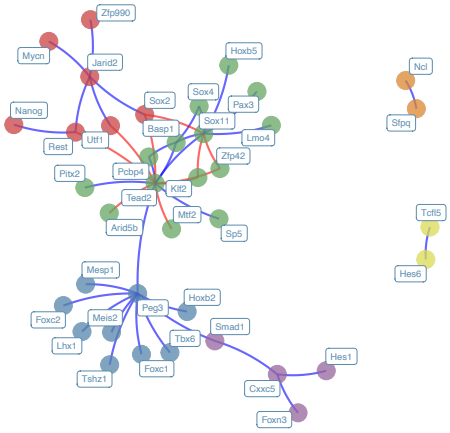
Supplementary Figure 7

Full Reconstruction

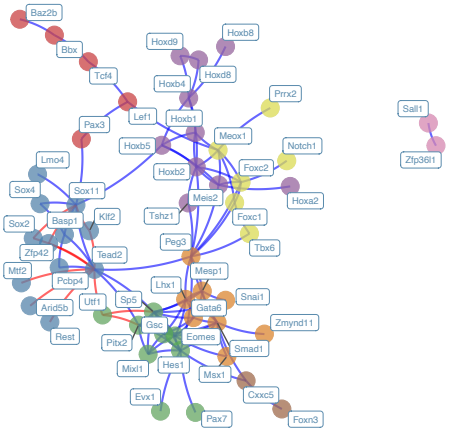
epoch1..epoch1



epoch2..epoch2

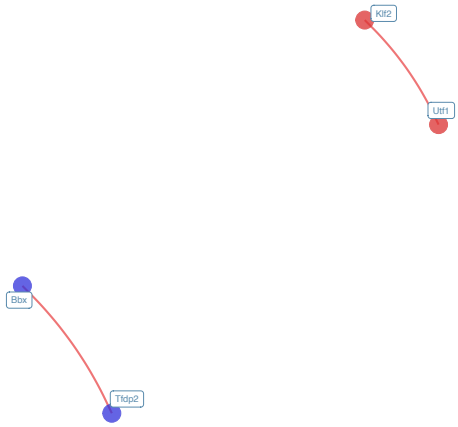


epoch3..epoch3

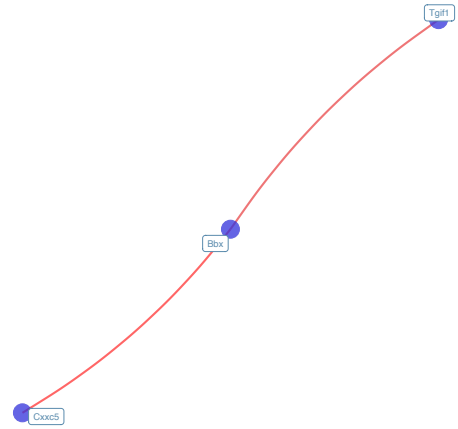


Unique to WAG (vs. WA)

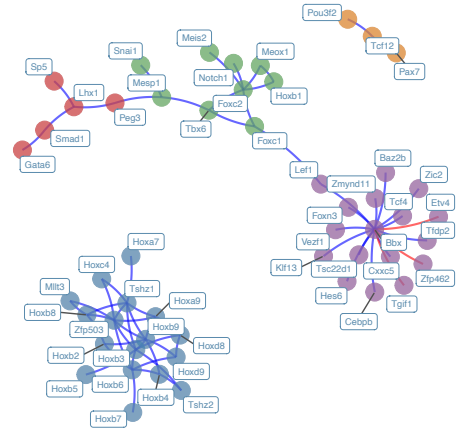
epoch1



epoch2

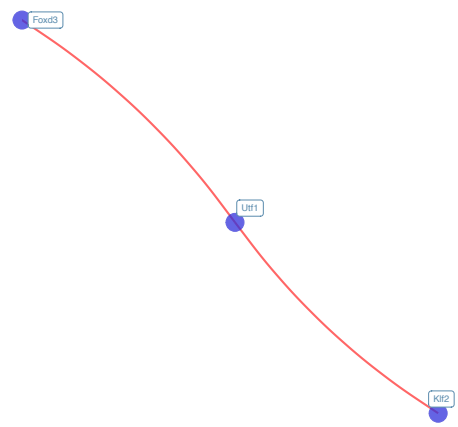


epoch3

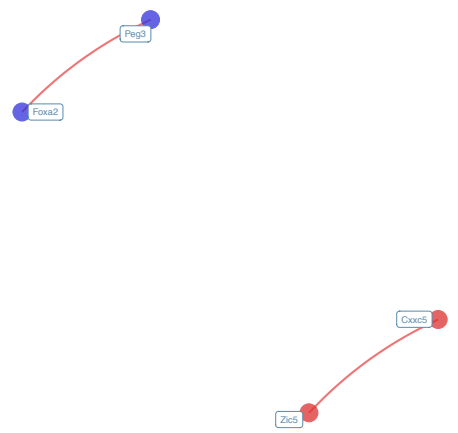


Unique to WA (vs. WAG)

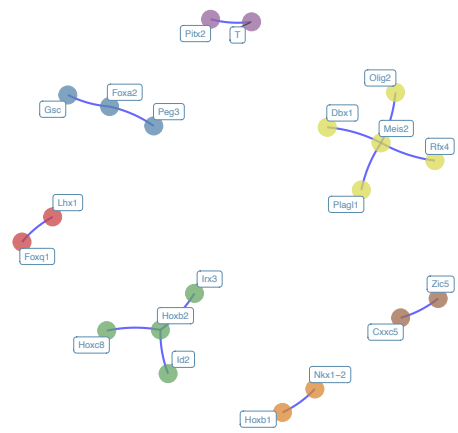
epoch1



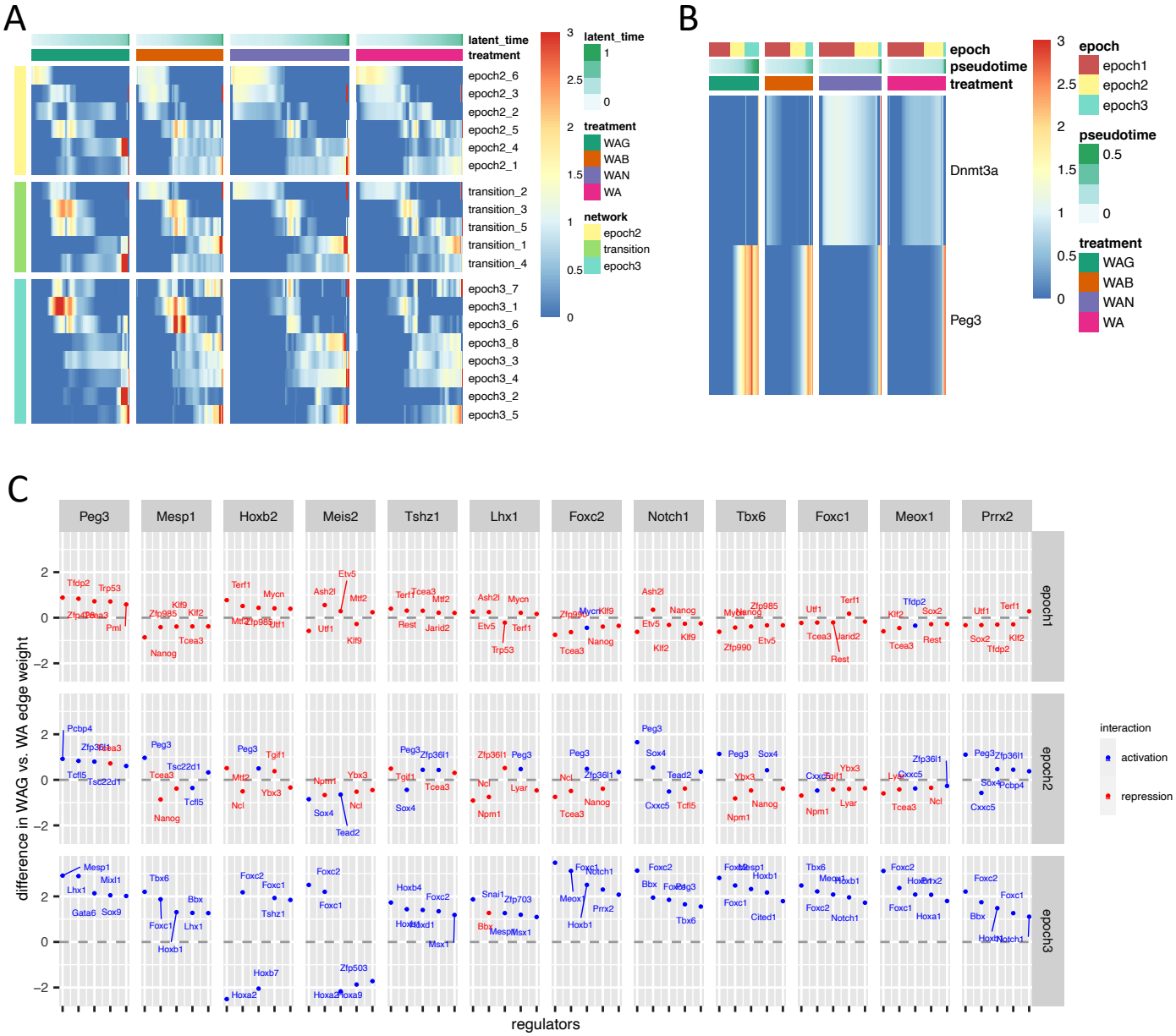
epoch2



epoch3



Supplementary Figure 8



Supplementary Table 1

Cluster	Cell Type Label	Evidence
1	Undifferentiated ESCs	This cluster classifies strongly as inner cell mass and epiblast. It expresses a number of pluripotency genes including Zfp42.
3	Epiblast	This cluster classifies as epiblast with some early ectodermal signature. Fgf5 expression begins here.
0	Primitive Streak	This cluster classifies strongly as primitive streak with some early ectodermal signature. It expresses brachyury.
8	(mesoderm) Presomitic Mesoderm/Somites	This cluster classifies as presomitic mesoderm and somites, with some residual primitive streak. It expresses a number of mesodermal markers including Mesp1.
7,1	(endoderm)	7,0 exhibits strong classification for definitive endoderm and gut endoderm. It expresses definitive endodermal markers including Foxa2 and Sox17. RNA velocity suggests 7,1 cells transition into 7,0, and largely classifies as anterior primitive streak and endoderm.
7,0		
7,2	Node	This cluster strongly classifies as node, and expresses Chordin.
0,4	(neuroectoderm) Early Ectoderm	This cluster classifies weakly as a number of ectodermal cell types, but based on RNA velocity seems to be most similar to early ectodermal cells.
0,2	(neuroectoderm) Early Ectoderm, Forebrain/ Midbrain fated	This cluster classifies weakly as a number of ectodermal cell types including early ectoderm. Based on RNA velocity, this cluster transitions into cluster 5 (forebrain/midbrain fated).
5	(neuroectoderm) Future Brain, Forebrain/ Midbrain	This cluster classifies strongly as forebrain/midbrain with some weaker future spinal cord classification. It expresses Zic1 and En2.
9	(neuroectoderm) Future brain	This cluster classifies weakly as future brain, future spinal cord, and neural crest. It expresses Crabp1 and Tubb3, and RNA velocity suggests it transitions from clusters 2 or 5.
0,3	Early Neuromesodermal Progenitor	This cluster classifies weakly as a number of cell types including primitive streak and early ectoderm, but more strongly as early neuromesodermal progenitor. RNA velocity shows cells in this cluster transition into a neuromesodermal progenitor. Cells in this cluster express T and some Sox2.
4	Neuromesodermal Progenitor	This cluster classifies strongly as neuromesodermal progenitor (late and early). It has some T and Sox2 expression.
2	(neuroectoderm) Future Spinal Cord	This cluster classifies mostly as future spinal cord, and weakly as fore/midbrain. It also expresses many patterning genes from the Hox family indicative of spinal cord fated cells such as Hoxb9.