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

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 on reconstruction of synthetically generated, *in vivo*, and *in vitro* data. To illustrate the usefulness of Epoch, we used it to identify the dynamic networks underpinning directed differentiation of mouse 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 the total existing 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. Uncovering topology within these GRNs is fundamental in answering a number of questions including: understanding how cellular identity is maintained and established, elucidating mechanisms of disease caused by dysfunctional gene regulation, and finding novel drug targets among others. 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. 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 specific regulatory relationships. However, experimental methods are generally limited in scope and scalability. For this reason, 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, often stemming from bloated false positives due to indirect interactions and a high degree of homogeneity from bulk RNA-sequencing data that does not offer enough perturbation for true relationships to be detected. These limitations ultimately result in imprecise networks (Marbach et al., 2012).

With the advent of droplet-based systems for single-cell RNA-sequencing (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 proved 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), and can be categorized similarly to their bulk counterparts into information theoretic, ensemble, Bayesian, and ODE approaches. Unfortunately, they also suffer from similar limitations, and aside from low precision and sensitivity, many of these methods are computationally burdensome, limiting users' ability to hone analysis through iterative application. Many of these methods, though they employ pseudotemporal analysis to aid in reconstruction, are limited in their ability to explain how network topology evolves over time. Further, it has been shown that the dynamic and noncommutative nature of such complex networks (e.g. via altering chromatin accessibility), 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, it's necessary for GRN models to encapsulate this dynamic behavior in order to gain a more nuanced view of how a changing GRN topology generates distinct cell fates in development or disease progression.

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 begin by undergoing embryoid body (EB) formation, which mimics *in vivo* gastrulation, and can be guided by the introduction of specific molecules to induce signaling pathway activity. However, it is unclear if the activation of these pathways influences cell fate by simply activating distinct targets, or by impacting GRN topology, thus altering the fate decision landscape and changing the capacity for, and dependence of, regulatory control over differing target states.

To this end, 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, following a protocol similar to one previously established (Spangler et al., 2018). We utilized Epoch reconstructed networks to quantify the extent of target gene differences amongst active TFs in all four treatments. We found that there were remarkably large differences between target genes of the same TF across treatments, and further applied gene set enrichment analysis to understand the function consequences of these differences. Overall, our results suggest that these topological differences may broadly alter cell fate potential.

We next applied Epoch to more specifically evaluate cell fate potential differences seen in the data, and focused on understanding 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, we used Epoch to integrate signal transduction pathways with reconstructed network topologies to demonstrate that cells treated with Wnt3a, ActivinA, and GSK inhibitor developed a network topology that allows for more effective mesodermal specification through activation of Wnt signaling. 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. 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 1**). 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 significantly reduce false positives that may result from indirect interactions or that may represent non-logical interactions, ultimately improving precision. 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 unlikely 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 essentially “time periods”, based on pseudotime, cell ordering, k-means clustering, or user-defined assignment. From here, genes are assigned to specific epochs based on their activity along pseudotime. 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 in driving or maintaining topology. Specifically, Epoch will compute and rank TFs according to their PageRank in the context of their epoch networks, and will further look for TFs that simultaneously score high in betweenness and degree centralities compared to all other TFs. Additionally, Epoch can integrate major signal transduction pathways with reconstructed networks and determine the shortest paths through the networks 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 designed in a stepwise workflow 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 benchmarked on synthetically generated data

To assess the performance of Epoch in generating static networks, we began by benchmarking its performance reconstructing networks from synthetically generated data. To this end, 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, up to 1000 cells. We benchmarked Epoch against original versions of GENIE3, which is the GRN reconstruction engine of SCENIC (Aibar et al., 2017), and CLR, 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), Epoch with mutual information (with and without cross-weighting).

We reconstructed individual networks across all methods 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 2a**). 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 2b, Supplementary Figure 1a**). 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 benchmarked 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 2c**). For this dataset, we used DPT to compute pseudotime for Epoch reconstruction. 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 2d, Supplementary Figure 1b**).

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

265 (Spangler et al., 2018). We limited our reconstruction to cells in earlier stages of differentiation
266 that had not yet transitioned into any committed germ layer state, as determined by
267 pseudotemporal analysis, RNA velocity analysis, expression of Zfp42, marking naive pluripotent
268 cells, and expression of Fgf5, marking cells in a primed state (**Supplementary Figure 1c**). The
269 filtering ensured the trajectory remained linear, which, while absolutely not a requirement for
270 Epoch reconstruction, was employed to minimize performance variability of trajectory inference
271 methods in which branched structures may be more difficult to order. This allowed us to narrow
272 our focus on reconstruction performance while minimizing impact from potentially inaccurate
273 pseudotemporal ordering. Based on precision-recall, we found that Epoch variations using
274 crossweighting and GENIE3 embedded within the Epoch framework performed best
275 (**Supplementary Figure 1d**). Overall, these benchmarking results demonstrate Epoch's utility
276 as a reconstruction tool, and further demonstrate the usefulness of Epoch as a flexible
277 framework that can improve reconstruction of other methods.
278
279 *Epoch elucidates regulatory networks driving lineage specification in early in vitro mouse ESC*
280 *differentiation*
281
282 We next wanted to apply Epoch to elucidate the underlying dynamic networks involved
283 in mesodermal fate specification in mESC directed differentiation. We collected scRNA-seq data
284 from day 0 through day 4 *in vitro* mESC directed differentiation optimized to induce primitive
285 streak formation (**Figure 3a**). Briefly, cells were allowed to differentiate in serum-free
286 differentiation media before being treated with one of four treatments on day 2: Wnt3a and
287 Activin A alone (WA), or with one of Bmp4 (WAB), GSK inhibitor (WAG), or Noggin (WAN).
288 Multiple rounds of differentiation were staggered such that we could harvest samples
289 representative of d0-d4 at one time for sequencing using the MULTI-seq protocol (McGinnis et
290 al., 2019). After barcode classification, samples were preprocessed using the SCANPY pipeline
291 (Wolf et al., 2018), and RNA velocity analysis was performed (Bergen et al., 2019; La Manno et
292 al., 2018) (**Figure 3b,c**). We found the presence of three distinct lineages, which we confirmed
293 to be representative of neuroectoderm, mesoderm, and endoderm based on gene expression
294 and further verified by singleCellNet analysis (Tan and Cahan, 2019) (**Figure 3d,**
295 **Supplementary Figure 2a-c, 3a**). A majority of the differentiating cells transitioned toward the
296 neuroectoderm (clusters labeled 0,2, 0,3, 4, 2, 5, 9) fate with smaller populations transitioning
297 toward mesoderm (cluster labeled 8) and endoderm fate (clusters labeled 7,1 and 7,0).

298 Interestingly, there were differences between treatments in terms of the distribution of
299 cells amongst the different populations (**Figure 3e,f, Supplementary Figure 3b**). Notably, cells
300 treated with WAG exhibited a stunted trajectory toward neuroectodermal fate but a fuller
301 trajectory toward mesodermal fate. This was in direct contrast to cells treated with the remaining
302 three treatments (WAB, WAN, WA), in which very few cells differentiated toward mesodermal
303 fate, but instead exhibited strong neuroectodermal or endodermal commitment. Further, while
304 WAB and WA cells tended to transition into cluster 7,0 (furthest along endodermal development
305 according to RNA velocity analysis), WAN and WAG cells were more likely to stay in cluster 7,1.
306 We then applied Epoch to this data, implementing latent time (Bergen et al., 2020) in
307 place of pseudotime, and analyzed the networks underlying each of the three lineages. Latent
308 time was divided into three epochs, and a dynamic network was extracted for each lineage
309 (**Figure 4a-c, Supplementary Figure 4**). Both top regulator prediction methods, PageRank and
310 Betweenness-Degree, had similar results (**Figure 4d, Supplementary Figure 4d**). In a more
311 nuanced analysis, we limited our scope to early latent time, corresponding to the inner cell mass
312 (ICM) to epiblast transition in an *in vivo* context, and bootstrapped the network reconstruction
313 (**Supplementary Figure 5**). In doing so, we were able to predict top regulators via a consensus
314 approach in which regulators were ranked by their frequency as a top regulator (PageRank) or
315 average importance (Betweenness-Degree) in the bootstrapped networks. The top regulators
316 recovered by both PageRank and Betweenness-Degree were largely overlapping, and included
317 TFs such as Zfp42, Klf2, Sox11, Pou3f1, and Klf4.
318
319 *Peg3 is a central regulator in mesodermal WAG networks and may drive differences in gene*
320 *module expression amongst treatments*
321
322 As noted above, cell fate potential amongst the differentiating cells appeared to be
323 dependent on treatment. We hypothesized that underlying these differences were treatment-
324 specific GRNs. To this end, we reconstructed and compared dynamic networks for each
325 treatment along the mesodermal path (which we refer to as the treatment-specific networks:
326 WAG, WAB, WAN, and WA networks), and further compared these against the full
327 reconstructed mesoderm network described in the previous section (which is reconstructed from
328 all cells regardless of treatment, and which we refer to as the mesoderm network). We then
329 aimed to understand why WAG treated cells were able to specify mesodermal fate while WAB,
330 WAN, and WA treated cells did not.

331 Because differences in state were largely apparent toward the end of latent time, we
332 performed community detection on epoch 2, epoch 3, and their transition, of the mesoderm
333 network and identified distinct TF communities, or modules, within the topology (**Figure 5a,**
334 **Supplementary Figure 7**). We assessed the activity of each module by looking at the average
335 expression of member genes (excluding genes predicted to be repressed in a given module)
336 across latent time for each treatment (**Figure 5b Supplementary Figure 8a**). Importantly, we
337 found the existence of certain modules that were strongly activated in the WAG treatment that
338 failed to activate or were weakly activated in the remaining three treatments. Three such
339 modules (one activated in each of the second epoch, third epoch, and transition) were
340 particularly interesting as they exhibited strong activity in the WAG treatment along not only the
341 mesoderm path, but also the full dataset. The TFs in these modules were Peg3, Lhx1, Hoxb2,
342 Foxc1, Tshz1, Meis2, Mesp1, Tbx6, Foxc2, Prrx2, Meox1, and Notch1. Upon extracting the
343 differential network (which we define as the network containing the interactions, or edges, that
344 have large weight differences between two treatments, and thus represent interactions most
345 likely to be unique to a treatment network; see methods) between WAG and WA, we recovered
346 the same modules and further identified the top differential regulators of these TFs themselves
347 (**Figure 5c, Supplementary Figure 8c**). Interestingly Peg3 was predicted to have a role in
348 activating many of these TFs in the WAG treatment networks, but had a notably absent role in
349 the WA treatment.
350 Peg3's apparent influence in the WAG networks suggested to us the potential existence
351 of a pathway linking GSK3 inhibition, Peg3 expression, and mesodermal gene expression.
352 Indeed, previous reports have demonstrated a role for GSK3 in maintaining DNA methylation at
353 imprinted loci, including Peg3, in mESCs (Meredith et al., 2015). Additional literature has
354 demonstrated that methylation at imprinted loci may be maintained by Dnmt3a2, which has
355 been suggested to be regulated by the GSK3-dependent transcription factor N-Myc (Popkie et
356 al., 2010). Unfortunately, our data was only able to capture Dnmt3a dynamics (representing
357 composite Dnmt3a and Dnmt3a2 expression), and not specifically Dnmt3a2, which is
358 transcribed from an alternative promoter within the Dnmt3a locus. However, since Dnmt3a2 is
359 the isoform predominant in ESCs (Chen et al., 2002), we roughly compared Dnmt3a temporal
360 dynamics and Peg3 expression between the different treatments (**Supplementary Figure 8b**).
361 Notably, Dnmt3a and Peg3 expression were largely mutually exclusive, and Dnmt3a expression
362 was low to none in the WAG treated cells. We then asked if this was a phenomenon observed in
363 *in vivo* gastrulation. In the single-cell mouse gastrulation dataset from Chan et al. (Chan et al.,
364 2019), we found relatively high expression of Dnmt3a in epiblast cells which was tapered in

365 primitive streak and mesodermal cells (**Figure 5d**). In contrast, Peg3 expression was low in
366 epiblast and increased in the primitive streak, mirroring our *in vitro* data.
367
368 *Target gene differences amongst TFs confirm network topology restructuring dependent on*
369 *treatment*
370
371 Given Peg3's apparent role in restructuring GRN topology, we next aimed to understand
372 the extent to which broad, network-wide topology changes are responsible for differences in cell
373 fate potential between treatments. To answer this, we performed two edge-level analyses
374 focusing on differences in targets of transcription factors across the four treatment networks.
375 First, we asked if transcription factors actively expressed in all four treatments regulated
376 the same set of target genes. To this end, we focused on the treatment-specific networks along
377 the mesodermal path. We narrowed our analysis to the 72 TFs that were active in all four
378 treatments during day 3 or day 4 along the mesodermal path. We quantified overlap between
379 targets of each TF in a pairwise comparison of the treatments using the Jaccard similarity
380 (**Figure 6a**). As a baseline, we implemented a bootstrapping method in which we reconstructed
381 10 networks for each treatment along the mesodermal path (using 400 sampled cells each); for
382 each TF we calculated the average Jaccard similarity of its targets amongst pairwise
383 comparisons of the reconstructed networks in a treatment. This gave us a baseline, or
384 expected, Jaccard similarity for each TF in each treatment. Overall, we found that
385 overwhelmingly, target differences between treatments were significantly greater than the
386 baselines, indicating vast network topology differences amongst the treatment networks.
387 Next, we asked if these differences in predicted targets had functional roles in impacting
388 fate potential. To this end, we focused on the 13 TFs amongst the 72 mentioned above that
389 exhibited the largest differences in predicted targets between treatments (**Figure 6b**). We
390 performed gene set enrichment analysis on the predicted targets of these TFs, and compared
391 the results between treatments (**Figure 6c**). Interestingly, we found enrichment for heart
392 development and mesoderm formation in only the WAG treatment, despite the fact that the
393 networks we analyzed were reconstructed along the mesodermal path. Of note, all treatments
394 showed enrichment for embryonic digestive tract morphogenesis. Our enrichment results thus
395 supported our hypothesis that modulating signaling results in a topological restructuring of the
396 GRN, ultimately guiding cell fate potential.
397

398 *Restrictive network topologies impede paths toward mesoderm gene activation and limit*
399 *mesodermal fate potential*
400
401 We next asked if we could trace pathways from signaling effectors to the activation of
402 meseodermal fate. To determine which signal transduction pathways were active at any given
403 point in latent time along the mesodermal lineage, we used ChIP-seq data from ChIP-Atlas (Oki
404 et al., 2018) to compile a list of binding targets for 18 signaling effector TFs (available as part of
405 the Epoch framework). We then computed average expression of these targets across latent
406 time broken down by treatment, and unsurprisingly found differences in the signaling pathways
407 that were activated in each treatment (**Figure 7a**). Of note, Notch signaling was activated in
408 WAB, WAN, and WA treatments, but not in the WAG treatment along the mesodermal lineage
409 (and weakly activated along the neuroectodermal lineage), consistent with literature
410 demonstrating that Notch signaling plays a role in specifying neuroectodermal fate in the
411 neuroectoderm-mesendoderm fate decision (Androutsellis-Theotokis et al., 2006; Aubert et al.,
412 2002).
413 Because Wnt signaling is essential for mesodermal fate, we looked for paths connecting
414 Wnt effectors to the TFs in the WAG-specific modules from the previous analysis. Specifically,
415 we computed the shortest paths from targets of these effectors to the module TFs in each
416 treatment-specific dynamic network (using edge lengths inversely proportional to the cross-
417 weighted score) (**Figure 7b**). Paths were assigned infinite length if such a path did not exist. We
418 normalized these path lengths against the average path length in a given network, and found
419 that for the WAB, WAN, and WA treatments, paths from Wnt to many of the module TFs were
420 non-existent. This was again consistent with our hypothesis that differing network topologies
421 underlie the treatment-specific differences we see in cell fate. Further, this analysis provides a
422 basis for identifying which signaling pathways must be targeted for directing certain fate
423 transitions, for example through an exhaustive search of effector targets.
424
425 *Suppression of PI3K signaling encourages endodermal fate*
426
427 Our signaling effector target analysis also revealed increased Foxo1 activity in WAB,
428 WAN, and WA treated cells as compared to WAG treated cells, indicating the suppression of
429 PI3K signaling along the WAB, WAN, and WA trajectories (**Figure 7a**). Importantly, it is known
430 that the establishment of definitive endoderm (DE) requires suppression of PI3K signaling
431 (McLean et al., 2007; Yu et al., 2015). Further, it has been recently shown in hiPSC to DE

differentiation that Foxo1 binds to DE-formation-related genes and that inhibition of Foxo1 during DE induction impedes DE establishment (Nord et al., 2020). We therefore hypothesized that a second fate choice between mesodermal and endodermal fates further exacerbated the uneven mesodermal fate preference between the treatments. Indeed, our analysis indicates that in the mesodermal-endodermal fate choice, the majority of WAG treated cells differentiate toward mesodermal fate. This is in contrast to the WAN and WA treatment in which cells preferentially differentiate toward definitive endoderm. WAB treated cells exhibited mixed potential, and split roughly equally along the two paths.

We sought to elucidate a possible explanation for Foxo1's role in this fate choice by searching for the shortest paths toward a set of endodermal genes. Specifically, similar to the previous mesodermal analysis, we applied community detection to the full endodermal network and identified the cluster containing Sox17 and Foxa2, known master regulators of endodermal fate (**Figure 7c**). This cluster included the transcription factors Sox17, Foxa2, Bmp2, Cited1, Foxa1, Gata6, Hhex, Lhx1, and Tcf7l2. For each treatment-specific network, we computed the shortest paths from Foxo1 targets (as determined in the same manner as the previous section) to these genes (**Figure 7d**). Importantly, we found that many of these regulators, including both Sox17 and Foxa2, were not reachable from Foxo1 in the WAG network, consistent with the observation that WAG treated cells preferentially differentiated toward mesodermal fate over endodermal fate. Interestingly, Foxa2 was not reachable in the WAN network, though Sox17 was. We believe this may explain the differences in distribution of cells between the endodermal clusters 7,1 and 7,0. Namely, of the cells that specify endodermal fate, WA and WAB treated cells more fully transitioned into cluster 7,0. In contrast, WAN treated cells remained mostly in cluster 7,1 and were less likely to differentiate further into 7,0. Previous studies in directed differentiation of human ESCs toward DE have implicated Sox17's regulatory role in establishing DE to be upstream of Foxa2, the loss of which impairs foregut and subsequent hepatic endoderm differentiation (Genga et al., 2019). This is consistent with our cell type annotations of cluster 7,1 (definitive endoderm) and 7,0 (gut endoderm), and offers a possible mechanism of the discrepancy in cell fate preferences between treatments.

Discussion

Methods of inferring regulatory network structures are essential for understanding cell identity transitions, uncovering mechanisms of disease progression due to dysfunctional gene regulation, and answering other regulatory questions in biology. Unfortunately, experimental methods are unscalable and existing computational methods suffer from poor performance or high computational burden. Here, we present a novel GRN reconstruction tool, Epoch, which

466 leverages single-cell transcriptomics and efficiently infers dynamic network structures. We
467 demonstrate that Epoch out performs commonly used methods CLR and GENIE3 in both
468 synthetically generated and real-world datasets. Furthermore, Epoch's relative efficiency allows
469 it to be easily applied iteratively for applications requiring fine-tuning of network topology or
470 iterative simulations of changes in network structure.

471 Further, the Epoch framework is flexible. There are no strict requirements on
472 pseudotemporal input as we demonstrate here using synthetic simulation time, diffusion
473 pseudotime, and RNA velocity-based latent time. We further validate on the in vivo muscle
474 development dataset that precision-recall is largely comparable for different pseudotime
475 methods (**Supplementary Figure 1f,g**). Additionally, Epoch's workflow is structured in discrete
476 steps allowing users to pick-and-choose as well as substitute portions of the workflow. We
477 anticipate that this flexibility will also be useful as advances in sequencing technology, such as
478 multi-modal methods, and analysis techniques, such as novel cell ordering or reconstruction
479 methods, can be easily adapted into Epoch's framework. Importantly, because of single-cell
480 transcriptomic data's inherent noisiness, we also asked if smoothing gene expression over time
481 could alleviate noise-related drops in reconstruction performance. We compared performance
482 between unsmoothed and kernel-smoothed data and found that the there was no significant
483 difference (**Supplementary Figure 1e**). However, smoothing functions remain in Epoch's code
484 base to aid in cleaner visualizations of gene dynamics.

485 Finally, we applied Epoch to scRNA-seq data from day 0 through day 4 *in vitro* mESC
486 directed differentiation to uncover underlying dynamic networks driving mesodermal lineage
487 specification. Our data consisted of cells transitioning into all three germ layers, and Epoch was
488 able to recover a number of key lineage drivers. Because our data consists of cells undergoing
489 four different treatments (WAG, WAB, WAN, WA), we were additionally able to reconstruct
490 treatment-specific GRNs. In direct contrast to the cells in the WAB, WAN, and WA treatment
491 groups, the cells in the WAG treatment group had higher mesodermal fate potential and
492 exhibited a relatively stunted neuroectodermal trajectory. To understand this observation we
493 applied a number of network comparison strategies to elucidate differences in network topology
494 between treatments responsible for this fate choice. Importantly, we found three sets of
495 topological differences that may play a role in specifying mesodermal fate.

496 First, we identified a set of TF modules that were preferentially activated in WAG treated
497 cells compared to cells undergoing the other three treatments along the mesodermal fate.
498 Specifically, these modules encompassed the transcription factors Peg3, Lhx1, Hoxb2, Foxc1,
499 Tshz1, Meis2, Mesp1, Tbx6, Foxc2, Prrx2, Meox1, and Notch1. Upon further analysis, we

500 observed that the transcription factor Peg3 was predicted by Epoch to have a role in regulating
501 a large fraction of these mesodermal TFs. This influence was specific to WAG treated-cells, and
502 indeed, prior studies have implicated a GSK3-dependent pathway for the epigenetic regulation
503 of imprinted loci including Peg3 (Meredith et al., 2015). Peg3 expression has been detected in
504 early somites and other mesodermal tissues (Kuroiwa et al., 1996). Interestingly, it has been
505 separately demonstrated that knockdown of Peg3 in mouse embryonic fibroblasts and neural
506 stem cells increases expression of pluripotency genes and enhances overall reprogramming
507 efficiency (Theka et al., 2017). Taken together these results suggest a role for Peg3 in
508 modulating the exit of pluripotency and mesodermal lineage specification. If so, this finding
509 could paint Peg3 as a highly enticing transcription factor to target with the goal of improving
510 efficiency of directed differentiation protocols toward certain mesodermal fates like somites.
511 Next, we compared topological differences in the treatment-specific networks by
512 examining the existence of, and computing the length of, shortest paths. Specifically, we sought
513 to trace paths from signaling effectors to the activation of mesodermal genes within these
514 dynamic networks, thereby providing possible mechanisms of how genetic programs are
515 activated from signal transduction pathways. A number of studies have pointed to the influence
516 of “Wntch” signaling on the neuroectoderm-mesendoderm cell fate decision in gastrulation, as
517 reviewed by (Hayward et al., 2008). Our data was consistent with these reports: WAB, WAN,
518 WA treated cells exhibited activated Notch pathways and exhibited full trajectories toward
519 neuroectoderm fate; WAG treated cells exhibited markedly diminished or non-existent Notch
520 signaling and reached mesodermal fates while seemingly displaying a stunted capacity for
521 neuroectodermal fate.
522 Our analysis indicated the existence of paths from Wnt signaling effectors to the
523 aforementioned mesodermal TFs within the WAG network. Analogous paths were markedly
524 sparser in WAB, WAN, and WA networks, and some mesodermal TFs were entirely
525 unreachable. This suggests topological network differences caused by treatment, though the
526 nature of the emergence of these differences remains unclear. One possibility is that the
527 network structures are dynamically altered by epigenetic events initiated through the various
528 signaling pathways activated in treating the cells. We point to GSK3’s reported effect over Peg3
529 methylation as an example of such a phenomenon. Further studies to elucidate differences in
530 chromatin accessibility between mESCs undergoing directed differentiation by way of distinct
531 treatments, or activation of distinct signaling pathways, will aid in clarifying if such a mechanism
532 is at play.

Finally, we focused our analysis on cells specifying mesendoderm fate, and outlined topological differences that could lead to differences in mesoderm vs. endoderm fate potential in cells undergoing different treatments. We applied an analogous shortest path analysis along the endodermal fated path to identify the link between suppression of PI3K signaling and endoderm specification. We again found topological differences between WAG and the remaining three treatments, this time indicating that many endodermal TFs, such as essential endodermal regulators Sox17 and Foxa2, were unreachable from PI3K suppression in the WAG network, explaining the preference of these cells to transition into mesodermal rather than endodermal fates. Interestingly, we found that Foxa2 was not reachable from PI3K suppression in the WAN network, which was consistent with our finding that of the WAN treated cells that specified endodermal fate, the majority tended to remain in earlier endodermal stages based on RNA velocity. This was contrast to WAB and WA treated cells which tended to transition more fully into a later endodermal stage. Taken together, these results offer explanation of differences in mesodermal vs. endodermal fate potential between the treatments.

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 and the Python module PyEpoch, 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 thank Chris McGinnis and the Gartner Lab for their help implementing the MULTI-seq protocol and for providing the required reagents.

567 **Methods**

568

569 **mESC maintenance and differentiation**

570

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

578

579 **Multi-Seq protocol**

580

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

591

592 **Library preparation and sequencing**

593

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

598

599 **Single-cell data processing**

600

601 Sequencing alignment to the mm10 reference genome was performed using CellRanger
602 (version 3.1), and spliced, unspliced, and ambiguous read counts were extracted using
603 Velocyto. Sample barcode classification was performed using the deMULTIplex R package
604 (https://github.com/chris-mcginnis-ucsf/MULTI-seq). Subsequent quality control filtering,
605 normalization, clustering, and differential gene expression analysis was performed using
606 Scanpy (version 1.4.5) (Wolf et al., 2018). Specifically, doublets and negatives from the sample
607 barcode classification were removed, along with 271 cells we identified as residual fibroblasts
608 (Thy1+). Next, genes were excluded if they were detected in less than 5 cells; cells were
609 excluded if their mitochondrial gene content exceeded 10% of their total reads or if they had
610 fewer than 500 unique genes. This left us with 5530 cells and 17528 genes. The data was then
611 normalized and log transformed before highly variable genes were identified (1938 genes). The
612 data was scaled and PCA was performed. Leiden clustering was performed and visualized on a
613 UMAP embedding. We further removed a cluster of possible dying cells expressing relatively
614 high mitochondrial gene content and low total read count, leaving a dataset with 5228 cells and
615 1938 variable genes. Leiden clustering was refined, and RNA velocity was computed using
616 scvelo (version 0.2.2) (Bergen et al., 2020). Latent time was computed with scvelo, using root
617 and end states that were computed through CellRank (version 1.0.0-rc.0). Cell type
618 classification was performed via SingleCellNet (Tan and Cahan, 2019), using complied embryo
619 data for training the classifier. For the E12.5 muscle development data and day 4 early directed
620 differentiation data, diffusion maps and diffusion pseudotime were computed through Scanpy.
621
622 **Synthetic benchmarking**
623
624 15 synthetic datasets were generated using the Dyngen package in R (Cannoodt et al., 2020).
625 Specifically networks were generated with 100-420 genes (including 20-70 TFs), containing
626 varied topological motifs, and cell trajectories were simulated over time. Simulated single-cell
627 sequencing was performed, resulting in datasets that included up to 1000 cells. CLR was
628 implemented with the minet package in R (Meyer et al., 2008), and GENIE3 was implemented
629 with the GENIE3 package in R (Aibar et al., 2017; Huynh-Thu et al., 2010). Variations of these
630 two methods were carried out using the Epoch framework. Performance (AUPR) was compared
631 against sets of 5 random networks generated for each reconstructed network via the random
632 permutation of edge weights across targets.
633
634 **Bootstrapped network reconstruction of early MULTI-seq data**

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

Differential network analysis and assessment

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

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

Transcription star site (TSS) annotations were downloaded from the USCS Genome Browser (mouse GRCm38/mm10 genome assembly) (Haeussler et al., 2019). Promoter regions were defined as -750 to +250 around the TSSs. We then acquired position weight matrices (PWMs) of transcription factors from the Catalog of Inferred Sequence Binding Preferences (CIS-BP

669 Database build version 2.00) (Weirauch et al., 2014). We used the Find Individual Motif
670 Occurrences tool (Grant et al., 2011) to obtain potential TF-promoter binding pairs for 648 TFs.
671 We kept all interactions for which p-value < 5 x10^{-5}. To generate path-unique gold standards, for
672 each lineage (mesoderm, endoderm, neuroectoderm) we filtered interactions for dynamically
673 expressed genes along each path as computed by Epoch. Further, when assessing the
674 "differential network" pairwise comparisons, we filtered out interactions that were redundant
675 between the gold standards, and only kept interactions unique to each path.
676
677 **TF-target level comparison**
678
679 To quantify the extent of topology differences between treatments along the mesodermal path,
680 we performed pairwise comparison on the four treatment-specific networks and computed
681 Jaccard similarity between predicted targets of TFs. The comparison was carried out on 72 TFs
682 that were selected based on their active expression on day 3 or day 4 (i.e. after induction
683 treatments). We computed baseline similarities for each TF in each treatment by bootstrapping
684 the network reconstruction (10 network reconstructions per treatment, 400 sampled cells per
685 network reconstruction) and averaging their pairwise Jaccard similarities. We selected the 13
686 TFs exhibiting the greatest differences in targets amongst the treatments and performed gene
687 set enrichment analysis using the Enrichr R package (Chen et al., 2013; Kuleshov et al., 2016)
688 and the GO Consortium Biological Processes database (Ashburner et al., 2000; Gene Ontology
689 Consortium, 2021) on the targets of each TF. We further filtered out terms that did not meet the
690 criteria of adjusted p-value<0.05. Finally, we summarized the results between the 13 TFs by
691 counting the frequency (out of 13) that a term was considered enriched and ranked the results
692 accordingly.
693
694 **Comprehensive signaling effector TF targets list**
695
696 We acquired binding score (MACS2) data for 18 signaling effector TFs from the ChIP-Atlas (Oki
697 et al., 2018), and processed each accordingly: target genes were ranked by maximum binding
698 score and the top 2000 targets were retained (or all retained, if less than 2000 targets). To
699 compute average effector activity, target lists were filtered for those appearing in the MULTI-seq
700 directed differentiation data, and not considered in the analysis if the effector had less than 10
701 targets after filtering.
702

Epoch Workflow

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

Identification of dynamically expressed genes

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

CLR

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

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

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, but can be further specified by the user. In both top regulator prediction methods Epoch will

return ranked lists of nodes, and further specify their corresponding PageRank or normalized betweenness, normalized degree, and betweenness-degree product.

Acknowledgements

We would like to thank Yuqi Tan, Ray Cheng, and David Johanson for helpful discussions and testing Epoch code. 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.

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., van den Oord, J., Atak, Z.K., Wouters, J., Aerts, S., 2017. SCENIC: single-cell regulatory network inference and clustering. Nat. Methods 14, 1083–1086. doi:10.1038/nmeth.4463

Androutsellis-Theotokis, A., Leker, R.R., Soldner, F., Hoeppner, D.J., Ravin, R., Poser, S.W., Rueger, M.A., Bae, S.-K., Kittappa, R., McKay, R.D.G., 2006. Notch signalling regulates stem cell numbers in vitro and in vivo. Nature 442, 823–826. doi:10.1038/nature04940

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., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29. doi:10.1038/75556

Aubert, J., Dunstan, H., Chambers, I., Smith, A., 2002. Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. Nat. Biotechnol. 20, 1240–1245. doi:10.1038/nbt763

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., Theis, F.J., 2019. Generalizing RNA velocity to transient cell states through dynamical modeling. BioRxiv. doi:10.1101/820936

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., Theis, F.J., 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. 38, 1408–1414. doi:10.1038/s41587-020-0591-3

Cannoodt, R., Saelens, W., Deconinck, L., Saeys, Y., 2020. dyngen: a multi-modal simulator for spearheading new single-cell omics analyses. BioRxiv. doi:10.1101/2020.02.06.936971

Chan, M.M., Smith, Z.D., Grosswendt, S., Kretzmer, H., Norman, T.M., Adamson, B., Jost, M., Quinn, J.J., Yang, D., Jones, M.G., Khodaverdian, A., Yosef, N., Meissner, A., Weissman, J.S., 2019. Molecular recording of mammalian embryogenesis. Nature 570, 77–82. doi:10.1038/s41586-019-1184-5

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128. doi:10.1186/1471-2105-14-128

Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F., Poidinger, M., 2016. Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. Nat. Commun. 7, 11988. doi:10.1038/ncomms11988

Chen, T., Ueda, Y., Xie, S., 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. doi:10.1074/jbc.M205312200

di Bernardo, D., Thompson, M.J., Gardner, T.S., Chobot, S.E., Eastwood, E.L., Wojtovich, A.P., Elliott, S.J., Schaus, S.E., Collins, J.J., 2005. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. Nat. Biotechnol. 23, 377–383. doi:10.1038/nbt1075

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., Gardner, T.S., 2007. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. 5, e8. doi:10.1371/journal.pbio.0050008

Gadue, P., Huber, T.L., Paddison, P.J., 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. doi:10.1073/pnas.0603916103

Gene Ontology Consortium, 2021. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, D325–D334. doi:10.1093/nar/gkaa1113

866 Genga, R.M.J., Kernfeld, E.M., Parsi, K.M., Parsons, T.J., Ziller, M.J., Maehr, R., 2019. Single-
867 Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early
868 Human Endoderm Development. Cell Rep. 27, 708–718.e10.
869 doi:10.1016/j.celrep.2019.03.076
870 Grant, C.E., Bailey, T.L., Noble, W.S., 2011. FIMO: scanning for occurrences of a given motif.
871 Bioinformatics 27, 1017–1018. doi:10.1093/bioinformatics/btr064
872 Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M.,
873 Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., Gibson, D., Diekhans, M., Clawson, H.,
874 Casper, J., Barber, G.P., Haussler, D., Kuhn, R.M., Kent, W.J., 2019. The UCSC
875 Genome Browser database: 2019 update. Nucleic Acids Res. 47, D853–D858.
876 doi:10.1093/nar/gky1095
877 Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., Theis, F.J., 2016. Diffusion pseudotime
878 robustly reconstructs lineage branching. Nat. Methods 13, 845–848.
879 doi:10.1038/nmeth.3971
880 Hartemink, A.J., 2005. Reverse engineering gene regulatory networks. Nat. Biotechnol. 23,
881 554–555. doi:10.1038/nbt0505-554
882 Hayward, P., Kalmar, T., Arias, A.M., 2008. Wnt/Notch signalling and information processing
883 during development. Development 135, 411–424. doi:10.1242/dev.000505
884 Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., Geurts, P., 2010. Inferring regulatory networks from
885 expression data using tree-based methods. PLoS One 5.
886 doi:10.1371/journal.pone.0012776
887 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S.,
888 Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D.,
889 Gundersen, G.W., Ma'ayan, A., 2016. Enrichr: a comprehensive gene set enrichment
890 analysis web server 2016 update. Nucleic Acids Res. 44, W90-7.
891 doi:10.1093/nar/gkw377
892 Kuroiwa, Y., Kaneko-Ishino, T., Kagitani, F., Kohda, T., Li, L.L., Tada, M., Suzuki, R.,
893 Yokoyama, M., Shiroishi, T., Wakana, S., Barton, S.C., Ishino, F., Surani, M.A., 1996.
894 Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. Nat.
895 Genet. 12, 186–190. doi:10.1038/ng0296-186
896 La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber,
897 K., Kastriti, M.E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L.E., Liu, Z., van Bruggen,
898 D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., Cramer, P.,
899 Adameyko, I., Linnarsson, S., Kharchenko, P.V., 2018. RNA velocity of single cells.
900 Nature 560, 494–498. doi:10.1038/s41586-018-0414-6
901 Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., Ma'ayan, A., 2010. ChEA:
902 transcription factor regulation inferred from integrating genome-wide ChIP-X
903 experiments. Bioinformatics 26, 2438–2444. doi:10.1093/bioinformatics/btq466
904 Letsou, W., Cai, L., 2016. Noncommutative biology: sequential regulation of complex networks.
905 PLoS Comput. Biol. 12, e1005089. doi:10.1371/journal.pcbi.1005089
906 Lu, T., Mar, J.C., 2020. Investigating transcriptome-wide sex dimorphism by multi-level analysis
907 of single-cell RNA sequencing data in ten mouse cell types. Biol. Sex Differ. 11, 61.
908 doi:10.1186/s13293-020-00335-2
909 Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R.,
910 DREAM5 Consortium, Kellis, M., Collins, J.J., Stolovitzky, G., 2012. Wisdom of crowds
911 for robust gene network inference. Nat. Methods 9, 796–804. doi:10.1038/nmeth.2016
912 Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R.,
913 Califano, A., 2006. ARACNE: an algorithm for the reconstruction of gene regulatory
914 networks in a mammalian cellular context. BMC Bioinformatics 7 Suppl 1, S7.
915 doi:10.1186/1471-2105-7-S1-S7
916 Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N., Hayashi, T.,

917 Nikaido, I., 2017. SCODE: an efficient regulatory network inference algorithm from
918 single-cell RNA-Seq during differentiation. Bioinformatics 33, 2314–2321.
919 doi:10.1093/bioinformatics/btx194
920 McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L.,
921 Murrow, L.M., Weissman, J.S., Werb, Z., Chow, E.D., Gartner, Z.J., 2019. MULTI-seq:
922 sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat.
923 Methods 16, 619–626. doi:10.1038/s41592-019-0433-8
924 McLean, A.B., D'Amour, K.A., Jones, K.L., Krishnamoorthy, M., Kulik, M.J., Reynolds, D.M.,
925 Sheppard, A.M., Liu, H., Xu, Y., Baetge, E.E., Dalton, S., 2007. Activin a efficiently
926 specifies definitive endoderm from human embryonic stem cells only when
927 phosphatidylinositol 3-kinase signaling is suppressed. Stem Cells 25, 29–38.
928 doi:10.1634/stemcells.2006-0219
929 Meredith, G.D., D'Ippolito, A., Dudas, M., Zeidner, L.C., Hostetter, L., Faulds, K., Arnold, T.H.,
930 Popkie, A.P., Doble, B.W., Marnellos, G., Adams, C., Wang, Y., Phiel, C.J., 2015.
931 Glycogen synthase kinase-3 (Gsk-3) plays a fundamental role in maintaining DNA
932 methylation at imprinted loci in mouse embryonic stem cells. Mol. Biol. Cell 26, 2139–
933 2150. doi:10.1091/mbc.E15-01-0013
934 Meyer, P.E., Kontos, K., Lafitte, F., Bontempi, G., 2007. Information-theoretic inference of large
935 transcriptional regulatory networks. EURASIP J. Bioinform. Syst. Biol. 79879.
936 doi:10.1155/2007/79879
937 Meyer, P.E., Lafitte, F., Bontempi, G., 2008. minet: A R/Bioconductor package for inferring large
938 transcriptional networks using mutual information. BMC Bioinformatics 9, 461.
939 doi:10.1186/1471-2105-9-461
940 Nord, J., Schill, D., Pulakanti, K., Rao, S., Cirillo, L.A., 2020. The transcription factor FoxO1 is
941 required for the establishment of the human definitive endoderm. BioRxiv.
942 doi:10.1101/2020.12.22.423976
943 Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R.,
944 Sese, J., Meno, C., 2018. ChIP-Atlas: a data-mining suite powered by full integration of
945 public ChIP-seq data. EMBO Rep. 19. doi:10.15252/embr.201846255
946 Popkie, A.P., Zeidner, L.C., Albrecht, A.M., D'Ippolito, A., Eckardt, S., Newsom, D.E., Groden,
947 J., Doble, B.W., Aronow, B., McLaughlin, K.J., White, P., Phiel, C.J., 2010.
948 Phosphatidylinositol 3-kinase (PI3K) signaling via glycogen synthase kinase-3 (Gsk-3)
949 regulates DNA methylation of imprinted loci. J. Biol. Chem. 285, 41337–41347.
950 doi:10.1074/jbc.M110.170704
951 Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Bruggner, R.V., Linderman, M.D., Sachs, K.,
952 Nolan, G.P., Plevritis, S.K., 2011. Extracting a cellular hierarchy from high-dimensional
953 cytometry data with SPADE. Nat. Biotechnol. 29, 886–891. doi:10.1038/nbt.1991
954 Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T., McFaline-Figueroa, J.L.,
955 Saunders, L., Trapnell, C., Kannan, S., 2020. Inferring Causal Gene Regulatory
956 Networks from Coupled Single-Cell Expression Dynamics Using Scribe. Cell Syst. 10,
957 265–274.e11. doi:10.1016/j.cels.2020.02.003
958 Spangler, A., Su, E.Y., Craft, A.M., Cahan, P., 2018. A single cell transcriptional portrait of
959 embryoid body differentiation and comparison to progenitors of the developing embryo.
960 Stem Cell Res. 31, 201–215. doi:10.1016/j.scr.2018.07.022
961 Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., Dudoit, S., 2018.
962 Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC
963 Genomics 19, 477. doi:10.1186/s12864-018-4772-0
964 Tan, Y., Cahan, P., 2019. SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq
965 Data Across Platforms and Across Species. Cell Syst. 9, 207–213.e2.
966 doi:10.1016/j.cels.2019.06.004
967 Theka, I., Sottile, F., Aulicino, F., Garcia, A.C., Cosma, M.P., 2017. Reduced expression of

Paternally Expressed Gene-3 enhances somatic cell reprogramming through mitochondrial activity perturbation. Sci. Rep. 7, 9705. doi:10.1038/s41598-017-10016-7

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., Rinn, J.L., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386. doi:10.1038/nbt.2859

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., Quake, S.R., 2014. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 509, 371–375. doi:10.1038/nature13173

Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., Boch, T., Hofmann, W.-K., Ho, A.D., Huber, W., Trumpp, A., Essers, M.A.G., Steinmetz, L.M., 2017. Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. 19, 271–281. doi:10.1038/ncb3493

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., Jardine, L., Dixon, D., Stephenson, E., Nilsson, E., Grundberg, I., McDonald, D., Filby, A., Li, W., De Jager, P.L., Rozenblatt-Rosen, O., Lane, A.A., Haniffa, M., Regev, A., Hacohen, N., 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356. doi:10.1126/science.aah4573

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., Zheng, H., Goity, A., van Bakel, H., Lozano, J.-C., Galli, M., Lewsey, M.G., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J.S., Govindarajan, S., Shaulsky, G., Walhout, A.J.M., Bouget, F.-Y., Ratsch, G., Larrondo, L.F., Ecker, J.R., Hughes, T.R., 2014. Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158, 1431–1443. doi:10.1016/j.cell.2014.08.009

Wolf, F.A., Angerer, P., Theis, F.J., 2018. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15. doi:10.1186/s13059-017-1382-0

Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., Jarvis, E.D., 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics 20, 3594–3603. doi:10.1093/bioinformatics/bth448

Yu, J.S.L., Ramasamy, T.S., Murphy, N., Holt, M.K., Czapiewski, R., Wei, S.-K., Cui, W., 2015. PI3K/mTORC2 regulates TGF-β/Activin signalling by modulating Smad2/3 activity via linker phosphorylation. Nat. Commun. 6, 7212. doi:10.1038/ncomms8212

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., Gregory, M.T., Shuga, J., Montesclaros, L., Underwood, J.G., Masquelier, D.A., Nishimura, S.Y., Schnall-Levin, M., Wyatt, P.W., Hindson, C.M., Bharadwaj, R., Wong, A., Ness, K.D., Beppu, L.W., Deeg, H.J., McFarland, C., Loeb, K.R., Valente, W.J., Ericson, N.G., Stevens, E.A., Radich, J.P., Mikkelsen, T.S., Hindson, B.J., Bielas, J.H., 2017. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049. doi:10.1038/ncomms14049

Figure legends

Figure 1. Epoch workflow. Epoch relies on a three-step process to reconstruct dynamic network structures: (1) extraction of dynamic genes and CLR, (2) application of crossweighting, and (3) extraction of the dynamic network.

Figure 2. Benchmarking Epoch. (A) Synthetic dataset generation, reconstruction, and assessment. (B) Fold improvement in AUPR over random of three versions of Epoch (Pearson, MI, and MI with crossweighting) against CLR and GENIE3 on GRN reconstruction from synthetic data. (C) Clustering and diffusion pseudotime for E12.5 muscle development data used in benchmarking. (D) Runtimes for reconstruction. Runtimes were measured on a MacBook Pro laptop with 2.9 GHz Intel Core i7 processor, 16GB 2133 MHz DDR3 of RAM). (E) Precision-recall curves for three versions of Epoch (Pearson, MI, and MI with crossweighting), CLR, and GENIE3 based on reconstruction from muscle development dataset.

Figure 3. mESC in vitro directed differentiation. (A) Directed differentiation protocol. Cells were harvested from d0 through d4 samples for MULTI-seq (B) Clustering for MULTI-seq data. (C) RNA velocity for MULTI-seq data. (D) Select marker gene expression. (E) Cell populations by treatment. (F) Quantification of cell fate distribution based on treatment comparing: mesendoderm vs. neuroectoderm, endoderm vs. mesoderm, cluster 7,0 vs. 7,1 along endoderm fate, and cluster distribution along neuroectoderm fate.

Figure 4. MULTI-seq directed differentiation network reconstruction. (A) Latent time annotation for MULTI-seq data. (B) population breakdown by cluster of epoch assignment along each germ layer path. (C) Reconstructed dynamic networks along mesoderm, endoderm, and neuroectoderm paths. (D) Top regulators as predicted by Epoch via betweenness and degree for the third epochs along each germ layer path.

Figure 5. Mesodermal network analysis. (A) The epoch 3 subnetwork of the mesodermal dynamic network. TFs are colored by community and faded by betweenness. Blue and red edges represent activating and repressive edges respectively. (B) Average community expression over time by treatment along the mesodermal path. Communities shown (each row) are from the epoch 2 subnetwork (yellow), transition (green), and epoch 3 subnetworks (aqua). (C) The epoch 3 differential network between WAG (on) and WA (off). Interactions in this network represent edges present in the WAG mesodermal network that are not present in the

1054 WA mesodermal network. TFs are colored by community and faded by betweenness. Blue and
1055 red edges represent activating and repressive edges respectively. (D) Peg3 and Dnmt3a
1056 expression in a sampled portion of relevant cell types in gastrulation data from Chan et al.
1057
1058 **Figure 6.** Similarity of targets of TFs in different treatments. (A) Jaccard similarity of targets of a
1059 given TF (each row) between treatments (as indicated by treatment 1 and treatment 2).
1060 Similarity is normalized by intra-treatment baseline (expected) similarity, such that values <1
1061 (green) indicate more-than-expected differences. (B) The 13 TFs selected for GSEA and their
1062 average pairwise Jaccard similarity amongst the treatment-specific networks. (C) Summarized
1063 results of performing GSEA on targets of TFs in B. Heatmap shows rankings (top 10 shown)
1064 based on frequency a term is considered enriched amongst the 13 TFs (1 = most frequent
1065 enriched term).
1066
1067 **Figure 7.** Integrating signaling effectors and tracing paths to target genes. (A) Average
1068 expression of targets of signaling effectors by treatment over time. Pathway and effector are
1069 labled on each row. (B) Shortest path analysis showing length of shortest paths from Wnt
1070 effector targets (each row) to mesodermal target genes (each column) within treatment-specific
1071 networks. Path lengths are normalized against average path length within a network. Blue and
1072 red indicate paths that activate and repress the target gene respectively. If such a path does not
1073 exist, it has length of infinity and is therefore white. (C) The epoch 3 subnetwork of the
1074 endodermal dynamic network. TFs are colored by community and faded by betweenness. Blue
1075 and red edges represent activating and repressive edges respectively. (D) Shortest path
1076 analysis showing length of shortest paths from PI3K suppression (targets of Foxo1 activation)
1077 (each row) to endodermal genes (each column) within treatment-specific networks. As before,
1078 path lengths are normalized against average path length within a network. Blue and red indicate
1079 activation and repression of the target gene respectively. If a path does not exist, it has length of
1080 infinity and is therefore white.
1081
1082 **Supplementary materials**
1083
1084 **Supplementary Figure 1.** Epoch benchmarking continued. (A) Full results of benchmarking
1085 reconstruction of synthetic data. Four versions of CLR (Pearson vs. MI, crossweighting vs. no
1086 crossweighting), four versions of Epoch (same as CLR), and three versions of GENIE3 (original,
1087 limited to dynamic genes, and with crossweighting) are compared. (B) Full results of

benchmarking reconstruction of the muscle development data. Methods are the same as in A. (C) Early mESC directed differentiation data (subsetted from Spangler et al. 2018). Zfp42 expression and velocity pseudotime are shown. (D) Full results of benchmarking reconstruction of the early Spangler et al. data. Methods are the same as in A. (E) Effect of smoothing vs. unsmoothing on AUPR of reconstruction using various methods. (F) Clustering, diffusion pseudotime, velocity pseudotime, and latent time of the E12.5 muscle development data. (G) Precision recall curves of four variations of Epoch using three different pseudotime annotations.

Supplementary Figure 2. MULTI-seq directed differentiation cell typing. (A) timepoint, treatment, total counts, and percent mitochondrial transcripts content of the MULTI-seq data. (B) Select marker gene expression of the MULTI-seq data. (C) Differential expression analysis of the MULTI-seq data by cluster.

Supplementary Figure 3. MULTI-seq cell type classification and cluster transitions. (A) SingleCellNet classification results. The classifier was trained from mouse embryo data found in literature, curated by members in the lab. (B) Sankey diagrams depicting cluster transitions (and cell fate transition proportions) as broken down by treatment.

Supplementary Figure 4. MULTI-seq network reconstruction and top regulators. Heatmaps 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.

Figure 1

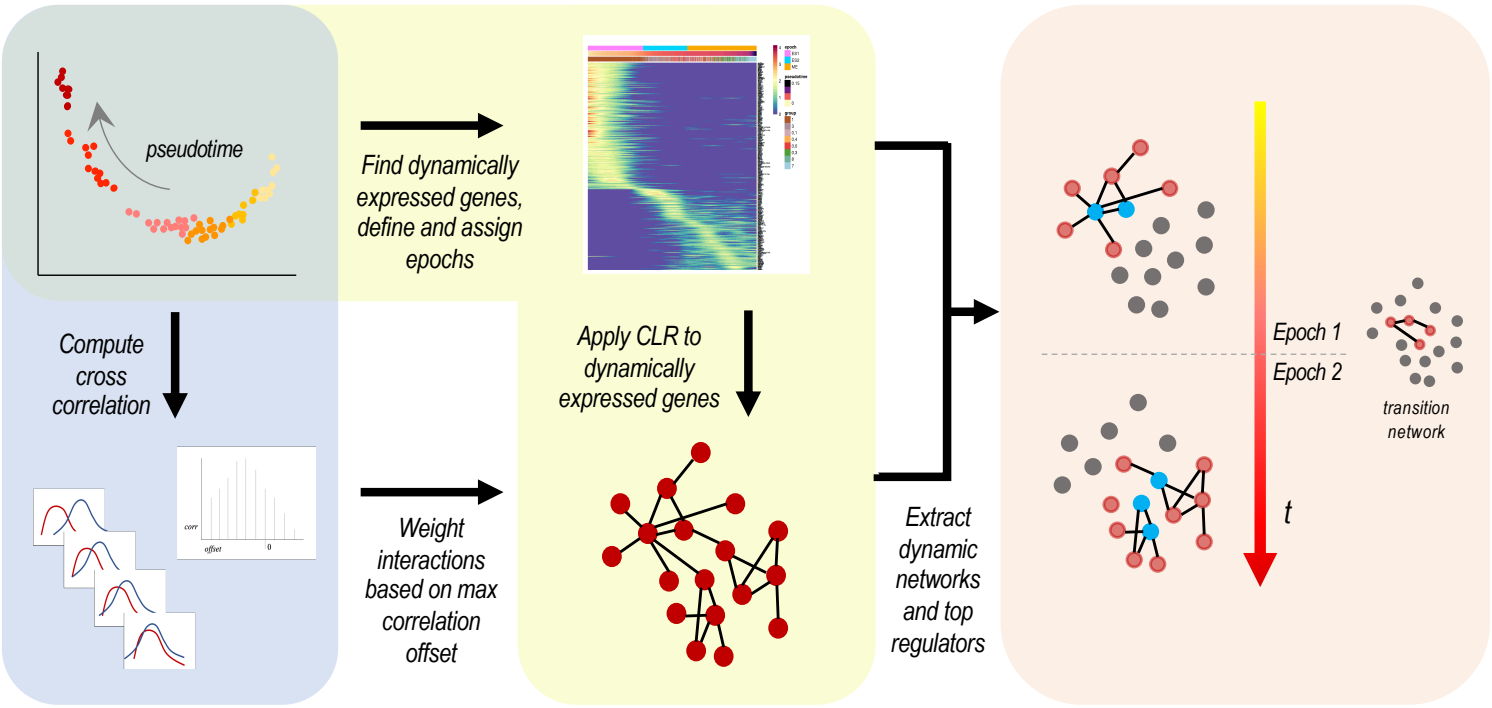
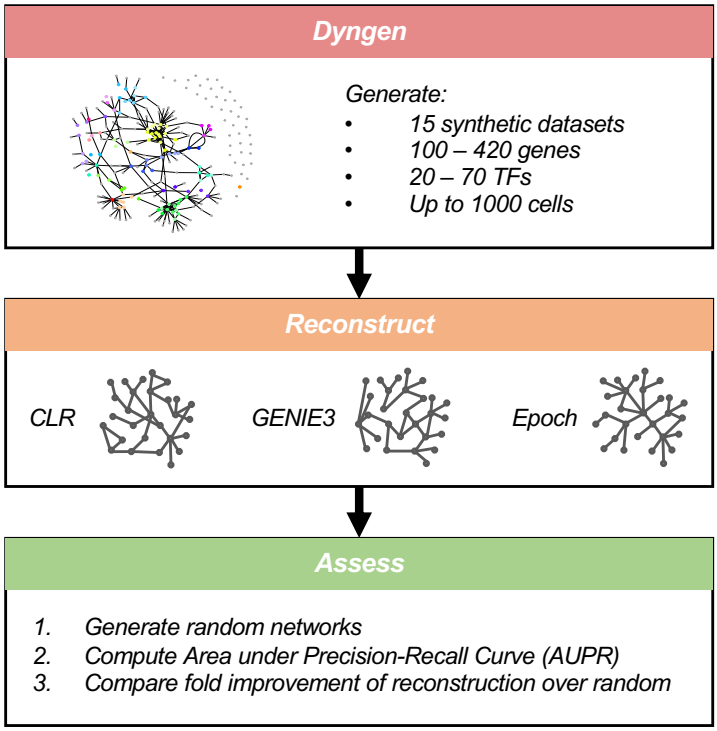
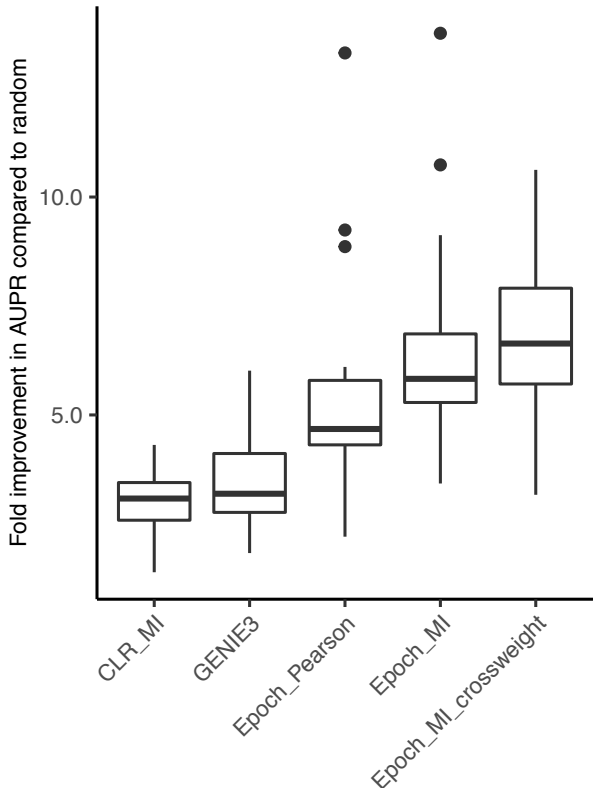

pseudotime
Find dynamically expressed genes, define and assign epochs
Compute cross correlation
Apply CLR to dynamically expressed genes
Weight interactions based on max correlation offset
Extract dynamic networks and top regulators
Epoch 1
Epoch 2
transition network
t

Figure 2

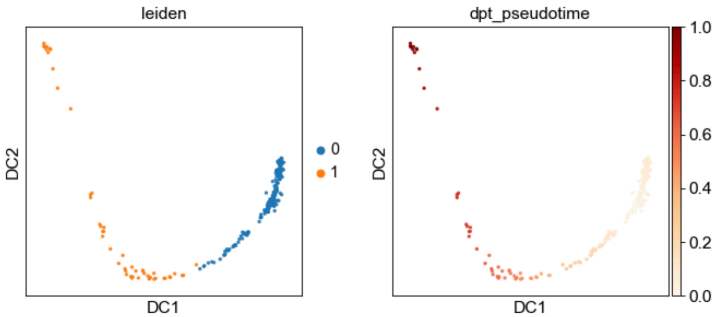
A



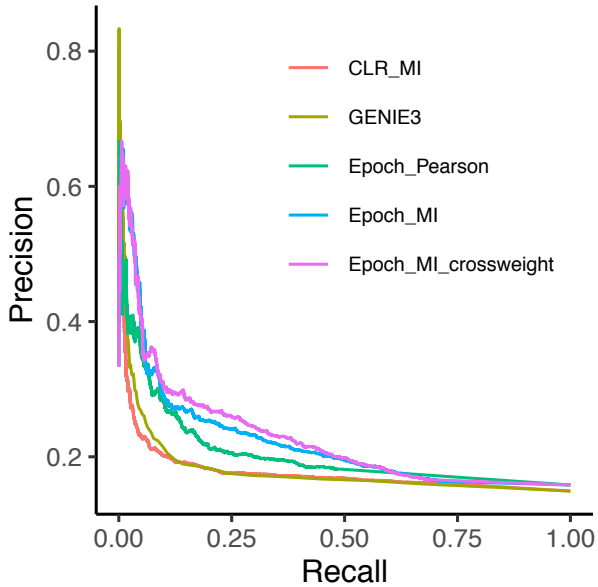
B



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 3

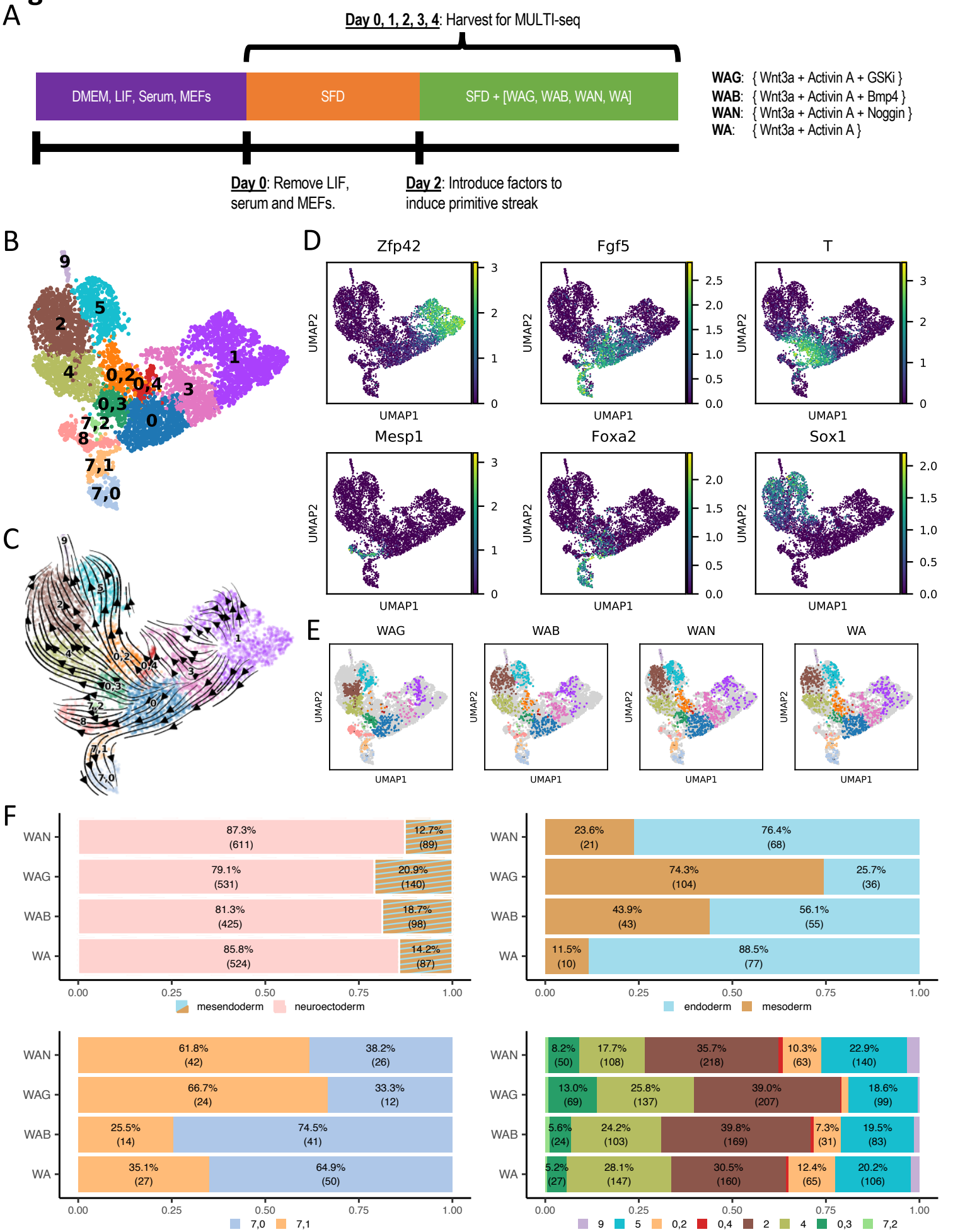
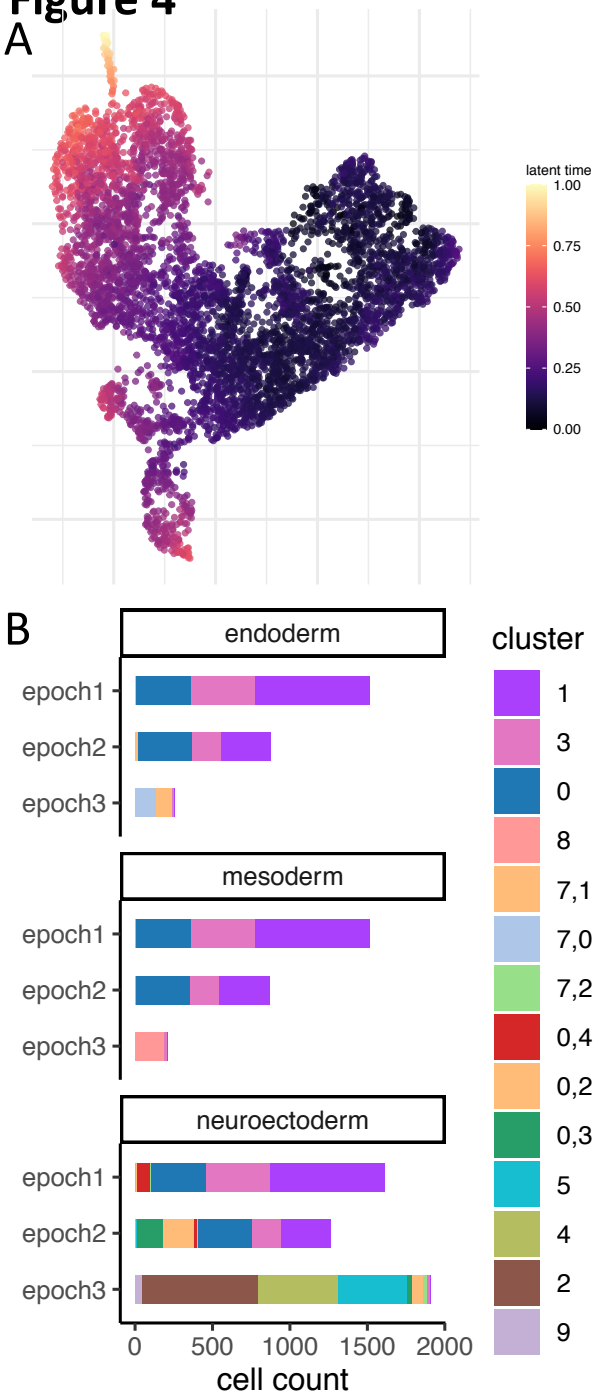
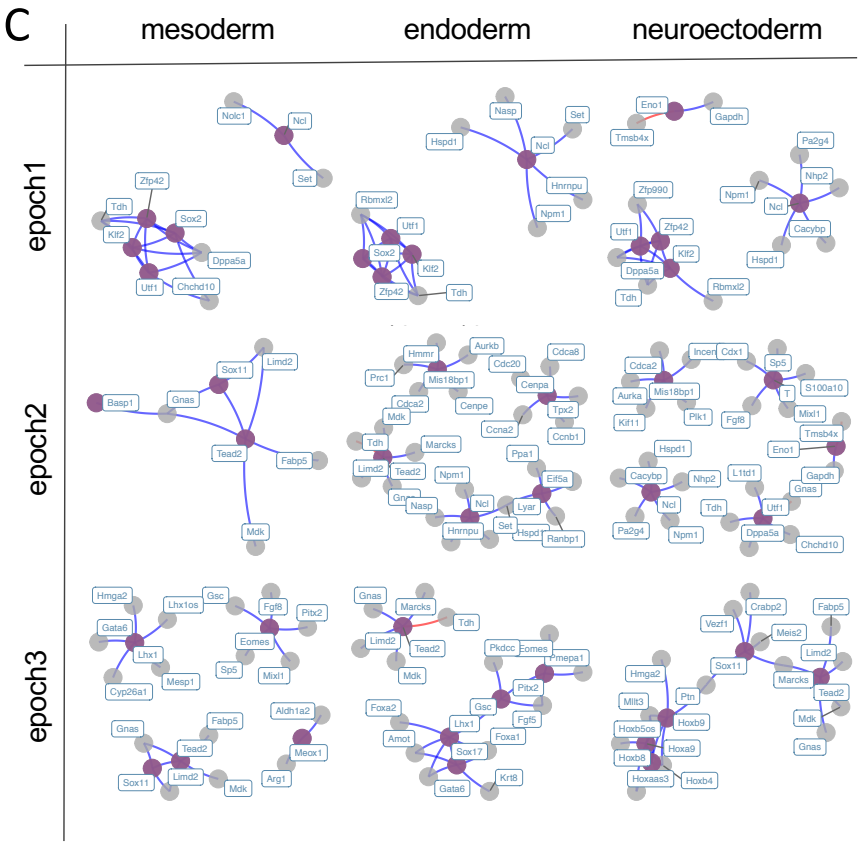

A
Day 0, 1, 2, 3, 4: Harvest for MULTI-seq
DMEM, LIF, Serum, MEFs
SFD
SFD + [WAG, WAB, WAN, WA]
WAG: { Wnt3a + Activin A + GSKi }
WAB: { Wnt3a + Activin A + Bmp4 }
WAN: { Wnt3a + Activin A + Noggin }
WA: { Wnt3a + Activin A }
Day 0: Remove LIF, serum and MEFs.
Day 2: Introduce factors to induce primitive streak
B
C
D
Zfp42
Fgf5
T
Mesp1
Foxa2
Sox1
UMAP1
UMAP2
E
WAG
WAB
WAN
WA
F
mesendoderm
neuroectoderm
endoderm
mesoderm
7,0
7,1
9
5
0,2
0,4
2
4
0,3
7,2

Figure 4


A
latent time
1.00
0.75
0.50
0.25
0.00
B
endoderm
mesoderm
neuroectoderm
epoch1
epoch2
epoch3
0
500
1000
1500
2000
cell count
cluster
1
3
0
8
7,1
7,0
7,2
0,4
0,2
0,3
5
4
2
9


C
mesoderm
endoderm
neuroectoderm
epoch1
epoch2
epoch3

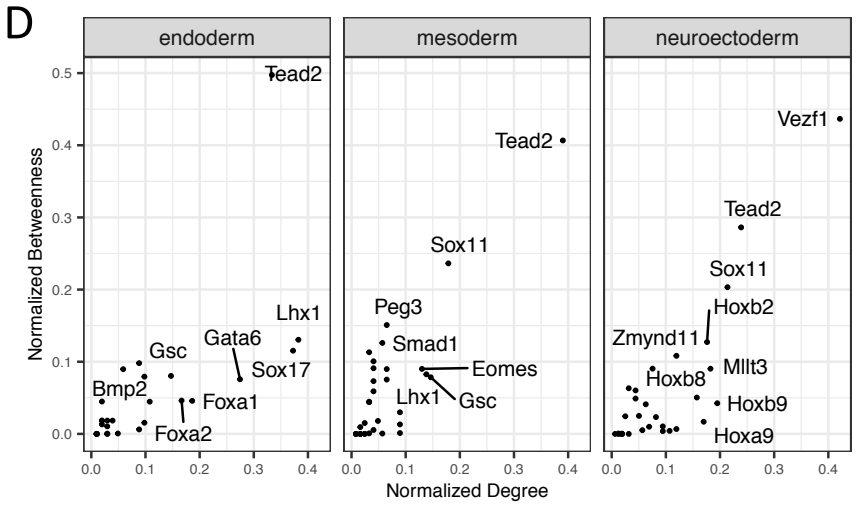
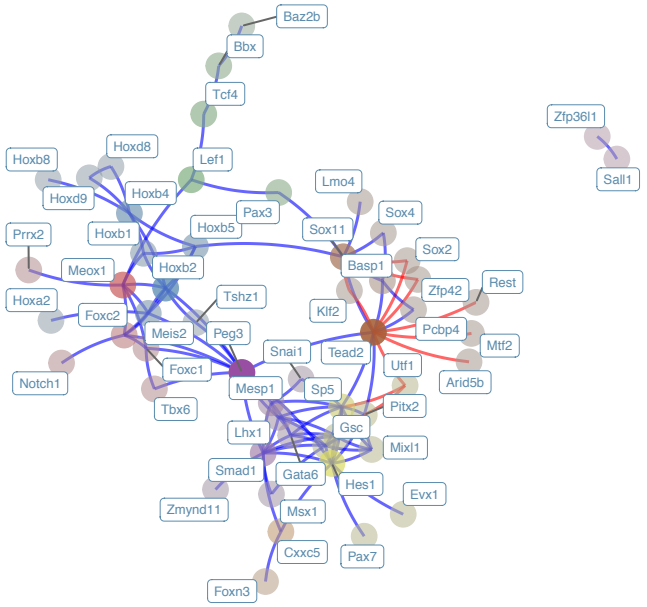
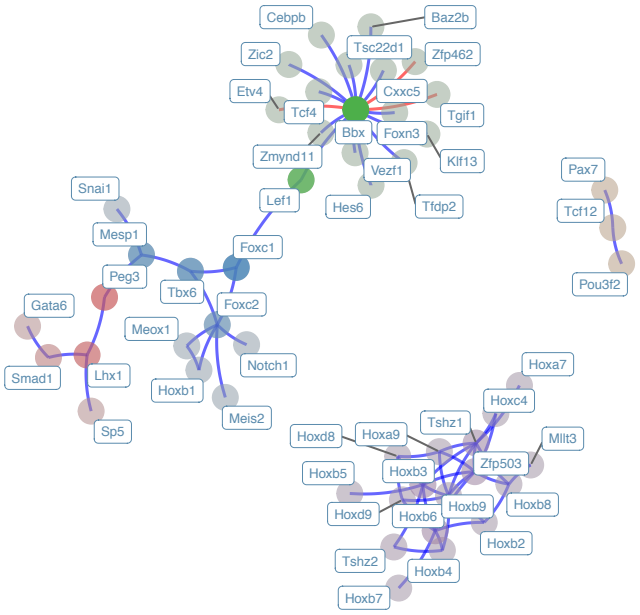

D
endoderm
mesoderm
neuroectoderm
Normalized Betweenness
Normalized Degree
Tead2
Lhx1
Gata6
Gsc
Sox17
Bmp2
Foxa1
Foxa2
Tead2
Sox11
Peg3
Smad1
Eomes
Lhx1
Gsc
Vezf1
Tead2
Sox11
Hoxb2
Zmynd11
Mllt3
Hoxb8
Hoxb9
Hoxa9

Figure 5

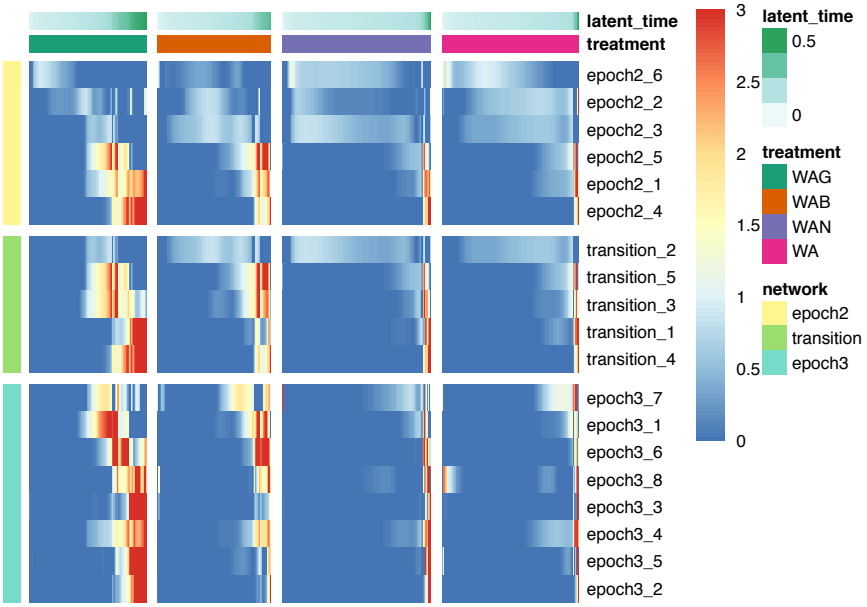
A



C



B



D

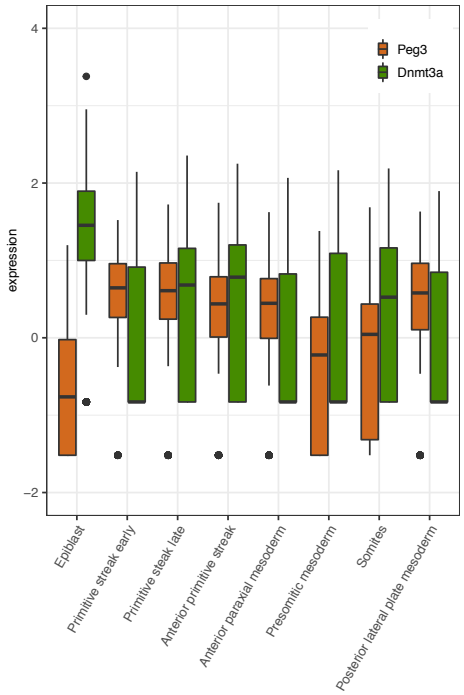
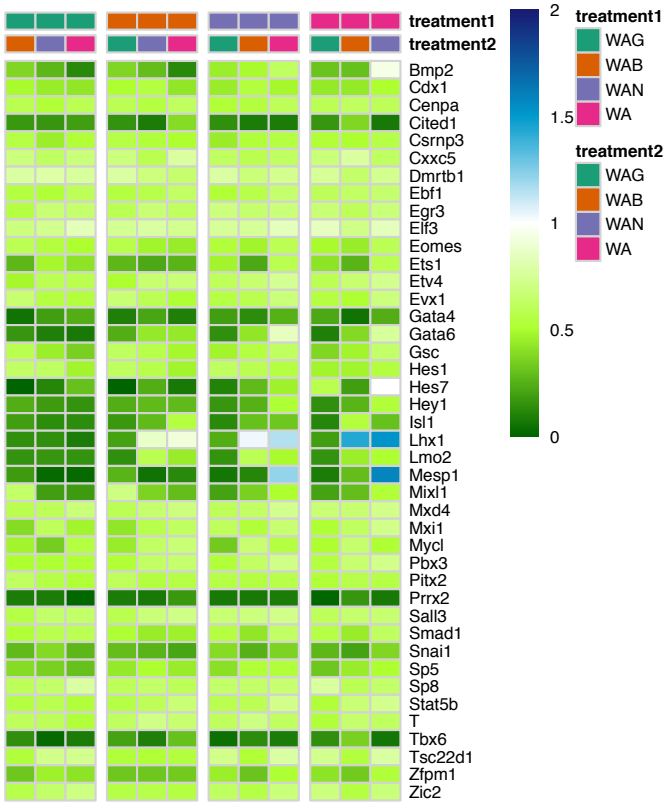


Figure 6

A



B

TF	avg(JS)	TF	avg(JS)	TF	avg(JS)
Prrx2	0.06175	Hes7	0.12431	Hey1	0.21341
Tbx6	0.06835	Cited1	0.12985	Lmo2	0.25975
Mesp1	0.09984	Gata6	0.19159	Mixl1	0.28015
Gata4	0.11422	Isl1	0.20265	Lhx1	0.28297
				Ets1	0.29662

C

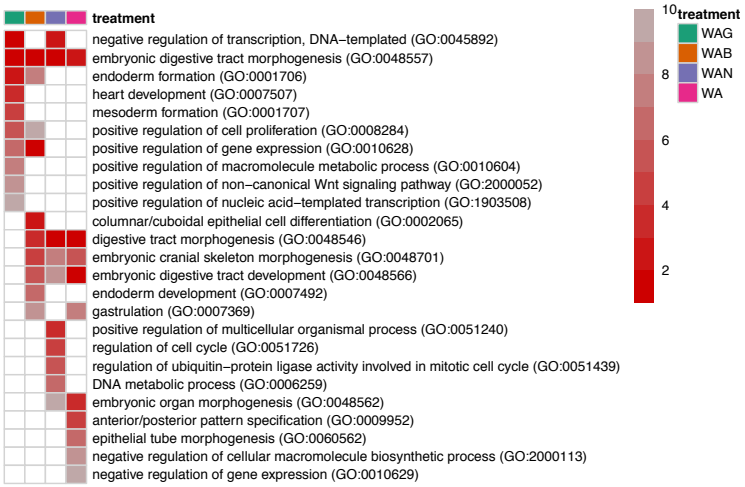
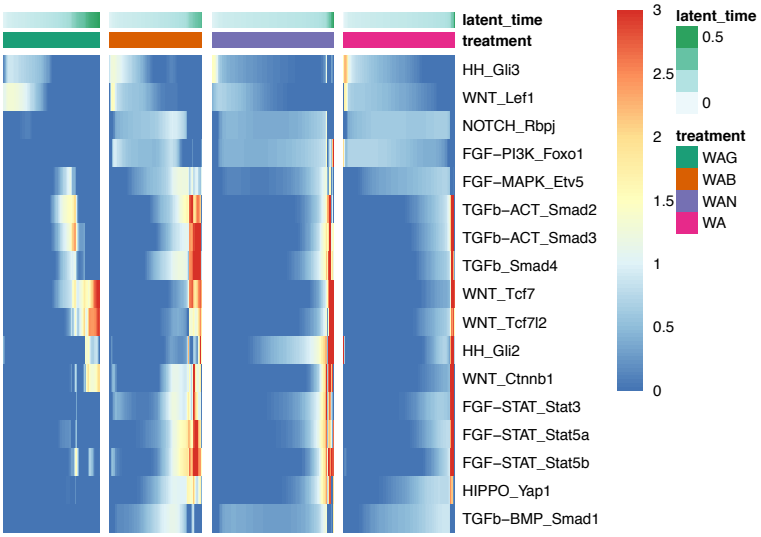
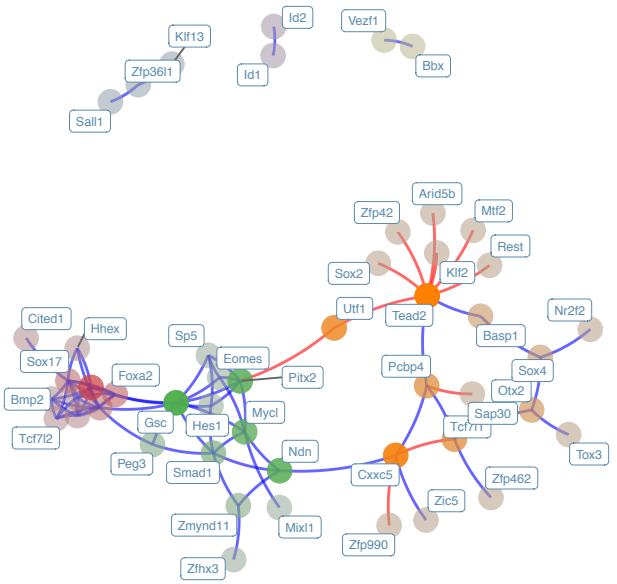


Figure 7

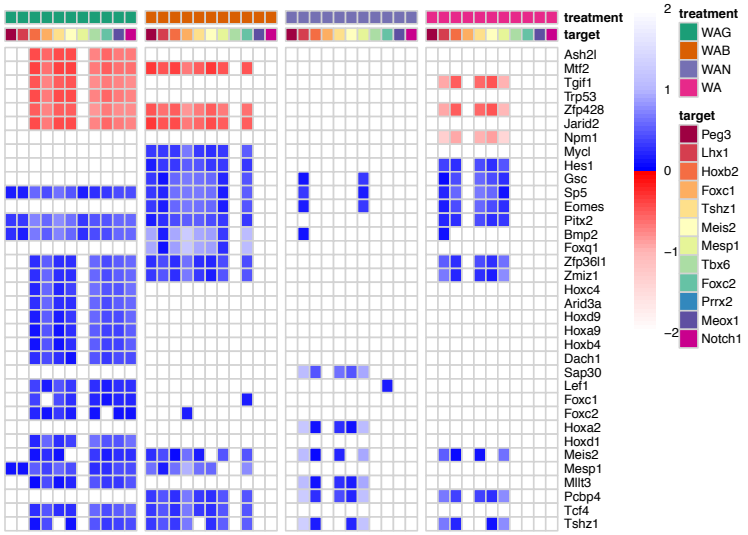
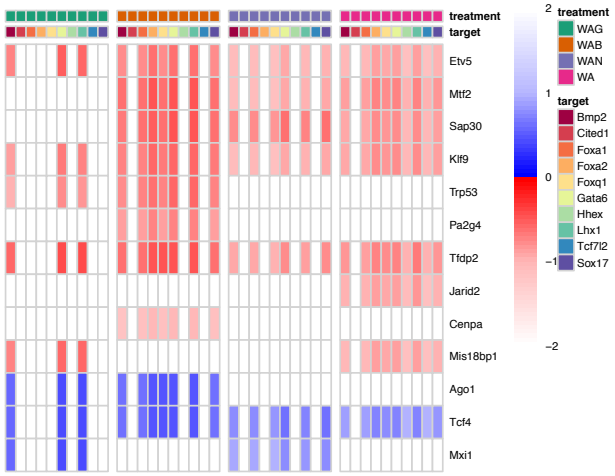
A



C



B



D