

SYSTEMS BIOLOGY

Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development

Kai Zhang¹, Mengchi Wang², Ying Zhao¹, Wei Wang^{1,2,3*}

Transcriptional regulation is pivotal to the specification of distinct cell types during embryonic development. However, it still lacks a systematic way to identify key transcription factors (TFs) orchestrating the temporal and tissue specificity of gene expression. Here, we integrated epigenomic and transcriptomic data to reveal key regulators from two cells to postnatal day 0 in mouse embryogenesis. We predicted three-dimensional chromatin interactions in 12 tissues across eight developmental stages, which facilitates linking TFs to their target genes for constructing transcriptional regulatory networks. To identify driver TFs, we developed a new algorithm, dubbed Taiji, to assess the global influence of each TF and systematically uncovered TFs critical for lineage-specific and stage-dependent tissue specification. We have also identified TF combinations that function in spatiotemporal order to form transcriptional waves regulating developmental progress. Furthermore, lacking stage-specific TF combinations suggests a distributed timing strategy to orchestrate the coordination between tissues during embryonic development.

INTRODUCTION

Transcription factors (TFs) are essential regulators of cell fate and play pivotal roles in development (1). Identification of critical TFs that drive tissue differentiation can provide key mechanistic insights into the developmental process. However, a complete catalog of driver TFs at each developmental stage in each tissue has not been established. Tackling this challenge requires dissecting the transcriptional networks and characterizing the genome-wide influences of key TFs.

The global influence of a TF in the cell is conveyed through its regulatory effect on the target genes, which is propagated over the transcriptional network. The TF's regulatory activity is affected by its own expression level and post-translational modification as well as other factors such as the presence of collaborative co-factors (2). Therefore, the expression level of a TF is not always correlated with its activity (3, 4). In light of this, many methods have been proposed to infer the activity of TFs using statistical or machine learning approaches. For instance, Schacht *et al.* (4) developed a statistical model to estimate the regulatory activity of TFs using their cumulative effects on their target genes. Arrieta-Ortiz *et al.* (5) used the linear model to infer the TF activity (TFA) by predicting target genes' expression levels. Although these methods were able to predict the local activity of a TF, i.e., the expression level of their direct target genes, measuring the system-wide influence of a given TF is not their focus. As genes rarely function alone but usually cross-talk with each other to form complex regulatory logic, it is reasonable to believe that the global influence of a TF can, in principle, better predict the cell-state change upon perturbing the TF than its local influence (6, 7).

Previously, we have shown that the personalized PageRank can successfully infer the global impact of TFs (8, 9). It clearly outperformed the motif enrichment analysis and the TFA approach (9). In addition, several predicted TFs were later validated experimentally

and demonstrated previously unappreciated roles (8, 9). Building upon our previous studies, we have made further improvements by incorporating spatial long-range interaction information into network construction, which helps assign distal regulatory elements to target genes. Furthermore, we have developed a software package, dubbed Taiji, to help biologist to perform integrated analysis and identify driver TFs using different genomic information, including chromatin state [from chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) or assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)], gene expression profile [from RNA sequencing (RNA-seq) or microarray], and chromatin long-range interactions (from Hi-C or computational prediction).

The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped the epigenomic dynamics during mouse embryonic development in 12 tissues and eight developmental stages. While this dataset provides an unprecedented opportunity to dissect the complex transcriptional regulatory logic during development, it also poses great challenges for integrated computational analyses. Using our framework, we have identified TFs that are crucial in defining tissue differentiation. Furthermore, we applied Taiji to the existing data in earlier development, from two-cell stage to embryonic stem cell, which complements the data generated by the mouse ENCODE project. Our analyses thus provide the first comprehensive catalog of key regulators for a variety of tissues during mouse embryogenesis. While many of the identified regulators are well supported by the literature, the newly found key TFs in development provide a valuable resource for follow-up biological study. We also uncovered TF combinations that activate in a spatiotemporal manner, which behave like transcriptional waves to orchestrate the developmental progress and tissue specification.

RESULTS

An overview of the Taiji framework

To identify key regulators, Taiji integrates various genomic information to build transcriptional regulatory networks by predicting

Copyright © 2019
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA. ²Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA. ³Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA.

*Corresponding author. Email: wei-wang@ucsd.edu.

regulations between TFs and genes. The TFs in the network are then ranked by the personalized PageRank algorithm (Fig. 1A) (10). Briefly, Taiji starts by identifying active regulatory elements, including active promoters and active enhancers, defined by ATAC-seq or H3K27ac ChIP-seq peaks. Enhancers were then assigned to their interacting promoters using chromatin interactions predicted by EpiTensor (11). To construct transcriptional regulatory networks, Taiji scans each regulatory element to identify putative TF binding sites using motifs from the CIS-BP database (12). TFs with putative binding sites in promoters or enhancers are then linked to their target genes in the networks. Lastly, the PageRank algorithm was used to assess the genome-wide influences of TFs. Furthermore, we used the node and edge weights to personalize the ranking algorithm. The node weights were determined by the z scores of gene expression levels, which allow the ranking algorithm to assign higher ranks to TFs that regulate more differentially expressed genes. The edge weights were set to be proportional to TFs' expression levels, which help filter out TFs that are not expressed or with low-expression levels (see Materials and Methods for details).

Predicting long-range chromatin interactions in mouse embryonic development

Hi-C data were not available for the tissues studied by the mouse ENCODE project. Therefore, we resorted to EpiTensor (11), an unsupervised learning method, to predict the enhancer-promoter interactions from histone modification data. As we have demonstrated in our previous study, EpiTensor not only shows high concordance with the experimental data, including Hi-C, chromatin interaction analysis by paired-end tag sequencing, and expression quantitative trait locus results, but also significantly outperforms other correlation-based methods and nearest gene assignment (11). Besides, a unique advantage of EpiTensor is that it can predict chromosome interactions at a high resolution of 200 base pairs (bp) within topologically associating domains (TADs) (Fig. 1B); this is much higher than the highest resolution (1000 bp) of the Hi-C data available in the public domain. Using EpiTensor, we predicted the chromatin interactions in 12 tissues at eight developmental stages. All predictions are available in the Supplementary Materials. The predicted interactions cover 25,358 transcription start sites (TSSs), >38% of all annotated transcripts [from GENCODE version M14 (13)], and 334 experimentally validated enhancers in the VISTA database (14) account for 55% of confirmed active enhancers during mouse embryonic development. To further assess the accuracy of EpiTensor's predictions, we compared the interactions predicted in mouse embryonic stem cells (mESCs) with Hi-C interactions from Dixon *et al.* (15). The result showed that EpiTensor's predictions are in great concordance with the Hi-C experiment (area under the receiver operating characteristic curve = 0.87) (fig. S1A). In Fig. 1C, we show one example of the predicted interactions. Specifically, locus iii is the promoter of *Tubb2b*, a critical gene for the cortical formation and brain morphogenesis (16). Locus iv is an experimentally validated active enhancer in embryonic mouse midbrain, hindbrain, and facial mesenchyme from the VISTA database (peak ID: mm1605). These two loci show correlated histone modification profiles across tissues/stages and were identified as interacting loci by EpiTensor. No interaction was found between locus iv and v despite their closer distance. This result, together with the large distance (160 kilo-base pairs) between locus i and v, indicates the power of EpiTensor for identifying long-range chromatin interactions.

Computational validation of the Taiji framework

Network construction and PageRank procedure are the two critical components deciding the overall accuracy of Taiji's results. To validate the network construction method, we trained a random forest model to predict the expression changes of target genes from the expression changes of their upstream regulators in the network. Using 10-fold cross-validation, we showed that the predicted changes is in good concordance with the actual values ($\rho = 0.756$) (fig. S1B). Next, to investigate whether the PageRank scores can be used to assess the TFs' importance, we used the GeneNetWeaver (17) to generate benchmark networks and associated them with dynamic models. Using the dynamic models, we generated the expression profiles and performed in silico knockout experiments for each regulator. The Euclidean distance between wild-type and knockout expression profiles was computed to represent the magnitude of the perturbation (fig. S1C). Regulators were ranked by their perturbation magnitudes, reflecting their global influences on the cells. We used this rank list as the ground truth to assess the performance of different algorithms. Among the five methods we have benchmarked, PageRank performed the best, showing a strong Spearman's correlation ($\rho = 0.914$) with the ground truth in the benchmark dataset of yeast (Fig. 2A). The gene expression of a TF is not a strong predictor of its importance, as indicated by the weak correlation ($\rho = 0.393$) with the perturbation magnitude (Fig. 2D). The regression-based TFA model (5) does not correlate with the ground truth ($\rho = -0.134$) (Fig. 2E). Another method, developed by Schacht and his colleagues (4), defines the TF activity as the weighted average of its target genes' expression (WATG) and it shows a weak correlation with TFs' overall importance ($\rho = 0.225$; Fig. 2C). We speculated that the averaging effect in WATG might contribute negatively to its performance on predicting global importance. Therefore, we developed a new metric called WSTG, which calculates the weighted sum of target genes' expression levels of a TF (see Materials and Methods for details). Note that the formulation of WSTG is conceptually equivalent to the network degree centrality. As expected, it performs much better than WATG ($\rho = 0.823$; Fig. 2B) because it takes into account the accumulated effects of all regulatees. In summary, algorithms that focus on measuring the "local" TF activities, i.e., TFA and WATG performed much worse than the network centrality-based methods (PageRank and WSTG), and PageRank showed the best performance.

A major advantage of PageRank algorithm, compared to other simpler centrality metrics such as WSTG, is that it is defined recursively and, hence, depends not only on the number of connected nodes but also on those nodes' PageRank values. Therefore, PageRank performed better than WSTG for the yeast network. To further confirm PageRank's superior performance, we conducted additional benchmarks using the *Escherichia coli* network and another 20 random subnetworks extracted from the yeast network (fig. S1D and Fig. 2F). In all the tested cases, PageRank consistently outperformed WSTG and other methods.

Transcriptional regulatory networks constructed from TF binding site predictions can be quite noisy. Therefore, assessing the robustness of PageRank algorithm against noises is critical for ensuring our results are reliable. For this purpose, we randomly inserted, deleted, or substituted a fraction of edges in the 20 benchmark networks mentioned above. We investigated whether the PageRank results were affected after introducing these noises. The result showed that the PageRank algorithm is resistant to random insertions

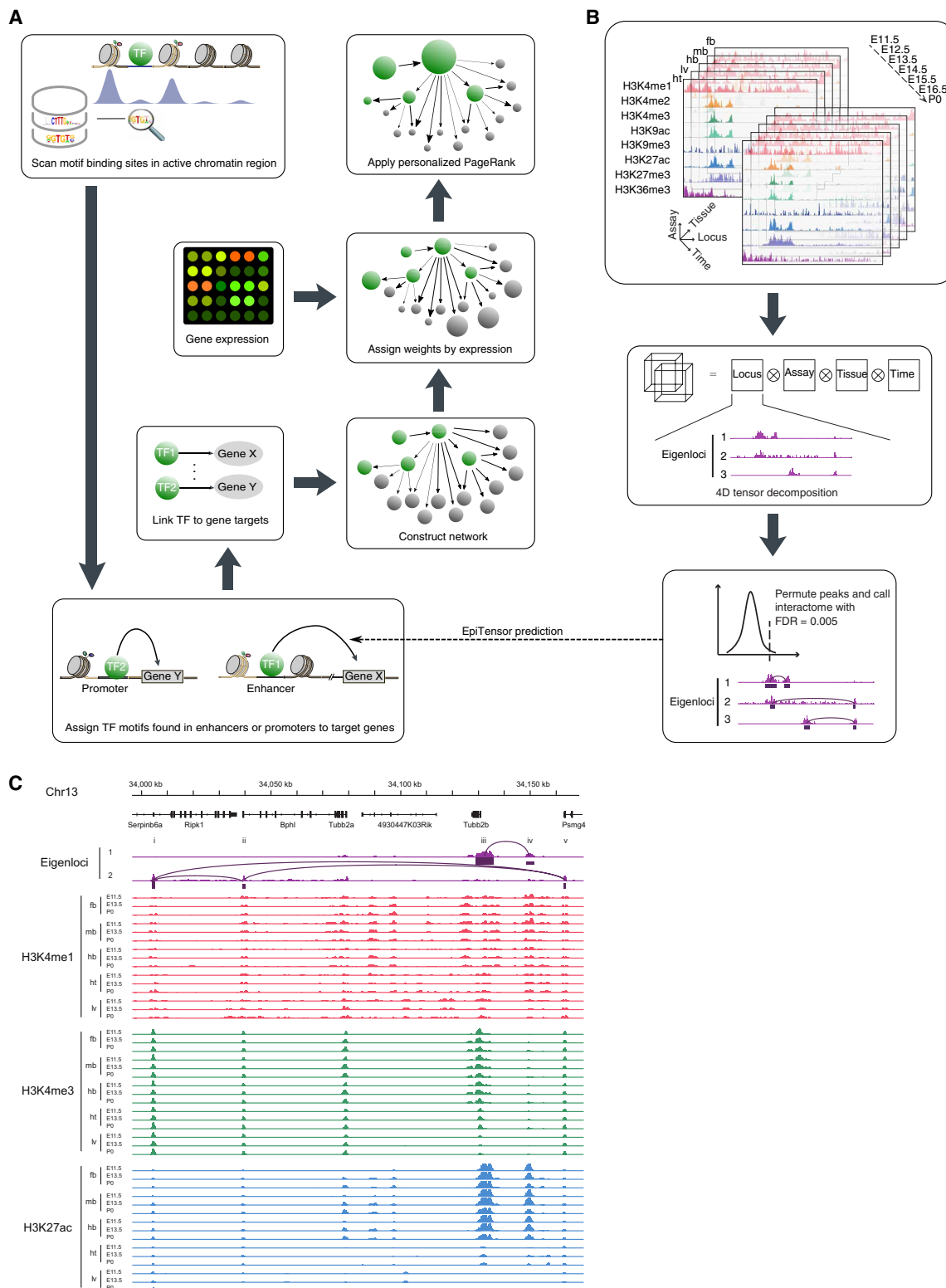


Fig. 1. The design of Taiji and EpiTensor algorithms. (A) The workflow of Taiji pipeline. (B) The workflow of EpiTensor algorithm. (C) An example of chromosome interactions predicted by EpiTensor. Predicted interacting loci are linked by lines. Locus i, ii, and v are three active promoters marked by H3K4me3 and H3K27ac in the tissue stages shown in the figure. Locus iv is a validated enhancer documented in the VISTA database. Locus iii is the promoter of *Tubb2b*, preferentially active in the brain. Eigenlocus 1 contains a strong peak representing the interaction between locus iii and iv; the covariation of histone modifications at the two loci can be exemplified by the coherent decrease of H3K4me1/H3K4me3 and H3K27ac signal levels across tissues/stages. Only H3K4me1, H3K4me3, and H3K27ac in E11.5, E13.5, and P0 are shown due to the limited space. ht, heart; lv, liver; hb, hindbrain; mb, midbrain; fb, forebrain.

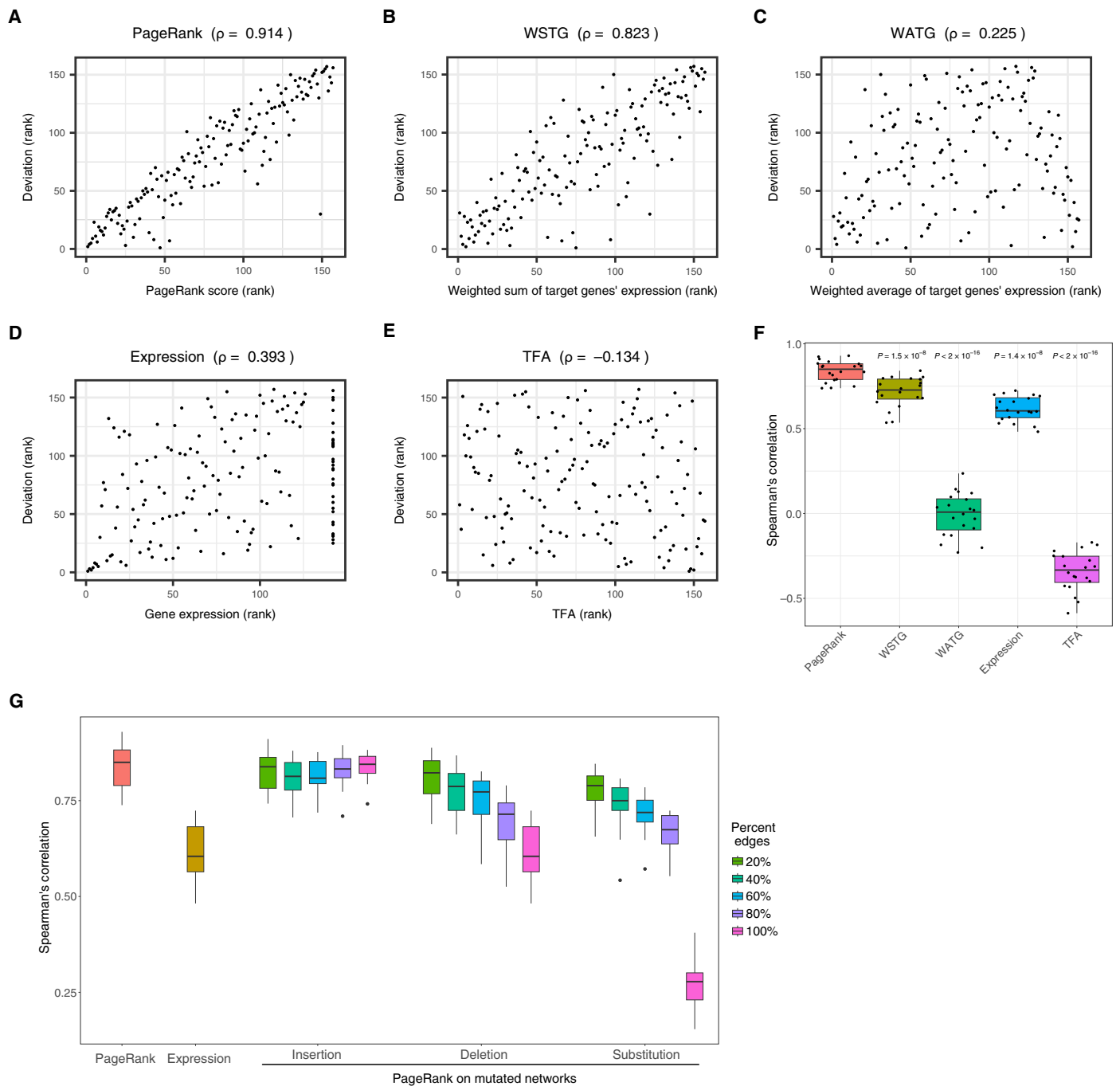


Fig. 2. Computational validation of the Taiji framework. (A to E) The performance of various metrics, PageRank (A), WSTG (B), WATG (C), gene expression (D), and TFA (E) on predicting TFs' importance, assessed by the Spearman's correlation with TFs' perturbation magnitudes resulted from simulated knockouts in yeast. The benchmark network contains 4441 genes and 12,873 edges. One hundred fifty-seven of the genes are regulators. (F) The performances of various metrics on predicting TFs' importance in 20 subnetworks extracted from the yeast gene regulatory network. Each subnetwork contains 1000 genes and at least 50 regulators. (G) Noises introduced by random insertions, deletions, or substitutions do not have large impact on PageRank's performance. Each box shows the results from 20 yeast subnetworks.

(Fig. 2G), suggesting that adding uniform noises to the network without modifying the existing edges is less likely to distort the real signal. Second, random deletion has only a minor effect on PageRank's performance. Removing a large percentage of the edges only induced a small drop of the average correlation. For example, when deleting 80% of the edges, the average correlation of the PageRank scores

with the ground truth drops to 0.691 from 0.839, which is still higher than that of gene expression ($\rho = 0.616$). When deleting all edges from the networks, the PageRank algorithm has the same performance as the gene expression ranking. The reason is that the gene expression levels are used to personalize the PageRank algorithm. After removing all edges, the ranking is then solely determined by the gene

expression levels. Among the three types of mutations, random substitutions pose the largest impact on PageRank's performance. However, as long as 20% of the edges are preserved, PageRank can still predict well ($\rho = 0.663$). Overall, we found that the PageRank algorithm is very robust against random noises. This feature is very important when applying this method to noisy biological networks.

The de novo construction of biological networks remains challenging. We indirectly assessed the accuracy of the constructed networks by predicting genes' expression changes. The predicted values are correlated with the experimental results despite the absolute values deviate slightly (fig. S1B), which suggests some unknown noise in the data or the constructed network. However, the random noise does not have major impact on the performance of our method, as shown in Fig. 2G.

Taiji reveals driver TFs during embryogenesis

We constructed transcriptional regulatory networks in 12 tissues at eight developmental stages. The average number of nodes and edges in these networks is 16,187 and 320,227, respectively. Six hundred thirty-nine (3.95%) of the nodes are TFs. On average, each TF regulates 501 genes, and each gene is regulated by 19 TFs (fig. S2). To characterize the dynamics of TFs' global influences across different tissues and stages, we removed TFs that do not present large variations [coefficients of variance (CVs) less than 1] in their PageRank scores, which gave us a list of 245 most variable TFs. We plotted the PageRank score matrix in Fig. 3A, ordering the samples (columns) by tissue types. We found that different tissues have quite distinct TF regulatory patterns (Fig. 3A, left). In contrast, when ordering the same data by developmental stages, no discernible patterns were observed (Fig. 3A, right), suggesting that transcriptional regulation largely takes place in a tissue-dependent fashion. To investigate the relationship between PageRank scores and expression levels, we calculated the Spearman's correlation between PageRank scores and expression levels for the 245 TFs (Fig. 3B). While 60% of TFs' expression levels strongly correlate with their PageRank scores ($\rho > 0.75$), a substantial portion (9.3%) of them show weak or no correlations ($\rho > 0.25$). The rest of the TFs' expression levels have moderate correlations with PageRank scores ($0.25 < \rho < 0.75$). The fact that about half of the TFs' expression levels are highly correlated with their PageRank scores may explain why gene expression achieves the third best in our *in silico* validation (Fig. 2F).

Using the FastME algorithm (18), we constructed a lineage tree based on the PageRank scores of 245 TFs (Fig. 3C). The result shows that the samples are mostly grouped by their tissue types. A mixture is only observed for four closely related samples (forebrain at E10.5, midbrain at E11.5, midbrain at E10.5, and craniofacial prominence at E10.5), which may be due to the difficulty of dissecting these tissues at their early stages. As a comparison, we found many unexpected clusters in the lineage tree constructed using the gene expression profile of the same 245 TFs (Fig. 3D). For instance, intestine at post-natal day 0 (P0) is grouped with liver samples; craniofacial prominence at E14.5 and E15.5 are clustered with limb samples. These results further demonstrate that PageRank is superior to gene expression as an indicator of TFs' activities.

We then asked whether there exist constitutively active TFs that exhibit high-ranking scores across all tissues and stages. For this purpose, we first selected TFs with average ranking scores larger than 3×10^{-3} , corresponding to the top 10% of all TFs. Next, we retained TFs of which the CVs are less than 0.5, giving us 35

constitutively active TFs in total (Fig. 4A). Similar to their transcriptional activities, the expression levels of these TFs remain relatively high across tissues and stages (fig. S3). Functional classification analysis revealed that these TFs are enriched in "metabolic process," "cellular process," and "developmental process," suggesting their general roles in basic cellular functions and embryonic development.

The earliest developmental stage that mouse ENCODE project has surveyed starts from E10.5. To obtain a complete view of mouse embryogenesis, we incorporated another published dataset that profiled the transcriptome- and genome-wide chromatin accessibility in earlier mouse embryos (19), spanning stages from E1 (two-cell) to E5 (blastocyst stage). Together, we have identified potential driver TFs in 12 tissues from eight stages, as well as those in two-, four-, eight-cell embryos, inner cell masses (ICMs), and mESCs. For the first time, the complex transcriptional regulation during mouse embryonic development is systematically mapped (figs. S4 and S5). The identified driver TFs include many well-known key regulators. For example, we successfully identified *Sox2*, *Nanog*, and *Pou5f1* (also known as *Oct4*) as crucial TFs in mESCs. To systematically validate our predictions, we performed literature search in five tissues that have been extensively studied, including the heart, lung, liver, kidney, and limb. Forty one of 56 (73.2%) TFs identified by our method are shown to be associated with either development or disease of these tissues (see the Supplementary Materials for details). For instance, 9 of 13 identified driver TFs in the lung have been previously reported to play a pivotal role during the bronchiole tree and terminal alveolar region formation of mouse lung (Fig. 4B) and another two TFs relate with lung cancer (20, 21). Apart from the spatial specificity (across tissues), the temporal pattern of TFs' activity (across stages) can also be accurately captured by our approach. For instance, *Nkx2-1* is initiated at the early stage of lung development, and *Nkx2-1* mutant embryos are arrested at early pseudoglandular (E11 to E15) stage (22), which is consistent with our analysis. Similarly, another key regulator, *Foxa2*, is present in the epithelial cells from the beginning of lung bud formation, and transgenic mice with *Foxa2* ectopically expressed in the lung epithelial cells exhibited defects in branching morphogenesis (23).

We have also found many previously unknown TFs. *Hlx*, identified as a key regulator for lung development, has not yet been studied in the lung. We showed that *Hlx* is regulating a number of high-rank TFs, i.e., *Sp1*, *Meis2*, and *Zfx* (Fig. 4C). The functional enrichment analysis of all *Hlx*'s regulatees shows that they likely participate in epithelial tubes development in the lung. Besides, some of the predicted TFs have been studied in closely related tissues. For instance, *Vsx1*, predicted as an important regulator in the hindbrain, was related to retina development (24).

During development, ESCs differentiate into three germ layers—ectoderm, endoderm, and mesoderm. To identify TFs that are specific to each layer, we grouped and compared tissues originated from the same layer against tissues from other layers (Fig. 5). The student *t* tests with a *P* value cutoff of 0.001 were used to identify TFs whose ranks change significantly in one germ layer compared with other layers. The functional relevance of the found layer-specific TFs is supported by the literature. For example, we have identified *Zic1*, *Zic4*, and *Zic5* as specific regulators for ectoderm, which is in agreement with the role of *Zic* family in neural development (25). In addition, many previously known layer-specific markers were also found

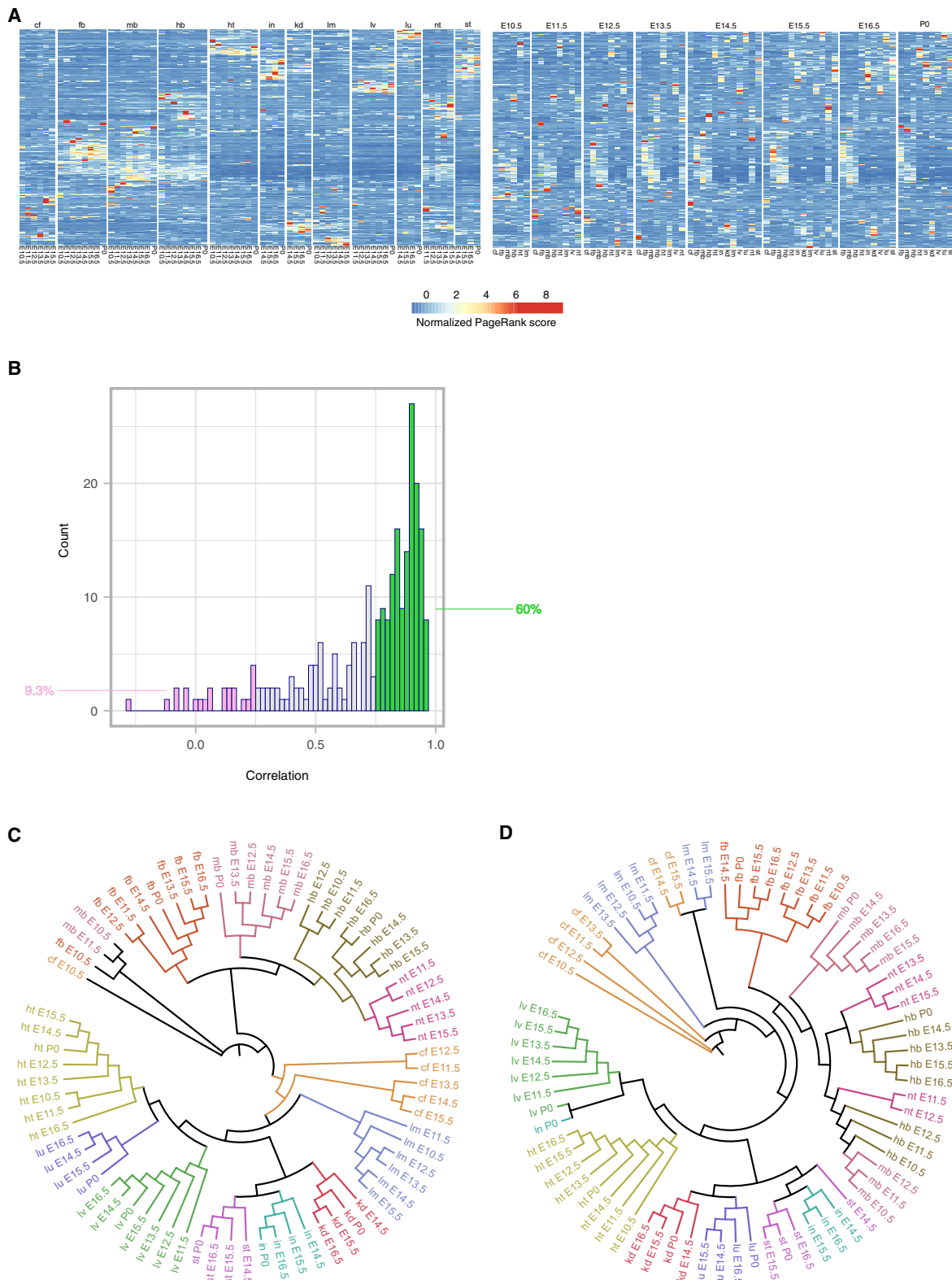
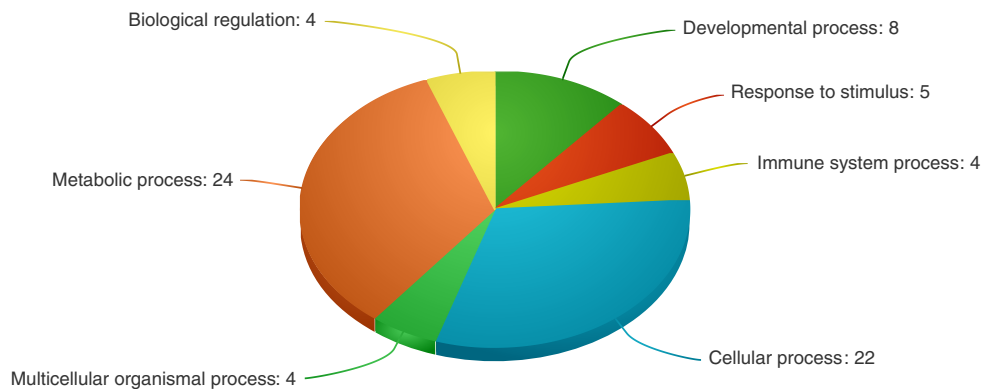


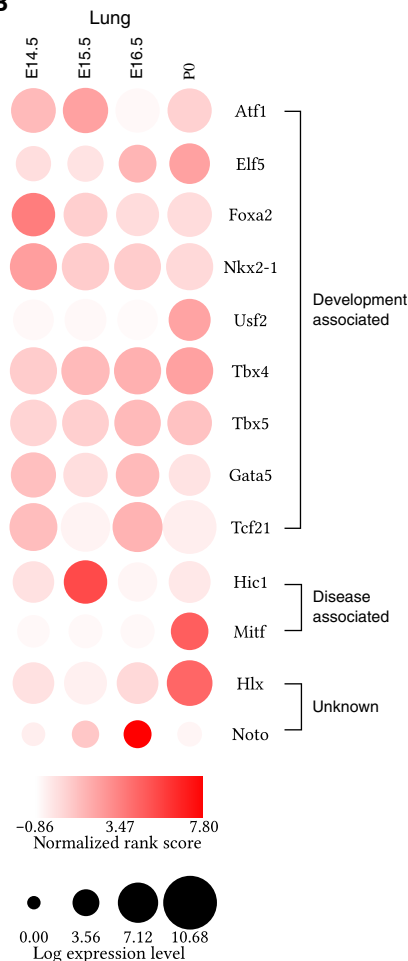
Fig. 3. TFA determined from Taiji accurately predicts tissue specification. (A) Two different views, arranged by tissue types (left) and arranged by stages (right), of the 245 most variable TFs' ranking scores during embryogenesis. (B) Histogram of the Spearman's correlations between PageRank scores and expression levels of 245 TFs across tissues and stages. Sixty percent of the TFs show strong correlations (>0.75), and 9.3% of the TFs show weak or no correlation (<0.25). (C) A lineage tree constructed from the ranking scores of the 245 TFs. (D) A lineage tree constructed from the gene expression levels, determined by RNA-seq, of the same 245 TFs in (C). cf, craniofacial prominence; in, intestine; kd, kidney; lm, limb; lu, lung; nt, neural tube; st, stomach.

A

E2f4, E2f6, Sp1, Sp2, Dnajc21, Zfp281, Glis2, Tcf12, Zeb1, Meis2, Zfp354c, Mzf1
 Sp4, Zfp523, Arnt, Ets1, Nfic, Rara, Nr2f6, Rxra, Bcl6, Arid3a, Rela, Stat2
 Stat3, Tcf3, Patz1, Ctcf, Zfp740, Rfx5, Foxj3, Foxk1, Zfp263, Nr2f2, Zfx



B



C

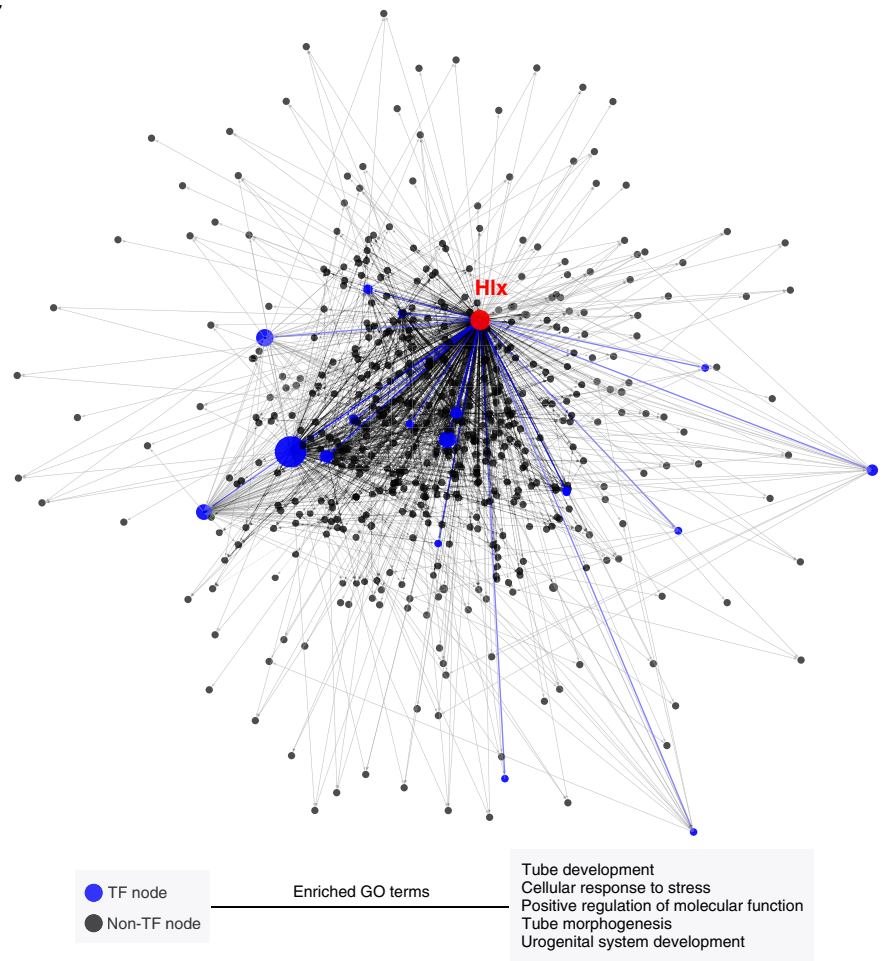


Fig. 4. Identification of driver TFs during mouse embryogenesis. (A) Functional classification of 35 constitutively active TFs reveals their functions in essential biological processes. (B) Identified driver TFs in the lung, including nine TFs related with lung development, two TFs related with lung diseases, and two TFs with unknown functions. (C) The network for *Hlx*, a previously unknown driver TF for the lung, and its regulatees in the lung at P0. Larger nodes represent TFs with higher rank. The bottom shows the top five enriched Gene Ontology (GO) terms for *Hlx*'s regulatees.

from our analysis, e.g., *Pax6* and *Otx2* for ectoderm (26, 27) and *Foxa2* for endoderm (28). To analyze the tissue specificity of the predicted driver TFs (listed in fig. S4), we further compared each tissue with other tissues originated from the same layer and used the student *t* tests with a *P* value cutoff of 0.001 to identify tissue-specific driver TFs. In total, 68 tissue-specific driver TFs were found, as shown in Fig. 5. Together, these results provide a comprehensive map for the future mechanistic study of embryonic development.

Transcriptional waves during embryogenesis

In addition to the tissue specificity, we have analyzed the temporal activity of TFs during the mouse embryogenesis. To identify clusters of TFs that show similar spatial-temporal patterns, we first performed the principal component analysis to reduce the dimension of the TF ranking score matrix. We reserved the first 20 components for clustering analysis as adding more components did not gain much explained variance (fig. S6A). We used the silhouette analysis (29) to select the clustering method and the number of clusters. Among the five algorithms we have tested, the *k*-means algorithm performed best, identifying 25 distinct dynamic patterns during embryogenesis (fig. S6B). We further performed functional enrichment analysis to identify enriched Gene Ontology (GO) terms for these clusters (listed in fig. S7). In general, we found that the enriched GO terms are consistent with the dynamic patterns of these clusters. For example, the cluster C12, showing a heart-specific dynamic pattern, is enriched with heart-specific functions, such as “embryonic heart tube development,” “blood vessel development,” and “cardiac ventricle development.”

The 25 clusters represent transcriptional waves that orchestrate the tissue differentiation. The first wave starts from as early as the two-cell stage, represented by the cluster C7 (Fig. 6A). In C7, TFs show the highest activity at two-cell stage (or possibly earlier as we do not have data in zygote). Example TFs from C7 include germ cell-specific factors such as *Obox1* and *Nr6a1*, both of which are essential for embryogenesis (30, 31), highlighting the roles of parental control in early development. As C7 wanes, C21 is turned on during four- and eight-cell stages, and C13 and C16 emerge in ICM and ESC, respectively. In C13 and C16, we found many well-known pluripotency regulators, such as *Pou5f1*, *Nanog*, and *Sox2*. These results provide valuable insights into the transcriptional program during early embryogenesis.

We also found clusters that are responsible for the differentiation of tissues from specific layers. For example, C5 is highly active in all four brain tissues from E10.5 to P0, suggesting its critical role in neural differentiation. Many TFs that are crucial for neural development were recovered in C5, such as *Zic5* (25), *Pax6* (27), and *Gbx2* (32). In contrast to C5, C1 is more active in mesoderm- and endoderm-derived tissues. TFs in this cluster are associated with functions specific to the development of these two layers. For example, *Hnf4a* was reported to play pivotal roles in the liver, colon, and kidney development (33–35).

Besides germ layer-specific transcriptional waves, there are also tissue-specific transcriptional programs that drive the differentiation of individual tissue. In Fig. 6C, we highlighted four clusters, C10, C12, C6, and C18, which are responsible for the stomach/intestine, heart, forebrain, and liver development, respectively. Note that we have found such a program for nearly every tissue, including the craniofacial/limb (C9), lung (C3), kidney (C4), midbrain (C24), and neural tube/hindbrain (C19). See fig. S7 for details.

Most of the tissue-specific TFs are long lived, playing roles in almost every stage of development. However, a few TFs exhibit transient transcriptional spikes, probably related to their stage-specific functions. In Fig. 6D, we show two such examples which activate at different stages in craniofacial development (C8 for E11.5 and C9 for E13.5 and E14.5). As the mechanism of craniofacial development remains largely unknown, our discovery, therefore, provides an invaluable resource for the future studies.

The mouse embryonic data we analyzed here provide the first comprehensive temporal transcriptomic and epigenomic dynamics in a diverse set of tissues, revealing “transcriptional waves” in which TFs are active in specific tissues at a particular development stage. The observed transcriptional waves provide new insights into understanding how the development of different tissues is coordinated. During embryonic development, the growth of tissues must be tightly coordinated through the control of developmental timing. This can be achieved by a series of central instructive signals that activate at different stages to synchronize the development of tissues. However, in this study, we did not observe any TF that is active in all tissues at a particular developmental stage, and hence, we conclude that central instructive signals may not exist at the transcriptional regulation level. Instead, we propose a “distributed coordination” model in which the coordination of tissue development is achieved by tissue-to-tissue communications; the timing information is communicated in a tissue-to-tissue fashion and then propagated through the entire body. Evolutionarily, the model of distributed coordination removes the burden of developing a “central authority,” which is fragile and inefficient. The distributed coordination has been shown to play an important role in *Drosophila* tissue development (36). Our analyses show that the same principle may also apply to mouse embryonic development.

DISCUSSION

The embryonic development is controlled by TFs that relay environmental signals through control of gene expression. Efficient and precise regulation of this process requires cooperation of many TFs to form complex transcriptional circuits. As a result, the global influence of a TF cannot be readily inferred from its local activity, i.e., its regulatory effect on direct target genes. To identify driver TFs in different tissues and at different stages during mouse embryogenesis, we developed a novel framework to assess the global influence or activity of a TF. In silico simulations show that our method is able to predict genes' influences with great accuracy, outperforming other methods for the similar purpose. Compared with metrics for defining TFs' local activities (4, 5), gene expression performs reasonably well in our benchmark data, and the PageRank scores correlate well with gene expressions for many TFs in real data. However, there exists a large portion (~40%) of TFs showing only weak or moderate correlation between their PageRank scores and gene expression levels. In our previous work, we showed this type of TFs can indeed play a pivotal role in CD8⁺ T cell differentiation (9). Because the correlation between gene expression and the activity of a TF may vary in different cell states, our method is more robust in predicting TF importance regardless of its activity correlated or not with its expression.

Our method Taiji is capable of flexibly integrating diverse genomic and epigenomic data, including ChIP-seq, RNA-seq, ATAC-seq, and Hi-C. Considering the current limitation of Hi-C experiments, we

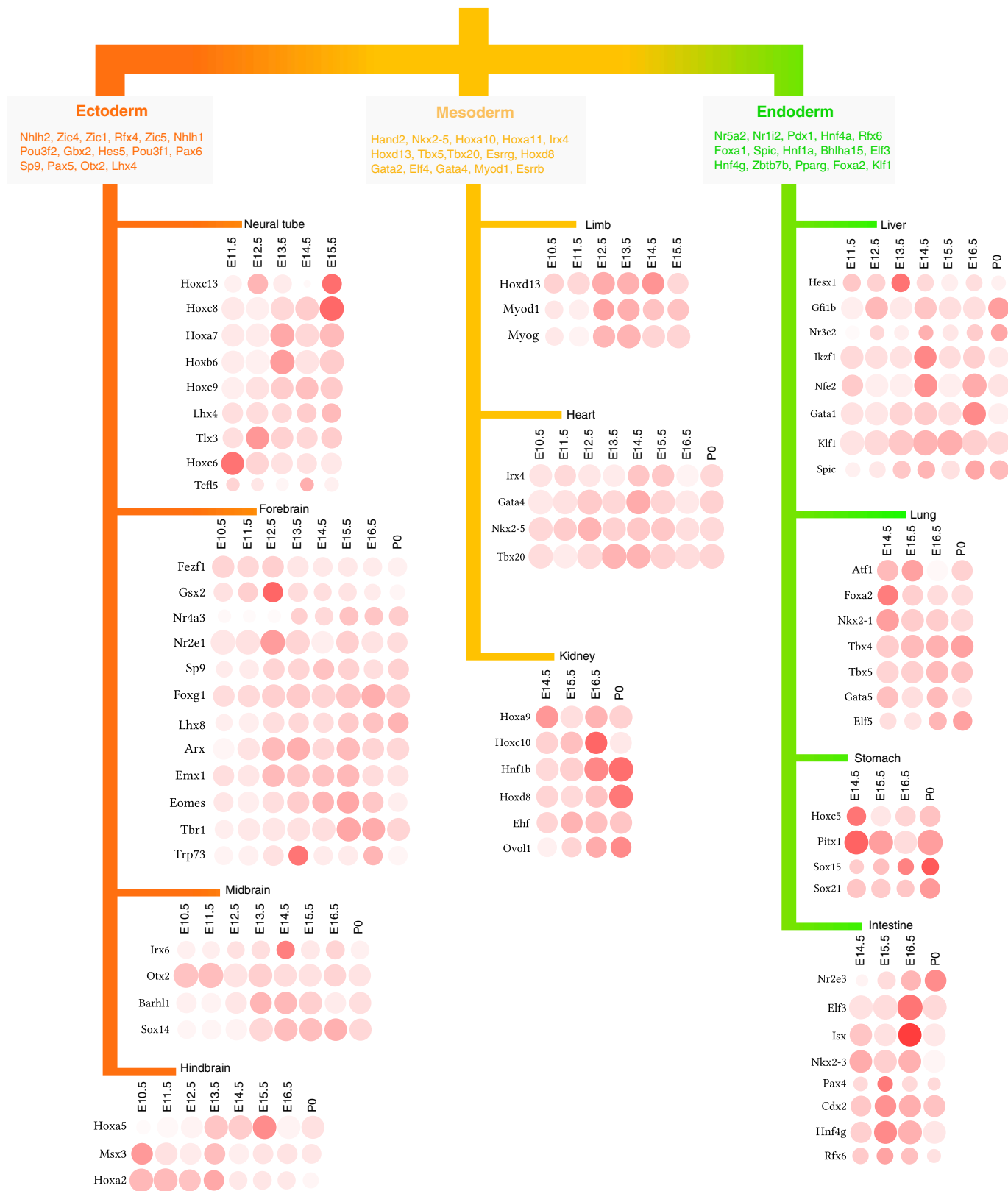


Fig. 5. Germ layer- and tissue-specific driver TFs in mouse embryonic development. The legends for the color scheme and the shape of the bubble are the same as those in fig. S4. Because of the limited space, we do not show them in this figure.

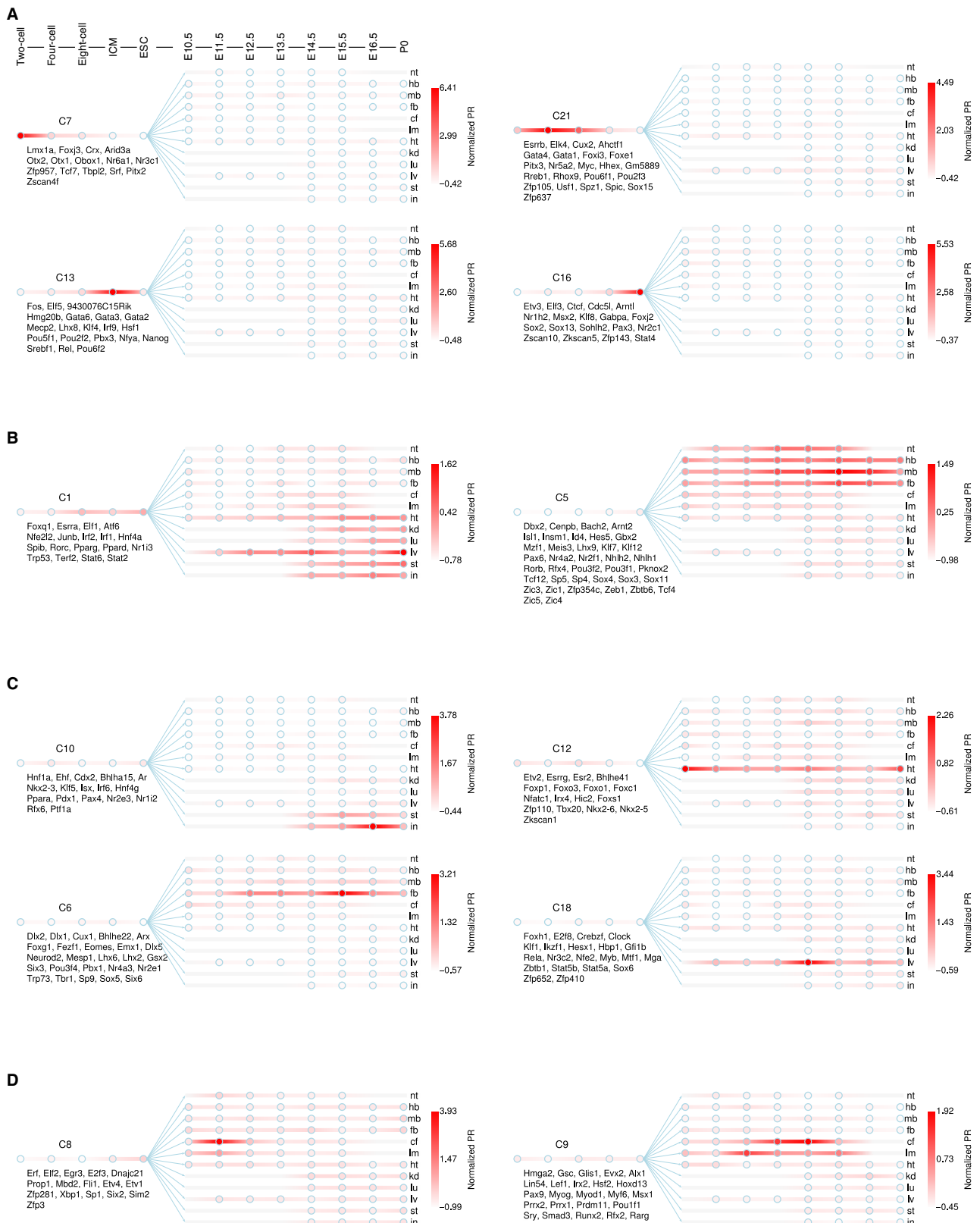


Fig. 6. Temporal transcriptional waves direct tissue differentiation during embryogenesis. (A) Four waves in early embryonic development. **(B)** Layer-specific transcriptional waves. **(C)** Four examples of tissue-specific transcriptional waves. **(D)** Stage-specific waves in craniofacial development. Circles in each panel represent developmental stages. From left to right, they are two-, four-, eight-cell, ICM, ESC, E10.5, E11.5, E12.5, E13.5, E14.5, E15.5, E16.5, and P0. TF members are shown below the names of clusters. PR, PageRank score.

provide a computational alternative when Hi-C experiment is infeasible or unavailable. In this work, we predicted three-dimensional (3D) chromatin interactions in 12 tissues and eight developmental stages of the mouse embryo, which provides the first 3D chromatin organization information. By leveraging the strength of various experiments, we have successfully mapped lineage-, tissue-, and stage-specific driver TFs throughout the mouse embryonic development from as early as the two-cell stage to P0. In addition to retrieving known key regulators, we have also identified previously unknown TFs responsible for tissue differentiation and development progress, which can guide the future experimental investigations to understand the regulatory mechanisms of development.

Particularly interesting is the observation of transcriptional waves represented by TF combinations that activate in a spatiotemporal fashion. We did not find stage-specific tissue-independent TFs, i.e., TFs that are active in all tissues at a specific stage (Fig. 3, A, C, and D). Therefore, we hypothesize that the coordination of tissue development is not achieved by global regulators, i.e., there lacks a “central timer,” but by sequential activations of regulators in individual tissues that function like a “distributed timer” and use tissue-to-tissue communication to ensure synchronization between tissues, i.e., tissues cross-talk and inform each other what is the current developmental stage. A distributed timer alleviates the burden of developing a central timer during evolution and also provides a more robust timing strategy relying on fewer pathways or loci.

MATERIALS AND METHODS

Prediction of chromatin interactions using EpiTensor

TADs were taken from the previous study (15). The input histone modification data were downloaded from the ENCODE data portal, including a common set of eight histone marks at seven time points in five tissues: H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K9me3 at developmental stages of E11.5, E12.5, E13.5, E14.5, E15.5, E16.5, and P0 in the heart, liver, forebrain, midbrain, and hindbrain. EpiTensor performed tensor analysis in the 4D space composed of tissue, time point, histone mark, and locus dimensions. The peaks in the eigenlocus vectors represent the covariation of histone modification signals that indicate 3D interactions of the corresponding loci. We considered the first 40 eigenlocus vectors, capturing, on average, 96.98% of total variance. The peaks in each eigenlocus vector were called by comparing to randomly shuffled background. The false discovery rate (FDR) and P value were calculated for each peak. In this study, we used 0.005 as the FDR cutoff. The EpiTensor predictions were validated in mESCs. First, we obtained the mESC-specific interactions by overlapping all predicted interactions in different tissues/stages with the H3K27ac peaks from mESC (ENCODE: ENCSR000CDE). We then compared the mESC-specific predictions with Hi-C interactions from Dixon *et al.* (15). We followed the same procedures of Zhu *et al.* (11) to draw the receiver operating characteristic curve.

Constructing TF regulatory networks

We used TF motifs, represented by position weight matrix, to predict TF binding sites at open chromatin regions within genes' promoters or enhancers and then connected these TFs to downstream genes. We chose this strategy because it is one of the most scalable approaches for studying TF gene regulations in a reasonably accurate way. Our network construction method contains three major steps listed below:

- 1) Identifying active promoters and enhancers: We define a gene's promoter as the 6-kb interval that covers the 5-kb upstream and the 1-kb downstream of the gene's TSS. As active promoters are usually indicated by active histone marks, we used the H3K27ac ChIP-seq peaks to determine the activity of a given promoter. In particular, we called a promoter active if it is overlapped with at least one peak. Any gene of which the promoter is inactive was excluded from the network. To increase the sensitivity, i.e., preserving more active genes in the analysis, in this step, we used a relaxed cutoff (q value equals to 0.1) in MAC's (model-based analysis of ChIP-Seq's) (37) peak calling procedure. Distal peaks that are not overlapped with promoters were considered as enhancers.

- 2) Identifying genes' regulatory domains: The definition of gene regulatory domain was borrowed from the GREAT (Genomic Regions Enrichment of Annotations Tool) software (38). Specifically, each gene was assigned a basal regulatory domain, which is the gene's promoter. The gene regulatory domain was then extended in both directions to the nearest gene's basal domain but no more than 1000 kb in one direction.

- 3) Scanning TF binding sites and linking TFs to target genes: We used the H3K27ac peaks called by ENCODE. For each peak, we identified its summit and scanned TF binding sites using FIMO's (Find Individual Motif Occurrences') algorithm (39) with the P value cutoff 1×10^{-5} in the 100-bp interval centered around the summit. In this study, we used the mouse TF motif database curated by the CIS-BP database (12) that contains the motifs of 639 TFs. To link enhancers to their target genes, we first used long-range chromosome interactions to identify the interacting promoters for each enhancer. TF binding sites in these enhancers were then linked to corresponding promoters/genes. For the rest of the predicted TF binding sites of which the assignments cannot be made using the 3D chromosome information, we assigned them to genes according to the regulatory domains defined in previous steps. Last, we formed a directed edge from a TF to a gene in the network if the TF has at least one binding site that is linked to the given gene.

Computing the ranking scores for TFs in the network

We used the personalized PageRank algorithm to calculate the ranking scores for TFs. We first assigned weights to the edges and nodes in TF regulatory networks. The weight of a node was calculated from the relative expression level of its representing gene. A gene's relative expression levels among different cell types were computed by applying the.

z score transformation to its absolute expression levels. Suppose a gene's relative expression level in cell type i is z_i , the node weight for this gene in cell type i was then given by e^{z_i} . The weight of an edge in a network was given by the logarithm of the gene expression level of the source (TF). Let \mathbf{s} be the vector containing node weights and \mathbf{W} be the edge weight matrix. The personalized PageRank score vector \mathbf{v} was calculated by solving a system of linear equations $\mathbf{v} = (1 - d)\mathbf{s} + d\mathbf{W}\mathbf{v}$, where d is the damping factor (default to 0.85). The above equation can be solved in an iterative fashion, i.e., setting $\mathbf{v}_{t+1} = (1 - d)\mathbf{s} + d\mathbf{W}\mathbf{v}_t$.

Validation of the network construction method

Considering the computational feasibility, we only selected 24 of 72 samples to perform validation. They are the embryonic facial prominence at E10.5 and E15.5, forebrain at E10.5 and P0, midbrain at E10.5 and P0, hindbrain at E10.5 and P0, heart at E10.5 and P0, limb at E10.5 and E15.5, stomach at E14.5 and P0, liver at E11.5 and

P0, neural tube at E11.5 and E15.5, kidney at E14.5 and P0, intestine at E14.5 and P0, and lung at E14.5 and P0. For every sample pair, we selected the top 500 most expressed genes from each sample. The log fold changes of these genes are the response variables of our prediction model. According to the constructed network, we identified the upstream regulators of those genes and computed their log fold changes. These were the predictors for the model. These procedures were repeated for all 276 sample pairs, and the results were combined. In total, we got 158,600 records for training a random forest model with 10-fold cross-validation.

Benchmarking different ranking algorithms

The gene expression profiles of wild-type and knockout experiments were simulated using the GeneNetWeaver (17). The yeast and *E. coli* networks were provided by GeneNetWeaver, and additional 20 sub-networks were sampled from the yeast network. Each contains at least 50 regulators of 1000 genes. The formula of WSTG is given by $WSTG(t) = \sum_{i \in \text{regulatees of } t} w_i g_i$, where t represents the TF, g_i is the expression level of gene i , and w_i represents the weight or confidence of the TF gene regulation between TF t and gene i .

Lineage tree construction

Lineage trees were constructed by running the FastME algorithm (18) on normalized ranking matrix, which was obtained from row-wise z score transformation of the original matrix.

Identification of driver TFs for tissues

The output of Taiji pipeline is a matrix, consisting of TFs' ranking scores from different experiments. The rows and columns of the matrix represent TFs and experiments, respectively. To identify driver TFs, we first removed the rows (TFs) with CVs less than 1. We then grouped the columns (experiments) from various stages together whether they are from the same tissue and averaged the ranking scores in each group. Assuming that the average ranking scores are normally distributed, we calculated the deviation from the center for each score and computed the P value. A P value cutoff of 0.01 was used for calling driver TFs.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/3/eaav3262/DC1>

Fig. S1. Validating the Taiji framework.

Fig. S2. The topological properties of genetic networks.

Fig. S3. PageRank scores (color shade) and expression levels (circle size) of 35 constitutively active TFs across 12 tissues and eight stages.

Fig. S4. Identification of driver TFs in 12 tissues.

Fig. S5. Identification of driver TFs in early mouse embryonic development, including two-, four-, and eight-cell stages, ICM, and ESC.

Fig. S6. Selecting algorithms and parameters for clustering analysis.

Fig. S7. Transcriptional waves direct tissue differentiation during mouse embryogenesis.

Table S1. Driver TFs in the heart.

Table S2. Driver TFs in the limb.

Table S3. Driver TFs in the liver.

Table S4. Driver TFs in the lung.

Table S5. Driver TFs in the kidney.

Data file S1. The accession numbers of datasets analyzed in this study.

Data file S2. Germ layer-specific TFs and their P values.

Data file S3. PageRank scores for all TFs across different stages and tissues.

Data file S4. Chromatin interaction predictions made by EpiTensor.

References (40–76)

REFERENCES AND NOTES

1. F. Spitz, E. E. M. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
2. T. M. Filtz, W. K. Vogel, M. Leid, Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol. Sci.* **35**, 76–85 (2014).
3. H. J. Bussemaker, H. Li, E. D. Siggia, Regulatory element detection using correlation with expression. *Nat. Genet.* **27**, 167–174 (2001).
4. T. Schacht, M. Oswald, R. Eils, S. B. Eichmüller, R. König, Estimating the activity of transcription factors by the effect on their target genes. *Bioinformatics* **30**, i401–i407 (2014).
5. M. L. Arrieta-Ortiz, C. Hafemeister, A. R. Bate, T. Chu, A. Greenfield, B. Shuster, S. N. Barry, M. Gallitto, B. Liu, T. Kacmarczyk, F. Santoriello, J. Chen, C. D. A. Rodrigues, T. Sato, D. Z. Rudner, A. Driks, R. Bonneau, P. Eichenberger, An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol. Syst. Biol.* **11**, 839 (2015).
6. E. Zotenko, J. Mestre, D. P. O'Leary, T. M. Przytycka, Why do hubs in the yeast protein interaction network tend to be essential: Reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* **4**, e1000140 (2008).
7. C. Boone, H. Bussey, B. J. Andrews, Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* **8**, 437–449 (2007).
8. J. J. Milner, C. Toma, B. Yu, K. Zhang, K. Omilusik, A. T. Phan, D. Wang, A. J. Getzler, T. Nguyen, S. Crotty, W. Wang, M. E. Pipkin, A. W. Goldrath, Runx3 programs CD8⁺ T cell residency in non-lymphoid tissues and tumours. *Nature* **552**, 253–257 (2017).
9. B. Yu, K. Zhang, J. J. Milner, C. Toma, R. Chen, J. P. Scott-Browne, R. M. Pereira, S. Crotty, J. T. Chang, M. E. Pipkin, W. Wang, A. W. Goldrath, Epigenetic landscapes reveal transcription factors that regulate CD8⁺ T cell differentiation. *Nat. Immunol.* **18**, 573–582 (2017).
10. L. Page, S. Brin, R. Motwani, T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web* (Stanford InfoLab, 1999).
11. Y. Zhu, Z. Chen, K. Zhang, M. Wang, D. Medovoy, J. W. Whitaker, B. Ding, N. Li, L. Zheng, W. Wang, Constructing 3D interaction maps from 1D epigenomes. *Nat. Commun.* **7**, 10812 (2016).
12. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
13. J. M. Mudge, J. Harrow, Creating reference gene annotation for the mouse C57BL/6/J genome assembly. *Mamm. Genome* **26**, 366–378 (2015).
14. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser—A database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
15. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
16. T. D. Cushion, W. B. Dobyns, J. G. L. Mullins, N. Stoodley, S.-K. Chung, A. E. Fry, U. Hehr, R. Gunny, A. S. Aylsworth, P. Prabhakar, G. Uyanik, J. Rankin, M. I. Rees, D. T. Pilz, Overlapping cortical malformations and mutations in *TUBB2B* and *TUBA1A*. *Brain* **136**, 536–548 (2013).
17. T. Schaffter, D. Marbach, D. Floreano, GeneNetWeaver: *In silico* benchmark generation and performance profiling of network inference methods. *Bioinformatics* **27**, 2263–2270 (2011).
18. R. Desper, O. Gascuel, Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J. Comput. Biol.* **9**, 687–705 (2002).
19. J. Wu, B. Huang, H. Chen, Q. Yin, Y. Liu, Y. Xiang, B. Zhang, B. Liu, Q. Wang, W. Xia, W. Li, Y. Li, J. Ma, X. Peng, H. Zheng, J. Ming, W. Zhang, J. Zhang, G. Tian, F. Xu, Z. Chang, J. Na, X. Yang, W. Xie, The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
20. X. Wang, Y. Wang, G. Xiao, J. Wang, L. Zu, M. Hao, X. Sun, Y. Fu, G. Hu, J. Wang, Hypermethylated in cancer 1 (HIC1) suppresses non-small cell lung cancer progression by targeting interleukin-6/Stat3 pathway. *Oncotarget* **7**, 30350–30364 (2016).
21. J. L. Simmons, C. J. Pierce, F. Al-Ejeh, G. M. Boyle, MITF and BRN2 contribute to metastatic growth after dissemination of melanoma. *Sci. Rep.* **7**, 10909 (2017).
22. P. Minoo, G. Su, H. Drum, P. Bringas, S. Kimura, Defects in tracheoesophageal and lung morphogenesis in *Nkx2.1*(^{-/-}) mouse embryos. *Dev. Biol.* **209**, 60–71 (1999).
23. H. Wan, S. Dingle, Y. Xu, V. Besnard, K. H. Kaestner, S.-L. Ang, S. Wert, M. T. Stahlman, J. A. Whitsett, Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J. Biol. Chem.* **280**, 13809–13816 (2005).
24. Z. Shi, D. Jervis, P. E. B. Nickerson, R. L. Chow, Requirement for the paired-like homeodomain transcription factor VSX1 in type 3a mouse retinal bipolar cell terminal differentiation. *J. Comp. Neurol.* **520**, 117–129 (2012).
25. C. S. Merzdorf, Emerging roles for *zic* genes in early development. *Dev. Dyn.* **236**, 922–940 (2007).

26. L. S. Gammill, H. Sive, *otx2* expression in the ectoderm activates anterior neural determination and is required for *Xenopus* cement gland formation. *Dev. Biol.* **240**, 223–236 (2001).
27. X. Zhang, C. T. Huang, J. Chen, M. T. Pankratz, J. Xi, J. Li, Y. Yang, T. M. LaVaute, X.-J. Li, M. Ayala, G. I. Bondarenko, Z.-W. Du, Y. Jin, T. G. Golos, S.-C. Zhang, *Pax6* is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90–100 (2010).
28. I. Burtscher, H. Lickert, *Foxa2* regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development* **136**, 1029–1038 (2009).
29. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
30. Z.-J. Lan, A. C.-K. Chung, X. Xu, F. J. DeMayo, A. J. Cooney, The embryonic function of germ cell nuclear factor is dependent on the DNA binding domain. *J. Biol. Chem.* **277**, 50660–50667 (2002).
31. P. Stein, S. Medvedev, R. M. Schultz, OBOX proteins are recruited during oocyte maturation and are essential for early development in mouse. *Biol. Reprod.* **87**, 210 (2012).
32. N.-E. Szabó, T. Zhao, X. Zhou, G. Alvarez-Bolado, The role of Sonic hedgehog of neural origin in thalamic differentiation in the mouse. *J. Neurosci.* **29**, 2453–2466 (2009).
33. B. Lucas, K. Grigo, S. Erdmann, J. Lausen, L. Klein-Hitpass, G. U. Ryffel, *HNf4a* reduces proliferation of kidney cells and affects genes deregulated in renal cell carcinoma. *Oncogene* **24**, 6418–6431 (2005).
34. W. Garrison, M. A. Battle, C. Yang, K. H. Kaestner, F. M. Sladek, S. A. Duncan, Hepatocyte nuclear factor 4a is essential for embryonic development of the mouse colon. *Gastroenterology* **130**, 1207–1220 (2006).
35. A. DeLaForest, M. Nagaoka, K. Si-Tayeb, F. K. Noto, G. Konopka, M. A. Battle, S. A. Duncan, *HNf4A* is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development* **138**, 4143–4153 (2011).
36. D. S. Andersen, J. Colombani, P. Léopold, Coordination of organ growth: Principles and outstanding questions from the world of insects. *Trends Cell Biol.* **23**, 336–344 (2013).
37. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
38. C. Y. McLean, D. Bristol, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
39. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
40. S.-Y. Oh, J. Y. Kim, C. Park, The ETS factor, *ETV2*: A master regulator for vascular endothelial cell development. *Mol. Cells* **38**, 1029–1036 (2015).
41. T. Li, C. Liu, Y. Xu, Q. Guo, S. Chen, K. Sun, R. Xu, Identification of candidate genes for congenital heart defects on proximal chromosome 8p. *Sci. Rep.* **6**, 36133 (2016).
42. J. J. Connelly, T. Wang, J. E. Cox, C. Haynes, L. Wang, S. H. Shah, D. R. Crosslin, A. B. Hale, S. Nelson, D. C. Crossman, C. B. Granger, J. L. Haines, C. J. H. Jones, J. M. Vance, P. J. Goldschmidt-Clermont, W. E. Kraus, E. R. Hauser, S. G. Gregory, *GATA2* is associated with familial early-onset coronary artery disease. *PLOS Genet.* **2**, e139 (2006).
43. C. T. Kuo, E. E. Morrisey, R. Anandappa, K. Sigrist, M. M. Lu, M. S. Parmacek, C. Soudais, J. M. Leiden, *GATA4* transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **11**, 1048–1060 (1997).
44. R. J. Schwartz, E. N. Olson, Building the heart piece by piece: Modularity of *cis*-elements regulating *Nkx2-5* transcription. *Development* **126**, 4187–4192 (1999).
45. X. Cai, W. Zhang, J. Hu, L. Zhang, N. Sultana, B. Wu, W. Cai, B. Zhou, C.-L. Cai, *Tbx20* acts upstream of *Wnt* signaling to regulate endocardial cushion formation and valve remodeling during mouse cardiogenesis. *Development* **140**, 3176–3187 (2013).
46. D.-H. Kwon, G. H. Eom, H. J. Kee, Y. S. Nam, Y. K. Cho, D.-K. Kim, J. Y. Koo, H.-S. Kim, K.-I. Nam, K. K. Kim, I.-K. Lee, S. B. Park, H.-S. Choi, H. Kook, Estrogen-related receptor gamma induces cardiac hypertrophy by activating *GATA4*. *J. Mol. Cell. Cardiol.* **65**, 88–97 (2013).
47. B. G. Bruneau, Z.-Z. Bao, D. Fatkin, J. Xavier-Neto, D. Georgakopoulos, C. T. Maguire, C. I. Berul, D. A. Kass, M. L. Kuroski-de Bold, A. J. de Bold, D. A. Conner, N. Rosenthal, C. L. Cepko, C. E. Seidman, J. G. Seidman, Cardiomyopathy in *Irx4*-deficient mice is preceded by abnormal ventricular gene expression. *Mol. Cell. Biol.* **21**, 1730–1736 (2001).
48. S. Ohgo, A. Itoh, M. Suzuki, A. Satoh, H. Yokoyama, K. Tamura, Analysis of *hoxa11* and *hoxa13* expression during patternless limb regeneration in *Xenopus*. *Dev. Biol.* **338**, 148–157 (2009).
49. Y. Hérault, S. Hrabá-Renevey, F. van der Hoeven, D. Duboule, Function of the *Evx-2* gene in the morphogenesis of vertebrate limbs. *EMBO J.* **15**, 6727–6738 (1996).
50. V. Bensoussan-Trigano, Y. Lallemand, C. S. Cloment, B. Robert, *Msx1* and *Msx2* in limb mesenchyme modulate digit number and identity. *Dev Dyn.* **240**, 1190–1202 (2011).
51. K. Koishi, M. Zhang, I. S. McLennan, A. J. Harris, MyoD protein accumulates in satellite cells and is neurally regulated in regenerating myotubes and skeletal muscle fibers. *Dev. Dyn.* **202**, 244–254 (1995).
52. E. Meadows, J.-H. Cho, J. M. Flynn, W. H. Klein, Myogenin regulates a distinct genetic program in adult muscle stem cells. *Dev. Biol.* **322**, 406–414 (2008).
53. E. Havis, P. Coumailleau, A. Bonnet, K. Bismuth, M.-A. Bonnin, R. Johnson, C.-M. Fan, F. Relaix, D.-L. Shi, D. Duprez, *Sim2* prevents entry into the myogenic program by repressing *MyoD* transcription during limb embryonic myogenesis. *Development* **139**, 1910–1920 (2012).
54. A. C. Tufan, K. M. Daumer, A. M. DeLise, R. S. Tuan, AP-1 transcription factor complex is a target of signals from both *WNT-7a* and *N-cadherin*-dependent cell–cell adhesion complex during the regulation of limb mesenchymal chondrogenesis. *Exp. Cell Res.* **273**, 197–203 (2002).
55. L. Vassen, H. Beauchemin, W. Lemsaddek, J. Krongold, M. Trudel, T. Möröy, *Growth factor independence 1b (Gfi1b)* is important for the maturation of erythroid cells and the regulation of embryonic globin expression. *PLOS ONE* **9**, e96636 (2014).
56. Q. Huo, C. Ge, H. Tian, J. Sun, M. Cui, H. Li, F. Zhao, T. Chen, H. Xie, Y. Cui, M. Yao, J. Li, Dysfunction of *IKZF1/MYC/MDIG* axis contributes to liver cancer progression through regulating *H3K9me3/p21* activity. *Cell Death Dis.* **8**, e2766 (2017).
57. Y. Hirotsu, N. Hataya, F. Katsuoka, M. Yamamoto, NF-E2-related factor 1 (*Nrf1*) serves as a novel regulator of hepatic lipid metabolism through regulation of the *Lipin1* and *PGC-1β* genes. *Mol. Cell. Biol.* **32**, 2760–2770 (2012).
58. L. Pevny, C. S. Lin, V. D'Agati, M. C. Simon, S. H. Orkin, F. Costantini, Development of hematopoietic cells lacking transcription factor *GATA-1*. *Dev. Camb. Engl.* **121**, 163–172 (1995).
59. Y. Y. Yien, J. J. Bieker, *EKLF/KLF1*, a tissue-restricted integrator of transcriptional control, chromatin remodeling, and lineage determination. *Mol. Cell. Biol.* **33**, 4–13 (2013).
60. M. Kohyama, W. Ise, B. T. Edelson, P. R. Wilker, K. Hildner, C. Mejia, W. A. Frazier, T. L. Murphy, K. M. Murphy, Role for *Sp1-C* in the development of red pulp macrophages and splenic iron homeostasis. *Nature* **457**, 318–321 (2008).
61. I. Delgado, O. Fresnedo, A. Iglesias, Y. Rueda, W.-K. Syn, A. M. Zubiaga, B. Ochoa, A role for transcription factor *E2F2* in hepatocyte proliferation and timely liver regeneration. *Am. J. Physiol. Gastrointest. Liver Physiol.* **301**, G20–G31 (2011).
62. Y. Xu, Y. Wang, V. Besnard, M. Ikegami, S. E. Wert, C. Heffner, S. A. Murray, L. R. Donahue, J. A. Whitsett, Transcriptional programs controlling perinatal lung maturation. *PLOS ONE* **7**, e37046 (2012).
63. E. Gao, Y. Wang, J. L. Alcorn, C. R. Mendelson, Transcription factor *USF2* is developmentally regulated in fetal lung and acts together with *USF1* to induce *SP-A* gene expression. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **284**, L1027–L1036 (2003).
64. B. Chen, T. V. Moore, Z. Li, A. I. Sperling, C. Zhang, J. Andrade, A. Rodriguez, N. Bahroos, Y. Huang, E. E. Morrisey, P. J. Gruber, J. Solway, *Gata5* deficiency causes airway constrictor hyperresponsiveness in mice. *Am. J. Respir. Cell Mol. Biol.* **50**, 787–795 (2014).
65. C. Chung, T. Kim, M. Kim, M. Kim, H. Song, T.-S. Kim, E. Seo, S.-H. Lee, H. Kim, S. K. Kim, G. Yoo, D.-H. Lee, D.-S. Hwang, T. Kinashi, J.-M. Kim, D.-S. Lim, *Hippo-Foxa2* signaling pathway plays a role in peripheral lung maturation and surfactant homeostasis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7732–7737 (2013).
66. Z. Yin, L. Gonzales, V. Kolla, N. Rath, Y. Zhang, M. M. Lu, S. Kimura, P. L. Ballard, M. F. Beers, J. A. Epstein, E. E. Morrisey, *Hop* functions downstream of *Nkx2.1* and *GATA6* to mediate *HDAC*-dependent negative regulation of pulmonary gene expression. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **291**, L191–L199 (2006).
67. M. Herriges, E. E. Morrisey, Lung development: Orchestrating the generation and regeneration of a complex organ. *Development* **141**, 502–513 (2014).
68. D. E. Metzger, M. T. Stahlman, J. M. Shannon, *Misexpression of ELF5* disrupts lung branching and inhibits epithelial differentiation. *Dev. Biol.* **320**, 149–160 (2008).
69. R. Arora, R. J. Metzger, V. E. Papaioannou, Multiple roles and interactions of *Tbx4* and *Tbx5* in development of the respiratory system. *PLOS Genet.* **8**, e1002866 (2012).
70. A. D. Bird, S. J. Flecknoe, K. H. Tan, P. F. Olsson, N. Antony, T. Mantamadiotis, R. Mollard, S. B. Hooper, T. J. Cole, cAMP response element binding protein is required for differentiation of respiratory epithelium during murine development. *PLOS ONE* **6**, e17843 (2011).
71. N. Di-Poi, J. Zákány, D. Duboule, Distinct roles and regulations for *Hoxd* genes in metanephric kidney development. *PLOS Genet.* **3**, e232 (2007).
72. P.-Y. Lam, C. N. Kamei, S. Mangos, S. Mudumana, Y. Liu, I. A. Drummond, *Odd-skipped related 2* is required for fin chondrogenesis in zebrafish. *Dev. Dyn.* **242**, 1284–1292 (2013).
73. A. Teng, M. Nair, J. Wells, J. A. Segre, X. Dai, Strain-dependent perinatal lethality of *Ovo1*-deficient mice and identification of *Ovo2* as a downstream target of *Ovo1* in skin epidermis. *Biochim. Biophys. Acta* **1772**, 89–95 (2007).
74. S. Adalat, A. S. Woolf, K. A. Johnstone, A. Wirsing, L. W. Harries, D. A. Long, R. C. Hennekam, S. E. Ledermann, L. Rees, W. van't Hoff, S. D. Marks, R. S. Trompeter, K. Tullus, P. J. Winyard, J. Cansick, I. Mushtaq, H. K. Dhillon, C. Bingham, E. L. Edghill,

- R. Shroff, H. Stanescu, G. U. Ryffel, S. Ellard, D. Bockenhauer, *HNFB* mutations associate with hypomagnesemia and renal magnesium wasting. *J. Am. Soc. Nephrol.* **20**, 1123–1131 (2009).
75. E. W. Brunskill, H. L. Lai, D. C. Jamison, S. S. Potter, L. T. Patterson, Microarrays and RNA-Seq identify molecular mechanisms driving the end of nephron production. *BMC Dev. Biol.* **11**, 15 (2011).
76. D. M. Wellik, *Hox* genes and kidney development. *Pediatr. Nephrol.* **26**, 1559–1565 (2011).

Acknowledgments

Funding: This project is partially supported by NIH (U54HG006997 and R01HG009626) and CIRM (RB5 07012). **Author contributions:** K.Z. and W.W. conceived the study and interpreted the results. K.Z. designed the Taiji software. K.Z. and M.W. performed the data analysis with contributions from Y.Z. K.Z. and W.W. wrote the manuscript with contributions from M.W. and Y.Z. W.W. supervised the project. **Competing interests:** The authors declare that they have no

competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All data analyzed during this study can be found in ENCODE portal or GEO database. A full list of accession numbers is included in data file S1. The Taiji software is available at <http://wanglab.ucsd.edu/star/taiji>. Additional data related to this paper may be requested from the authors.

Submitted 5 September 2018

Accepted 30 January 2019

Published 27 March 2019

10.1126/sciadv.aav3262

Citation: K. Zhang, M. Wang, Y. Zhao, W. Wang, Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci. Adv.* **5**, eaav3262 (2019).