Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/3/eaav3262/DC1

Supplementary Materials for

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

Kai Zhang, Mengchi Wang, Ying Zhao, Wei Wang*

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

Published 27 March 2019, *Sci. Adv.* **5**, eaav3262 (2019)
DOI: 10.1126/sciadv.aav3262

The PDF file includes:

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.
References (*40*–*76*)

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/3/eaav3262/DC1)

Data file S1 (Microsoft Excel format). The accession numbers of datasets analyzed in this study.
Data file S2 (Microsoft Excel format). Germ layer–specific TFs and their *P* values.
Data file S3 (Microsoft Excel format). PageRank scores for all TFs across different stages and tissues.
Data file S4 (.zip format). Chromatin interaction predictions made by EpiTensor.

Supplementary Materials

Literature evidence supports identified driver TFs

Here we list the supporting evidence for the identified driver TFs in five tissues, including heart, limb, liver, lung and kidney. The rest of the tissues (neural tube, forebrain, hindbrain, midbrain, craniofacial, stomach and intestine) are either less studied or closely related to each other (functionally and spatially), which makes it very difficult to assess their tissue specificity.

8 out of 11 (72.7%) predicted driver TFs in heart are supported by literature (Table S1). 8 out of 10 (80%) predicted driver TFs in limb are supported by literature (Table S2). 7 out of 11 (63.6%) predicted driver TFs in liver are supported by literature (Table S3). 11 out of 13 (84.6%) predicted driver TFs in lung are supported by literature (Table S4). 7 out of 10 (70%) predicted driver TFs in kidney are supported by literature (Table S5).

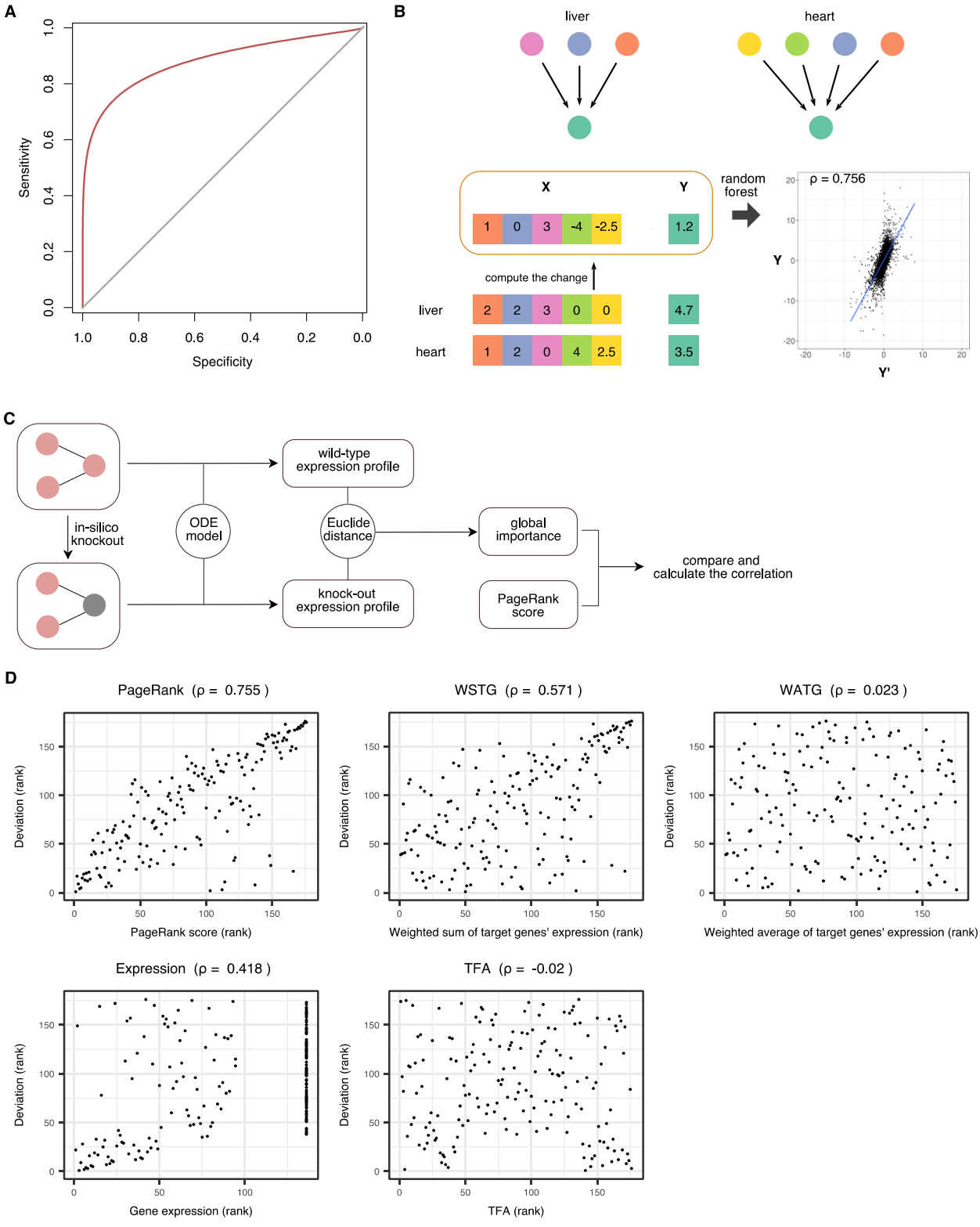

A
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
B
liver
heart
X
Y
1 0 3 -4 -2.5
1.2
random forest
ρ = 0.756
compute the change
liver
2 2 3 0 0
4.7
heart
1 2 0 4 2.5
3.5
Y'
C
wild-type expression profile
in-silico knockout
ODE model
Euclide distance
global importance
compare and calculate the correlation
knock-out expression profile
PageRank score
D
PageRank (ρ = 0.755)
WSTG (ρ = 0.571)
WATG (ρ = 0.023)
Deviation (rank)
PageRank score (rank)
Weighted sum of target genes' expression (rank)
Weighted average of target genes' expression (rank)
Expression (ρ = 0.418)
TFA (ρ = -0.02)
Gene expression (rank)
TFA (rank)

Fig. S1. Validating the Taiji framework. (A) Comparing EpiTensor's predictions with Hi-C interactions in mESC. The AUC is 0.87. **(B)** An illustration of the network validation process. The log-fold expression changes of genes and their regulators between two tissues were computed and used to train a random forest model. The performance was assessed by the Spearman's correlation between predictions and "ground truth" with 10-fold cross-validation. The average correlation is 0.756. **(C)** The workflow of the in-silico benchmark process for assessing PageRank's performance on predicting TFs' global importance. **(D)** Comparing different ranking methods in E. coli network.

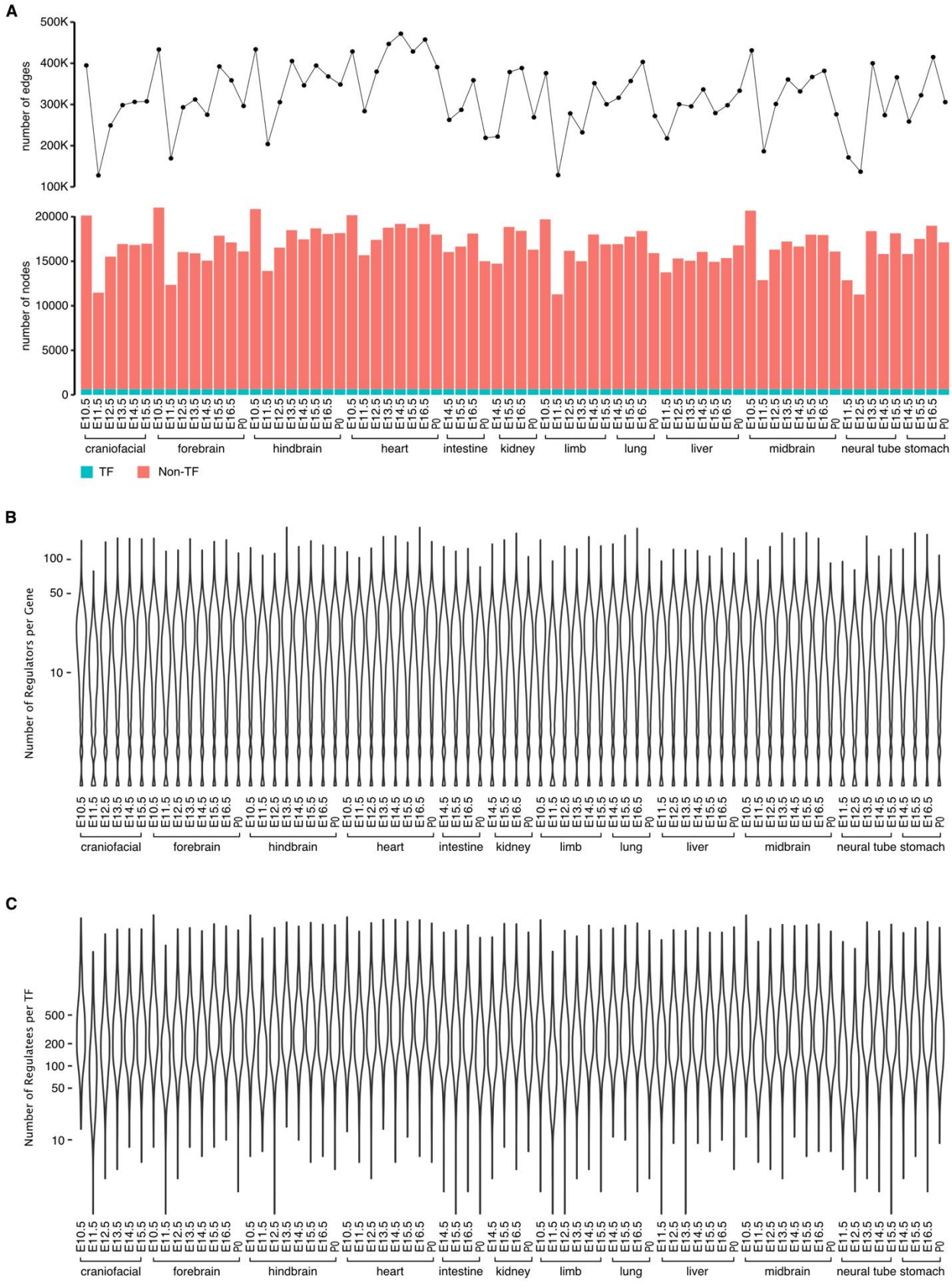
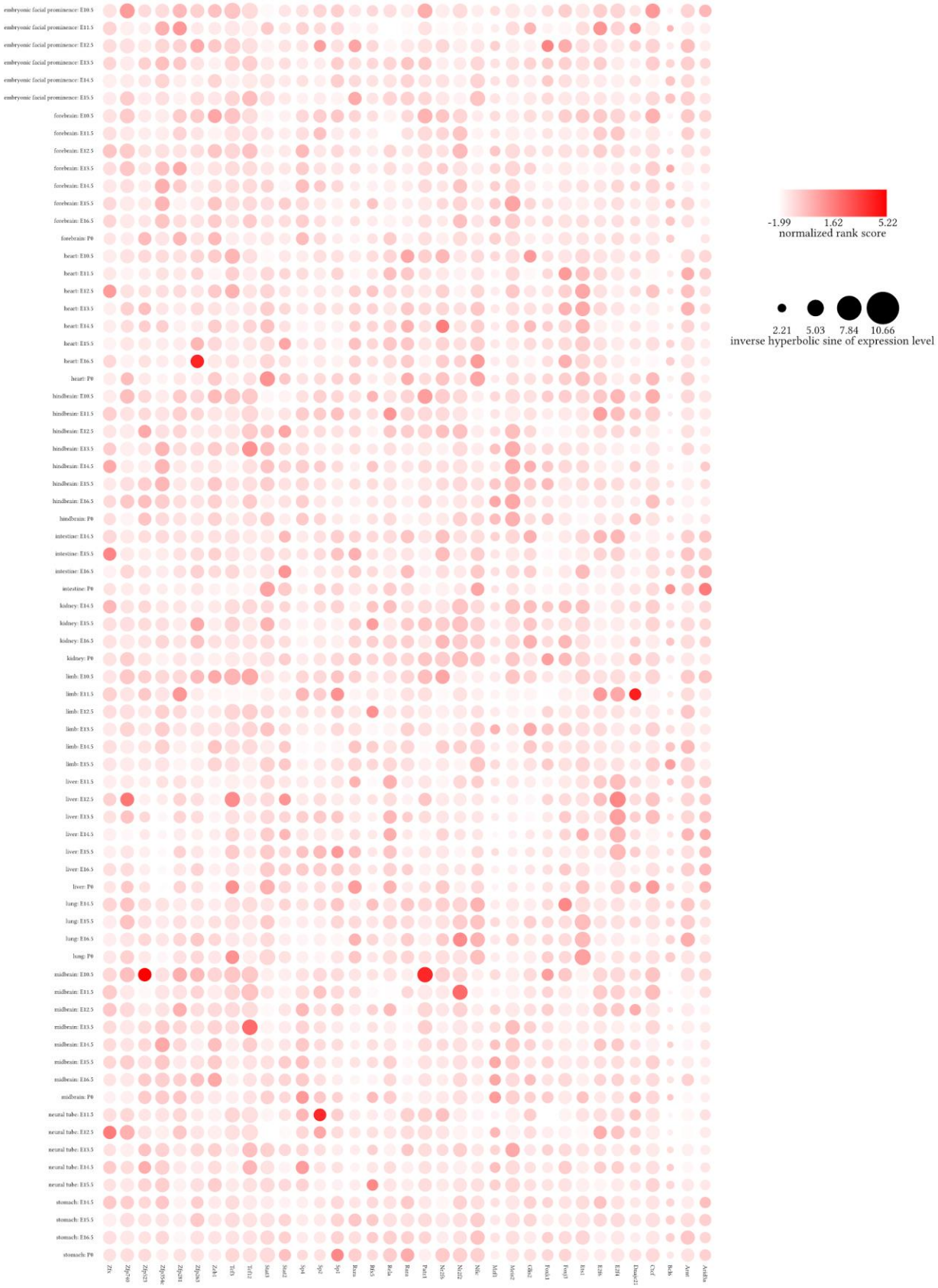

A
number of edges
500K
400K
300K
200K
100K
number of nodes
20000
15000
10000
5000
0
TF
Non-TF
B
Number of Regulators per Gene
100
50
10
C
Number of Regulatees per TF
500
200
100
50
10
craniofacial
forebrain
hindbrain
heart
intestine
kidney
limb
lung
liver
midbrain
neural tube
stomach

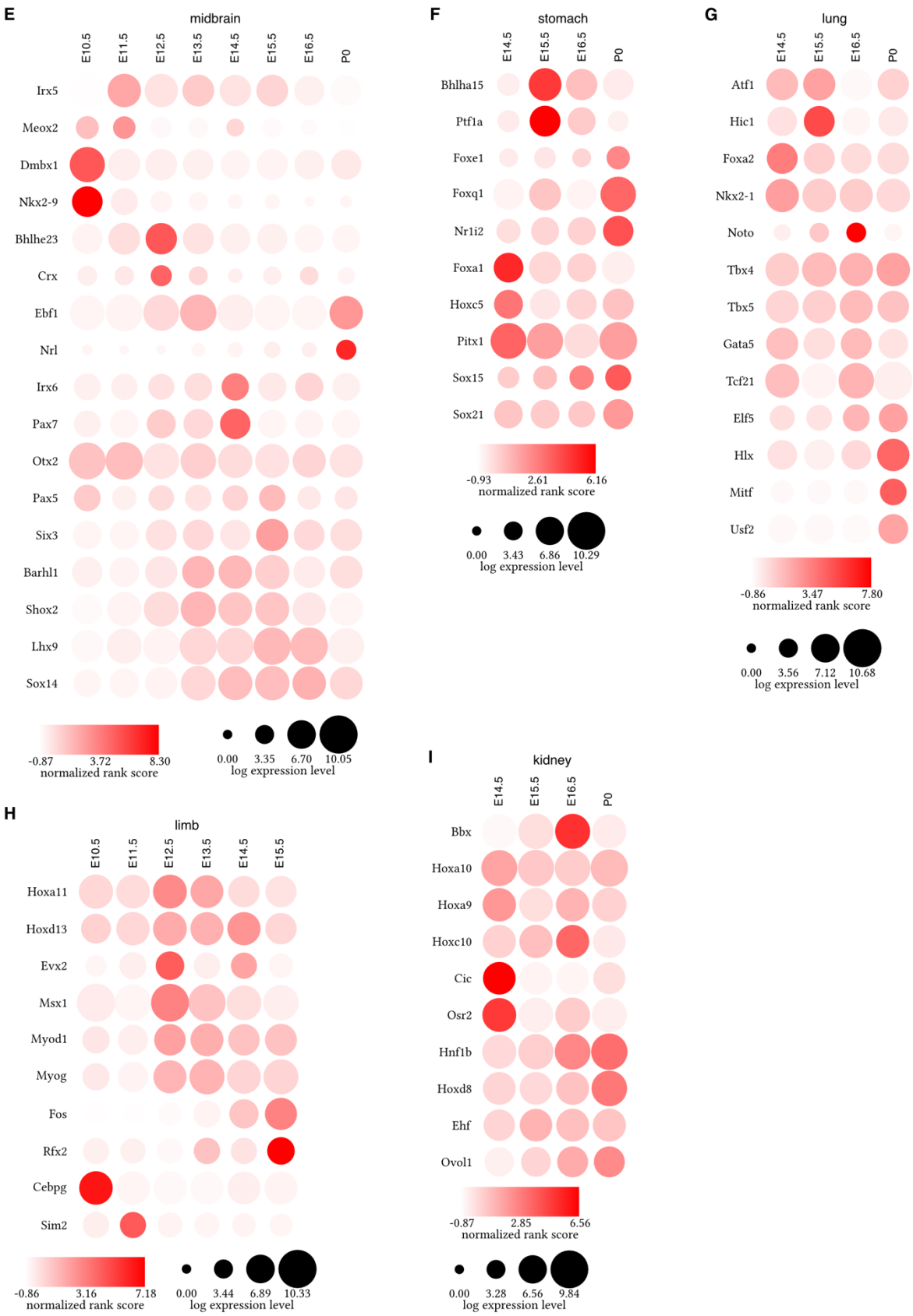
Fig. S2. The topological properties of genetic networks. (A) The number of edges (top) and nodes (bottom) in each genetic network. **(B)** The distribution of the number of regulators per gene for each genetic network. **(C)** The distribution of the number of regulatees per TF for each genetic network.



embryonic facial prominence: E10.5
embryonic facial prominence: E11.5
embryonic facial prominence: E12.5
embryonic facial prominence: E13.5
embryonic facial prominence: E14.5
embryonic facial prominence: E15.5
forebrain: E10.5
forebrain: E11.5
forebrain: E12.5
forebrain: E13.5
forebrain: E14.5
forebrain: E15.5
forebrain: E16.5
forebrain: P0
heart: E10.5
heart: E11.5
heart: E12.5
heart: E13.5
heart: E14.5
heart: E15.5
heart: E16.5
heart: P0
hindbrain: E10.5
hindbrain: E11.5
hindbrain: E12.5
hindbrain: E13.5
hindbrain: E14.5
hindbrain: E15.5
hindbrain: E16.5
hindbrain: P0
intestine: E14.5
intestine: E15.5
intestine: E16.5
intestine: P0
kidney: E14.5
kidney: E15.5
kidney: E16.5
kidney: P0
limb: E10.5
limb: E11.5
limb: E12.5
limb: E13.5
limb: E14.5
limb: E15.5
liver: E11.5
liver: E12.5
liver: E13.5
liver: E14.5
liver: E15.5
liver: E16.5
liver: P0
lung: E14.5
lung: E15.5
lung: E16.5
lung: P0
midbrain: E10.5
midbrain: E11.5
midbrain: E12.5
midbrain: E13.5
midbrain: E14.5
midbrain: E15.5
midbrain: E16.5
midbrain: P0
neural tube: E11.5
neural tube: E12.5
neural tube: E13.5
neural tube: E14.5
neural tube: E15.5
stomach: E14.5
stomach: E15.5
stomach: E16.5
stomach: P0
Zfx
Zfp740
Zfp523
Zfp354c
Zfp281
Zfp263
Zeb1
Tcf3
Tcf12
Stat3
Stat2
Sp4
Sp2
Sp1
Rxra
Rfx5
Rela
Rara
Pat1
Nr2f6
Nr2f2
Nfic
Mef1
Mesi2
Glis2
Foxk1
Foxj3
Etv1
Ezh6
Ezh1
Dnaje21
Ccf
Bdk
Arnt
Ariafa
-1.99
1.62
5.22
normalized rank score
2.21
5.03
7.84
10.66
inverse hyperbolic sine of expression level

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

A
forebrain
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
Dmrt3
Dmrta1
Hsfy2
Fezf1
Dmrt1
Gsx2
Mesp1
Egr4
Tbx21
Dlx1
Dlx2
Dlx5
Bhlhe22
Neurod2
Nr4a3
Lhx2
Nr2e1
Sp9
Foxg1
Lhx8
Arx
Emx1
Eomes
Tbr1
Cux1
Trp73
-0.97 3.69 8.34
normalized rank score
0.00 3.31 6.62 9.93
log expression level
B
neural tube
E11.5
E12.5
E13.5
E14.5
E15.5
Hoxc12
Hoxc13
Gbx1
Hoxc8
Hoxa7
Hoxb6
Hoxc9
Mnx1
Pax2
Lhx4
Tlx3
Cdc5l
Hoxa1
Hoxc6
Sohlh2
Tcfl5
Rora
Zfp128
-0.87 3.70 8.27
normalized rank score
0.00 3.12 6.24 9.36
log expression level
C
intestine
E14.5
E15.5
E16.5
P0
Bhlhe40
Hes2
Irf7
Nr2e3
Mafg
Elf3
Isx
Nkx2-3
Pax4
Cdx2
Hnf4g
Rfx6
-0.84 3.53 7.90
normalized rank score
0.00 3.37 6.74 10.10
log expression level
D
hindbrain
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
Lmx1b
Mafb
Pou2f1
Hoxa5
Evx1
Hoxa2
Msx3
Hoxb1
Olig3
Lcor
Vsx1
-0.79 3.50 7.78
normalized rank score
0.00 3.26 6.51 9.77
log expression level


E
midbrain
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
Irx5
Meox2
Dmbx1
Nkx2-9
Bhlhe23
Crx
Ebf1
Nrl
Irx6
Pax7
Otx2
Pax5
Six3
Barhl1
Shox2
Lhx9
Sox14
-0.87 3.72 8.30
normalized rank score
0.00 3.35 6.70 10.05
log expression level
F
stomach
E14.5
E15.5
E16.5
P0
Bhlha15
Ptf1a
Foxe1
Foxq1
Nr1i2
Foxa1
Hoxc5
Pitx1
Sox15
Sox21
-0.93 2.61 6.16
normalized rank score
0.00 3.43 6.86 10.29
log expression level
G
lung
E14.5
E15.5
E16.5
P0
Atf1
Hic1
Foxa2
Nkx2-1
Noto
Tbx4
Tbx5
Gata5
Tcf21
Elf5
Hlx
Mitf
Usf2
-0.86 3.47 7.80
normalized rank score
0.00 3.56 7.12 10.68
log expression level
H
limb
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
Hoxa11
Hoxd13
Evx2
Msx1
Myod1
Myog
Fos
Rfx2
Cebpg
Sim2
-0.86 3.16 7.18
normalized rank score
0.00 3.44 6.89 10.33
log expression level
I
kidney
E14.5
E15.5
E16.5
P0
Bbx
Hoxa10
Hoxa9
Hoxc10
Cic
Osr2
Hnf1b
Hoxd8
Ehf
Ovol1
-0.87 2.85 6.56
normalized rank score
0.00 3.28 6.56 9.84
log expression level

J
heart
e10.5
e11.5
e12.5
e13.5
e14.5
e15.5
e16.5
p0
Etv2
Esx1
Nkx2-6
Esrrg
Spz1
Gata2
Bhlhe41
Irx4
Gata4
Nkx2-5
Tbx20
-0.78
3.80
8.37
normalized rank score
0.00
3.35
6.70
10.05
log expression level
K
liver
e11.5
e12.5
e13.5
e14.5
e15.5
e16.5
p0
Foxh1
E2f2
Hesx1
Gfi1b
Nr3c2
Ikzf1
Nfe2
Gata1
Klf1
Spic
Zfp652
-0.82
3.74
8.30
normalized rank score
0.00
3.58
7.17
10.75
log expression level
L
craniofacial prominence
e10.5
e11.5
e12.5
e13.5
e14.5
e15.5
Irx2
Myf6
Pou1f1
Sry
Tbx19
Hmga2
Alx1
Glis1
Pax9
Prdm11
Dlx3
Prop1
Zfp3
Zscan10
-0.89
3.74
8.37
normalized rank score
0.00
3.40
6.80
10.20
log expression level

Fig. S4. Identification of driver TFs in 12 tissues. (A) forebrain. **(B)** neural tube. **(C)** intestine. **(D)** hindbrain. **(E)** midbrain. **(F)** stomach. **(G)** lung. **(H)** limb. **(I)** kidney. **(J)** heart. **(K)** liver. **(L)** craniofacial prominence.

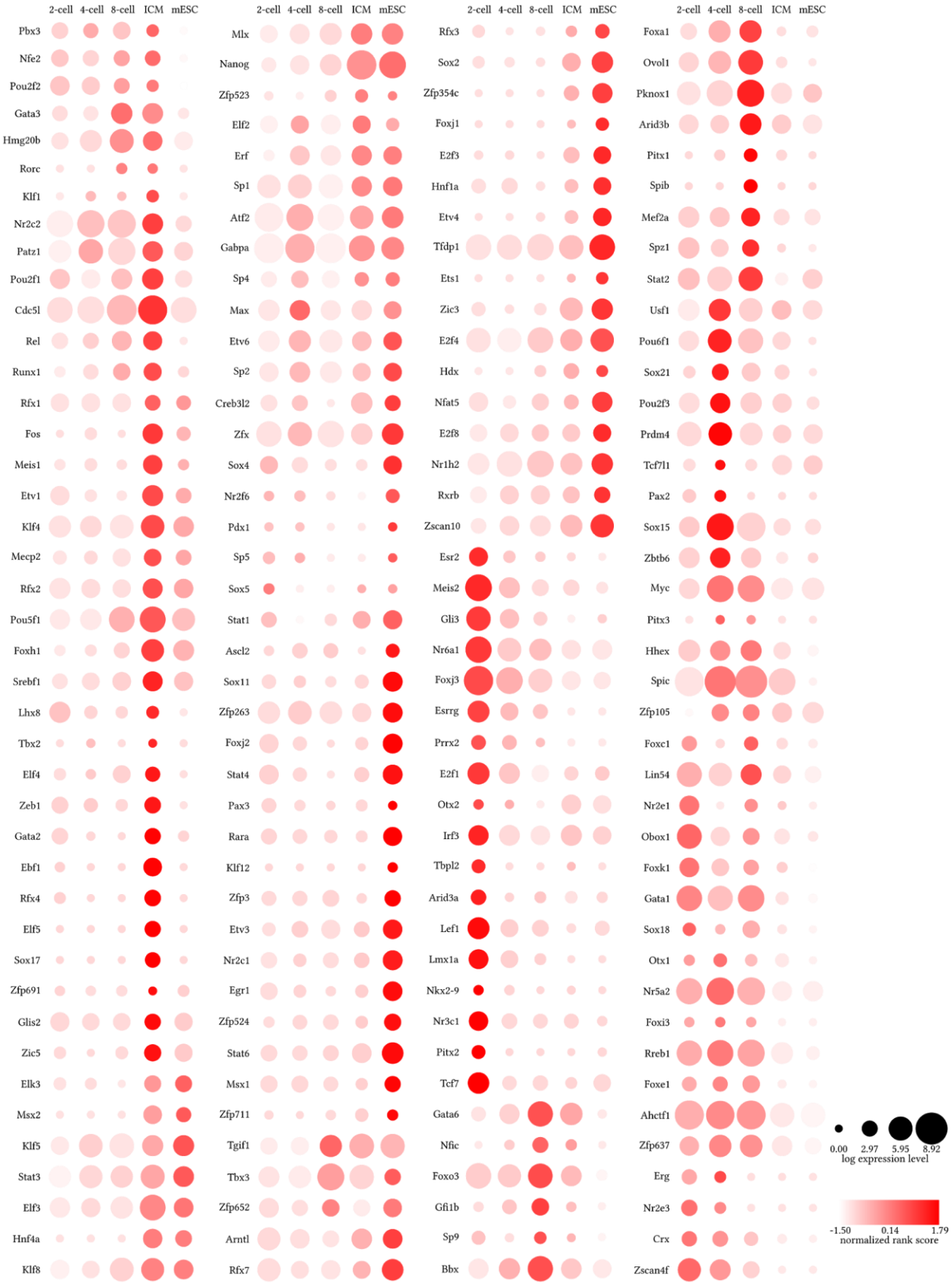
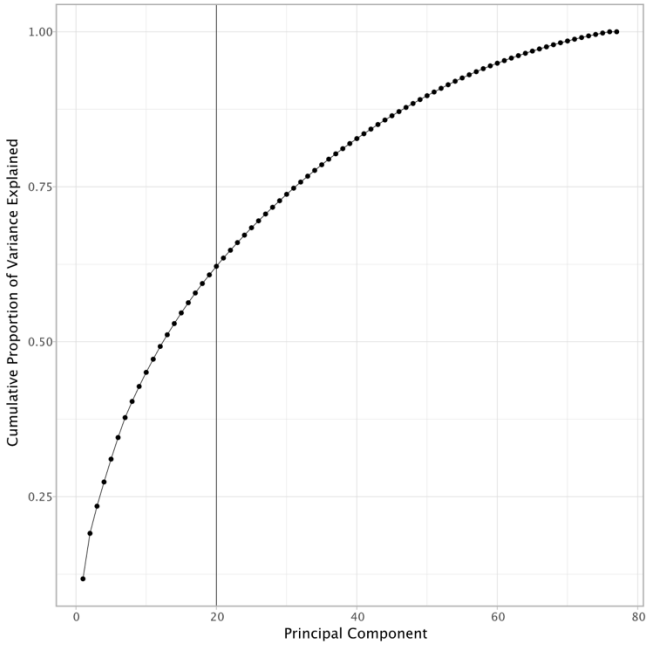

2-cell 4-cell 8-cell ICM mESC
Pbx3 Nfe2 Pou2f2 Gata3 Hmg20b Rorc Klf1 Nr2c2 Patz1 Pou2f1 Cdc5l Rel Runx1 Rfx1 Fos Meis1 Etv1 Klf4 Mecp2 Rfx2 Pou5f1 Foxh1 Srebf1 Lhx8 Tbx2 Elf4 Zeb1 Gata2 Ebf1 Rfx4 Elf5 Sox17 Zfp691 Glis2 Zic5 Elk3 Msx2 Klf5 Stat3 Elf3 Hnf4a Klf8
2-cell 4-cell 8-cell ICM mESC
Mlx Nanog Zfp523 Elf2 Erf Sp1 Atf2 Gabpa Sp4 Max Etv6 Sp2 Creb3l2 Zfx Sox4 Nr2f6 Pdx1 Sp5 Sox5 Stat1 Ascl2 Sox11 Zfp263 Foxj2 Stat4 Pax3 Rara Klf12 Zfp3 Etv3 Nr2c1 Egr1 Zfp524 Stat6 Msx1 Zfp711 Tgif1 Tbx3 Zfp652 Arntl Rfx7
2-cell 4-cell 8-cell ICM mESC
Rfx3 Sox2 Zfp354c Foxj1 E2f3 Hnf1a Etv4 Tfdp1 Ets1 Zic3 E2f4 Hdx Nfat5 E2f8 Nr1h2 Rxrb Zscan10 Esr2 Meis2 Gli3 Nr6a1 Foxj3 Esrrg Prrx2 E2f1 Otx2 Irf3 Tbpl2 Arid3a Lef1 Lmx1a Nkx2-9 Nr3c1 Pitx2 Tcf7 Gata6 Nfic Foxo3 Gfi1b Sp9 Bbx
2-cell 4-cell 8-cell ICM mESC
Foxa1 Ovol1 Pknox1 Arid3b Pitx1 Spib Mef2a Spz1 Stat2 Usf1 Pou6f1 Sox21 Pou2f3 Prdm4 Tcf7l1 Pax2 Sox15 Zbtb6 Myc Pitx3 Hhex Spic Zfp105 Foxc1 Lin54 Nr2e1 Obox1 Foxk1 Gata1 Sox18 Otx1 Nr5a2 Foxi3 Rreb1 Foxe1 Ahctf1 Zfp637 Erg Nr2e3 Crx Zscan4f
0.00 2.97 5.95 8.92
log expression level
-1.50 0.14 1.79
normalized rank score

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

A


Cumulative Proportion of Variance Explained
1.00
0.75
0.50
0.25
0
20
40
60
80
Principal Component

B

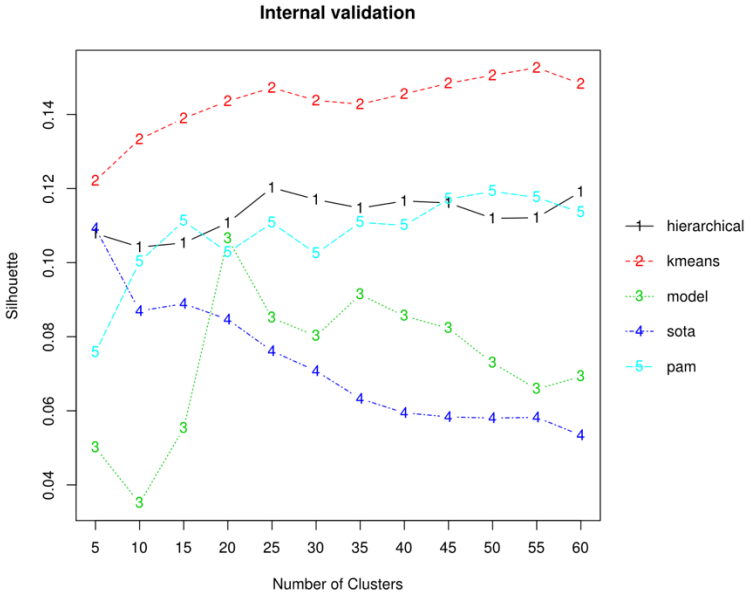
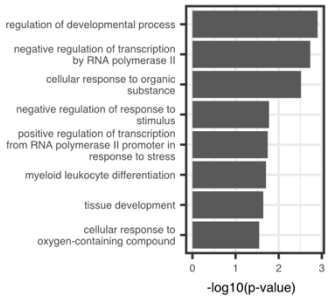
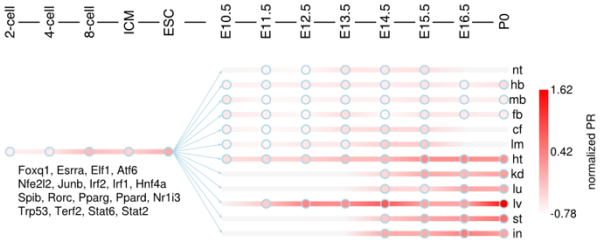

Internal validation
Silhouette
0.14
0.12
0.10
0.08
0.06
0.04
5
10
15
20
25
30
35
40
45
50
55
60
Number of Clusters
hierarchical
kmeans
model
sota
pam

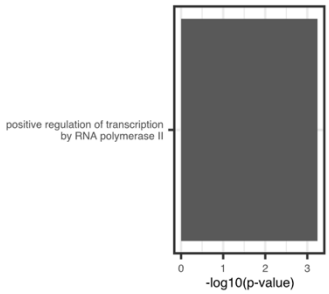
Fig. S6. Selecting algorithms and parameters for clustering analysis. (A) Plotting the cumulative proportion of variance explained against the number of principal components (PCs). The first 20 PCs were kept according to the "elbow" method. **(B)** Selecting the best clustering algorithm and number of clusters according to the Silhouette metric. The k-means algorithm (k = 25) was picked using the "elbow" method. hierarchical, hierarchical clustering algorithm; kmeans, k-means algorithm; model, model-based algorithm; sota, self-organizing tree algorithm; pam, partitioning around medoids algorithm.

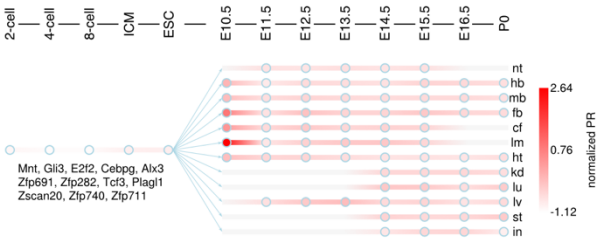
C1


regulation of developmental process
negative regulation of transcription by RNA polymerase II
cellular response to organic substance
negative regulation of response to stimulus
positive regulation of transcription from RNA polymerase II promoter in response to stress
myeloid leukocyte differentiation
tissue development
cellular response to oxygen-containing compound
0
1
2
3
-log10(p-value)

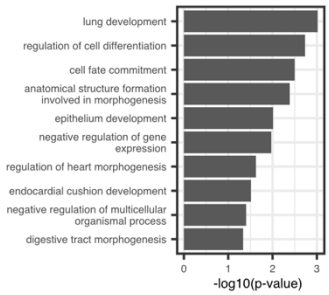

2-cell
4-cell
8-cell
ICM
ESC
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
nt
hb
mb
fb
cf
lm
ht
kd
lu
lv
st
in
Foxq1, Esrra, Elf1, Atf6
Nfe2l2, Junb, Irf2, Irf1, Hnf4a
Spib, Rorc, Pparg, Ppard, Nr1i3
Trp53, Terf2, Stat6, Stat2
1.62
0.42
-0.78
normalized PR

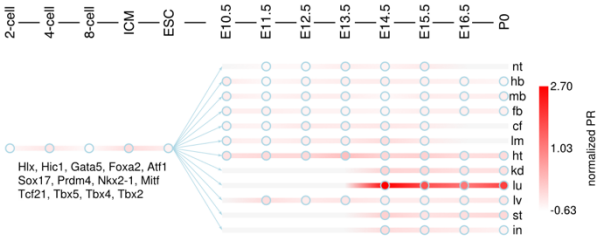
C2


positive regulation of transcription by RNA polymerase II
0
1
2
3
-log10(p-value)

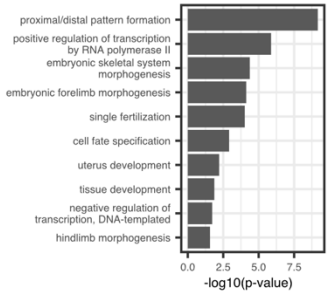

2-cell
4-cell
8-cell
ICM
ESC
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
nt
hb
mb
fb
cf
lm
ht
kd
lu
lv
st
in
Mnt, Gli3, E2f2, Cebpg, Alx3
Zfp691, Zfp282, Tcf3, Plagl1
Zscan20, Zfp740, Zfp711
2.64
0.76
-1.12
normalized PR

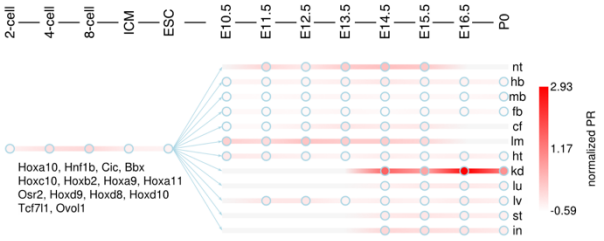
C3


lung development
regulation of cell differentiation
cell fate commitment
anatomical structure formation involved in morphogenesis
epithelium development
negative regulation of gene expression
regulation of heart morphogenesis
endocardial cushion development
negative regulation of multicellular organismal process
digestive tract morphogenesis
0
1
2
3
-log10(p-value)

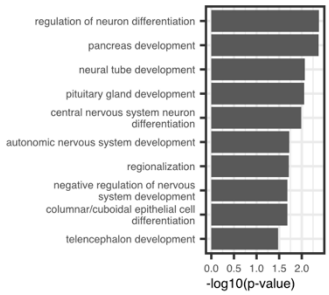

2-cell
4-cell
8-cell
ICM
ESC
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
nt
hb
mb
fb
cf
lm
ht
kd
lu
lv
st
in
Hlx, Hic1, Gata5, Foxa2, Atf1
Sox17, Prdm4, Nkx2-1, Mitf
Tcf21, Tbx5, Tbx4, Tbx2
2.70
1.03
-0.63
normalized PR

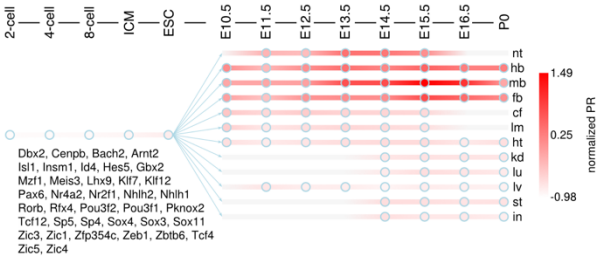
C4


proximal/distal pattern formation
positive regulation of transcription by RNA polymerase II
embryonic skeletal system morphogenesis
embryonic forelimb morphogenesis
single fertilization
cell fate specification
uterus development
tissue development
negative regulation of transcription, DNA-templated
hindlimb morphogenesis
0.0
2.5
5.0
7.5
-log10(p-value)

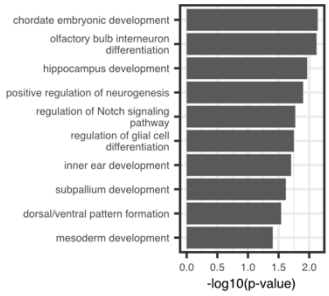

2-cell
4-cell
8-cell
ICM
ESC
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
nt
hb
mb
fb
cf
lm
ht
kd
lu
lv
st
in
Hoxa10, Hnf1b, Cic, Bbx
Hoxc10, Hoxb2, Hoxa9, Hoxa11
Osr2, Hoxd9, Hoxd8, Hoxd10
Tcf7l1, Ovol1
2.93
1.17
-0.59
normalized PR

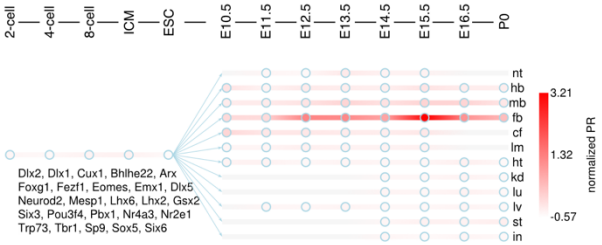
C5


regulation of neuron differentiation
pancreas development
neural tube development
pituitary gland development
central nervous system neuron differentiation
autonomic nervous system development
regionalization
negative regulation of nervous system development
columnar/cuboidal epithelial cell differentiation
telencephalon development
0.0
0.5
1.0
1.5
2.0
-log10(p-value)

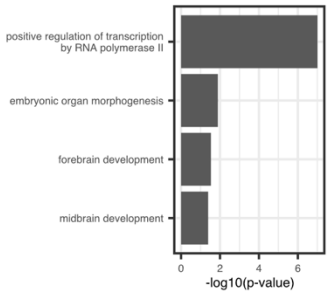

2-cell
4-cell
8-cell
ICM
ESC
E10.5
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
P0
nt
hb
mb
fb
cf
lm
ht
kd
lu
lv
st
in
Dbx2, Cenpb, Bach2, Arnt2
Isl1, Insm1, Id4, Hes5, Gbx2
Mzf1, Meis3, Lhx9, Klf7, Klf12
Pax6, Nr4a2, Nr2f1, Nhlh2, Nhlh1
Rorb, Rfx4, Pou3f2, Pou3f1, Pknox2
Tcf12, Sp5, Sp4, Sox4, Sox3, Sox11
Zic3, Zic1, Zfp354c, Zeb1, Zbtb6, Tcf4
Zic5, Zic4
1.49
0.25
-0.98
normalized PR

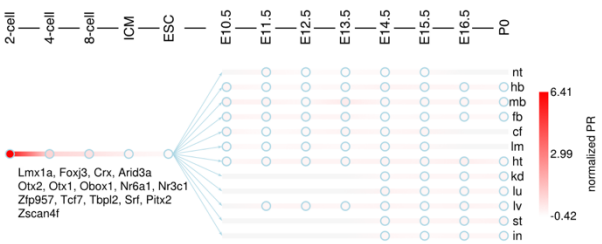
C6



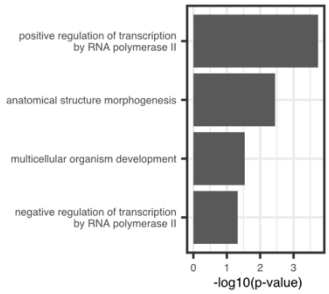


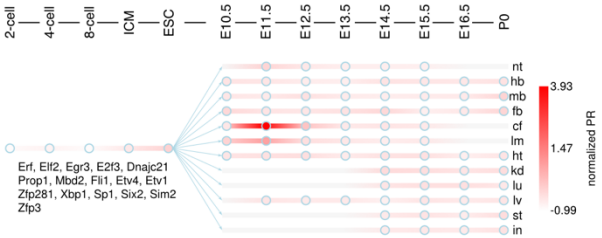
C7



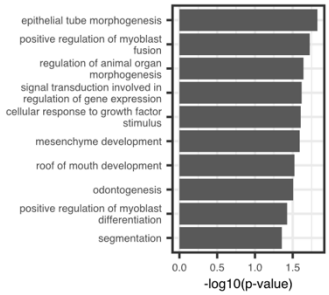


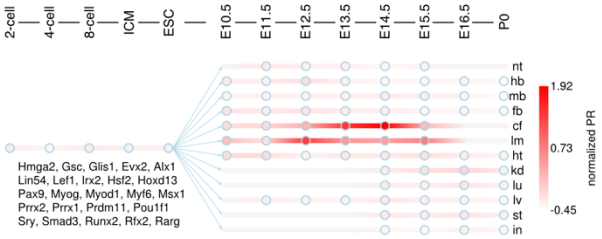
C8



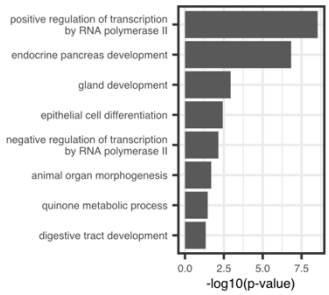


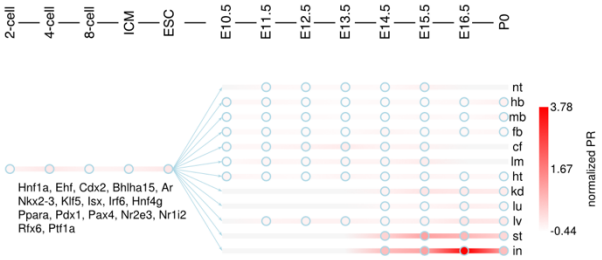
C9



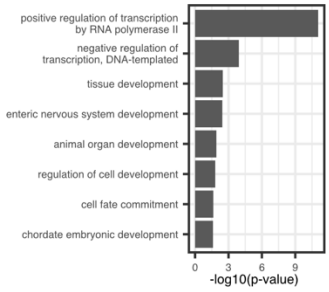


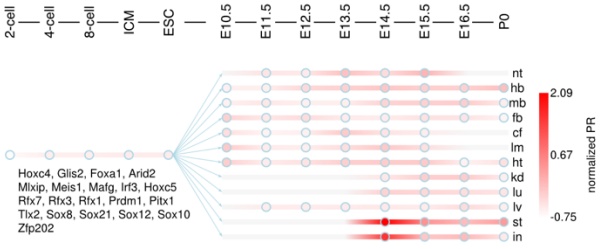
C10



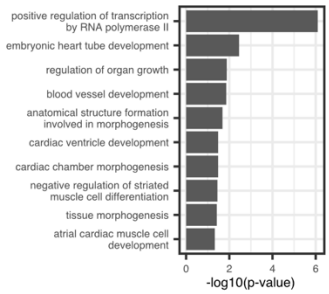


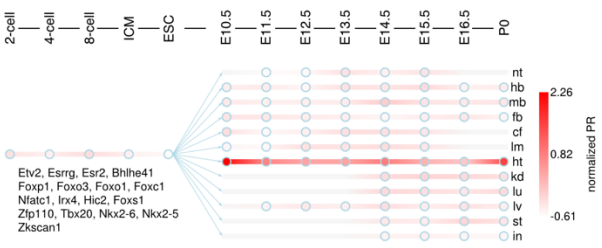
C11



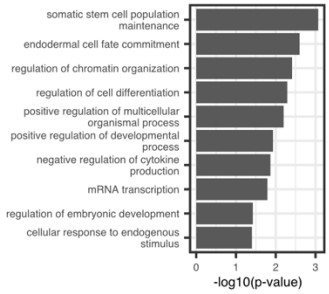


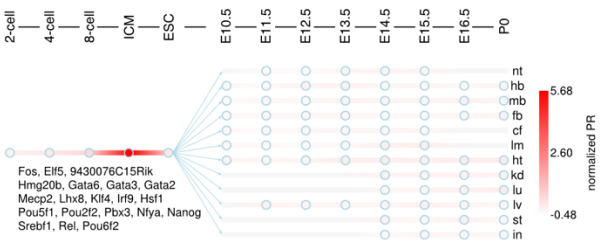
C12



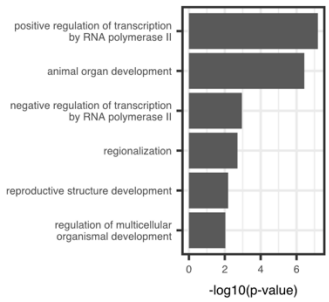


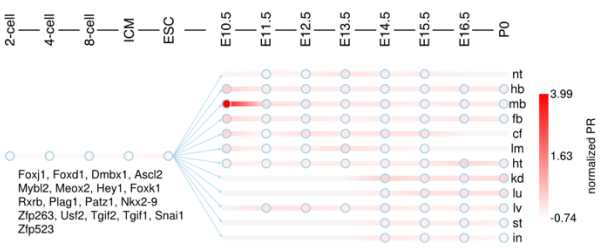
C13



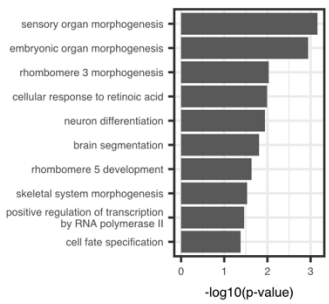


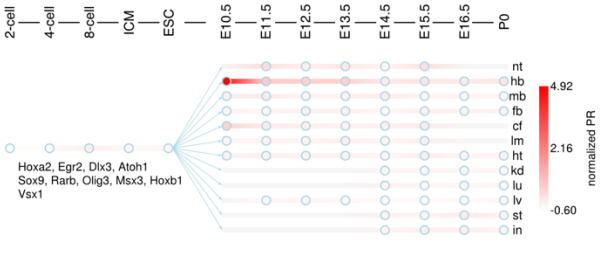
C14



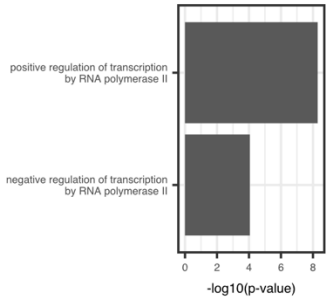


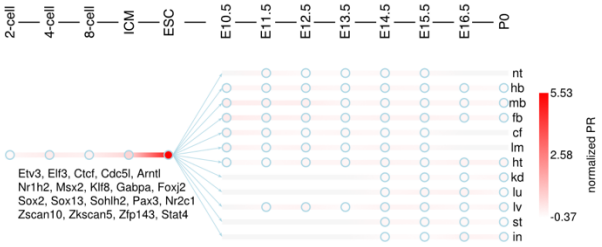
C15



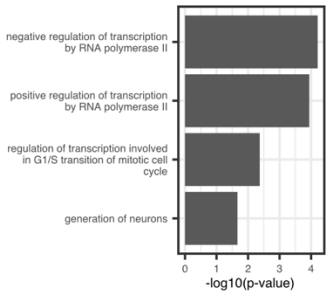


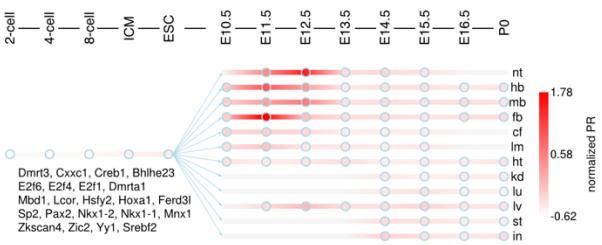
C16



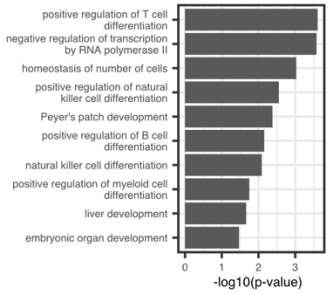


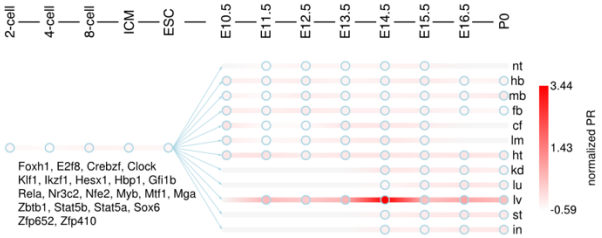
C17



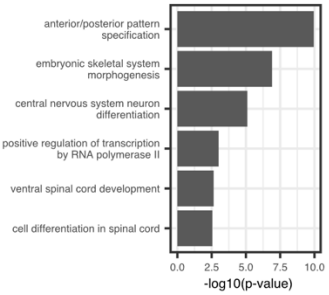


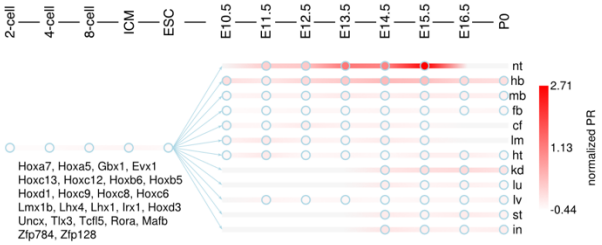
C18



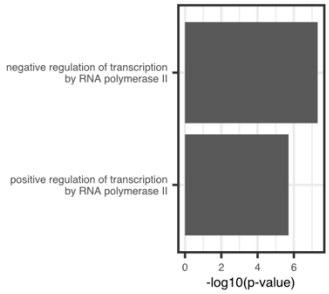


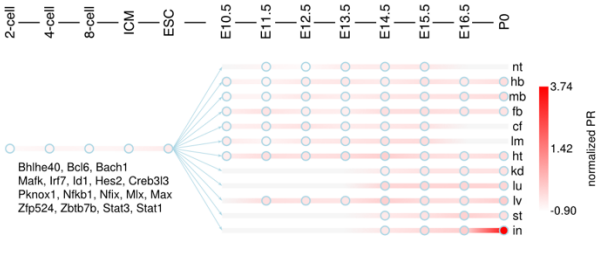
C19



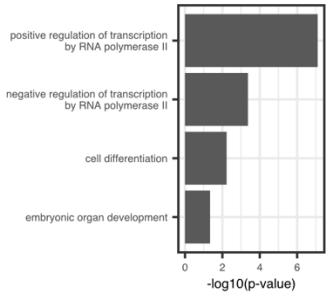


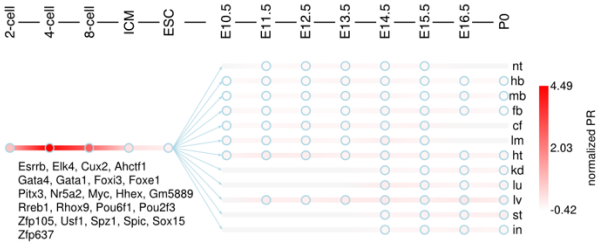
C20



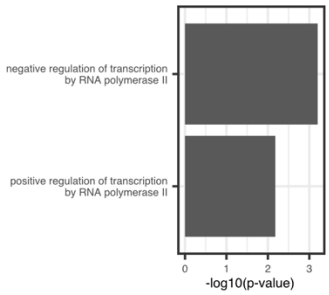


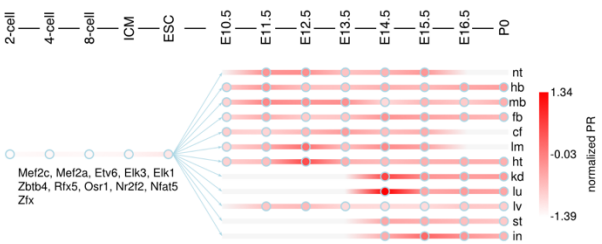
C21



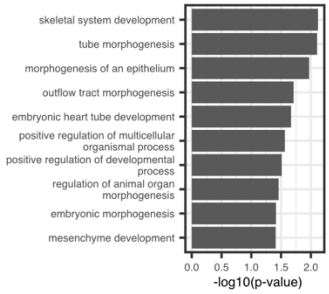


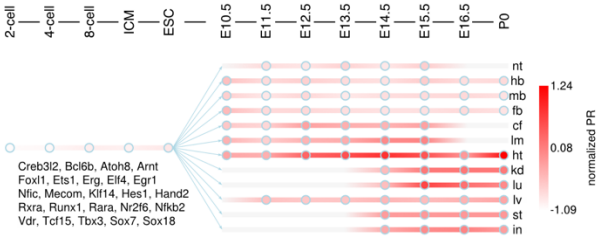
C22



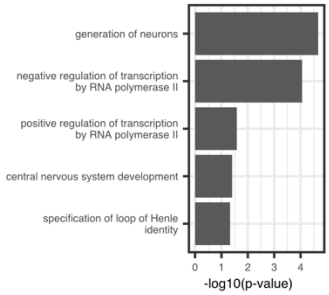


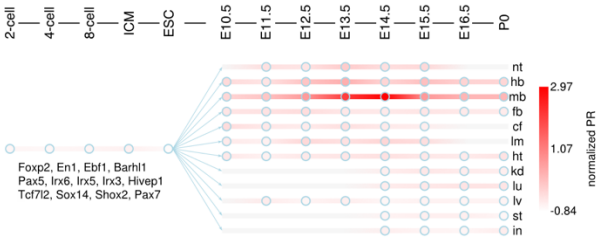
C23



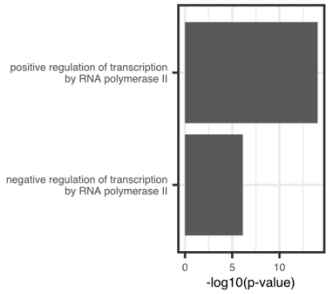


C24





C25



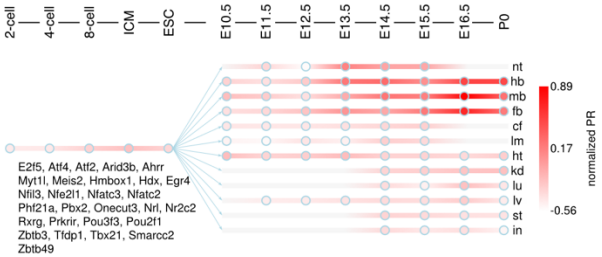


Fig. S7. Transcriptional waves direct tissue differentiation during mouse embryogenesis.

Table S1. Driver TFs in the heart.

Predicted driver TFs	Evidence
Etv2	Affect heart development (*40*)
Esx1	Unknown
Nkx2-6	Relate to congenital heart disease (*41*)
Bhlhe41	Unknown
Gata2	Relate to coronary artery disease (*42*)
Gata4	Essential for heart development (*43*)
Nkx2-5	Relate to heart development (*44*)
Tbx20	Essential for heart development (*45*)
Esrrg	Affect heart development (*46*)
Spz1	Unknown
Irx4	Affect heart development (*47*)

Table S2. Driver TFs in the limb.

Predicted driver TFs	Evidence
Hoxa11	Relate to limb development (*48*)
Hoxa13	Relate to limb development (*48*)
Evx2	Important for limb development (*49*)
Msx1	Important for limb development (*50*)
Myod1	Important for limb development (*51*)
Myog	Important for limb development (*52*)
Sim2	Affect limb development (*53*)
Fos	Relate to limb mesenchymal chondrogenesis (*54*)
Rfx2	Unknown
Cebpg	Unknown

Table S3. Driver TFs in the liver.

Predicted driver TFs	Evidence
Gfi1b	Relate to liver development (*55*)
Nr3c2	Unknown
Ikzf1	Relate to liver cancer (*56*)
Nfe2	Important for liver function (*57*)
Gata1	Important for liver development (*58*)
Klf1	Relate to liver development (*59*)
Spic	Relate to liver development (*60*)
E2f2	Important for liver function (*61*)
Hesx1	Unknown
Zfp652	Unknown
Foxh1	Unknown

Table S4. Driver TFs in the lung.

Predicted driver TFs	Evidence
Tcf21	Important for heart development (*62*)
Usf2	Relate to heart development (*63*)
Gata5	Important for heart development (*64*)
Foxa2	Important for heart development (*65*)
Nkx2-1	Essential for heart development (*66*, *67*)
Mitf	Relate to heart cancer (*20*)
Elf5	Relate to heart development (*68*)
Tbx4	Important for heart development (*69*)
Hic1	Relate to heart cancer (*19*)
Atf1	Important for heart development (*70*)
Tbx5	Important for heart development (*69*)
Noto	Unknown
Hlx	Unknown

Table S5. Driver TFs in the kidney.

Predicted driver TFs	Evidence
Hoxd8	Relate to kidney development (*71*)
Osr2	Relate to early pectoral fin specification and pronephric development (*72*)
Ovol1	Relate to kidney development (*73*)
Hnf1b	Relate to kidney disease (*74*)
Bbx	Relate to kidney development (*75*)
Hoxa10	Important for kidney development (*76*)
Hoxc10	Important for kidney development (*76*)
Cic	Unknown
Hoxa9	Unknown
Ehf	Unknown