Supplementary Material

Time course regulatory analysis based on paired expression and chromatin accessibility data

Zhana Duren[1, #], Xi Chen[1, #], Jingxue Xin[1], Yong Wang[2, 3,*], Wing Hung Wong[1, 4, *]

[1] Department of Statistics, Stanford University, Stanford, CA 94305, USA; [2] CEMS, NCMIS, MDIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China; [3] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, 650223, China; [4] Department of Biomedical Data Science, Bio-X Program, Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA; # Authors contributed equally to this work; * Corresponding authors.

Supplementary Material :

Supplementary Figure S1

TF\Method	Correlation	BO(200k)	BI	BOI	BOI+Correlation	BOI+Correlation+TFTG (PECA2)
Pou5f1	0.5866	0.6914	0.7791	0.8175	0.8203	0.8840
Sox2	0.6632	0.7022	0.8579	0.8946	0.9048	0.9162
Nanog	0.6372	0.7010	0.8364	0.8882	0.8903	0.8946
Esrrb	0.6241	0.7422	0.7718	0.7863	0.8051	0.8872
Stat3	0.6736	0.7893	0.8235	0.8708	0.8851	0.8803
Mean	0.6370	0.7252	0.8138	0.8515	0.8611	0.8924

Fig. S1. Comparison of TF-TG prediction on 5 key regulators in mESC. The values represent AUC. A gene is considered as positive sample if it is a target from both ChIP-seq experiment and knocking-down experiment. Negative samples are neither ChIP-seq target nor knocking-down target.

Supplementary Figure S2

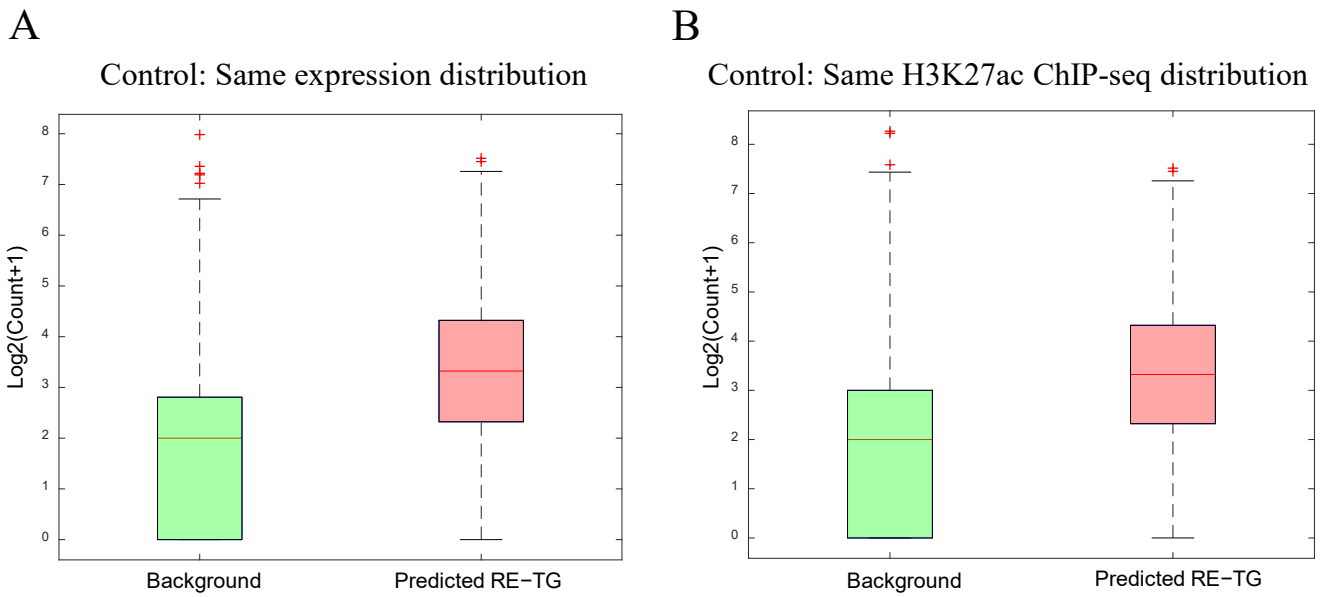

Fig. S2. Validation of PECA2 predicted RE-TG pairs by the HiChIP experiment on mESC. (A) Background RE-TG pairs are randomly selected to have the same distribution of target gene expression as the predicted RE-TG pairs. Fold represent fold change of average read count of predicted RE-TG pairs verse background RE-TG pairs. (B) Background RE-TG pairs are randomly selected to have the same distribution of number of two ends covered by H3K27ac ChIP-seq peak as the predicted RE-TG pairs.

Supplementary Figure S3

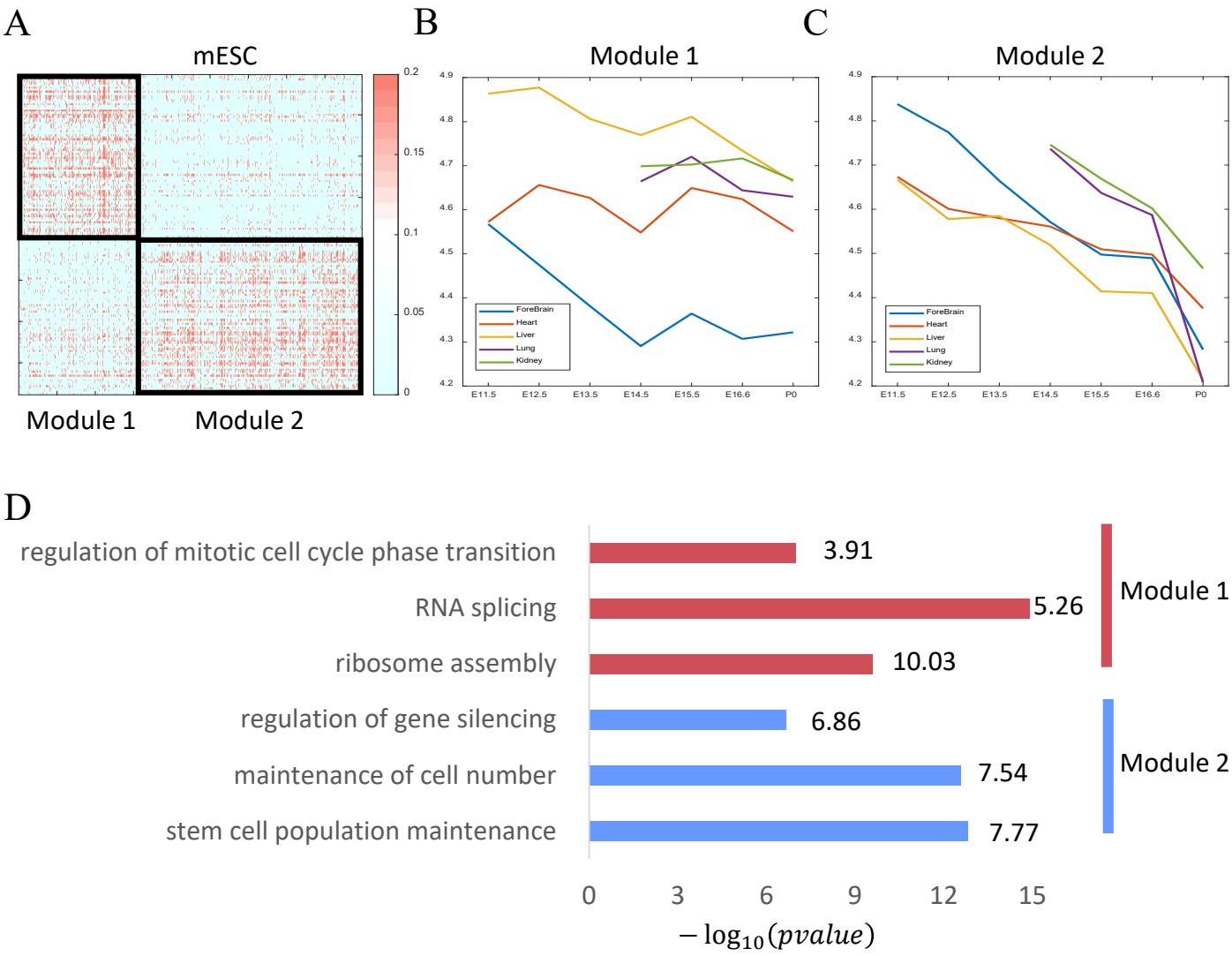

Fig. S3. Core module analysis in mESC. (A) Heatmap of reordered normalized TRS scores in mESC. The black line represents the detected modules from NMF. (B-C) Mean expression pattern of genes from two different modules of mESC on 7 tissues' developmental stage. (D) Enriched GO terms in the top 500 module specific genes of two modules. The horizontal axis is $-\log_{10}$(p-value) and the number behind the bar represents fold enrichment.

Supplementary Figure S4

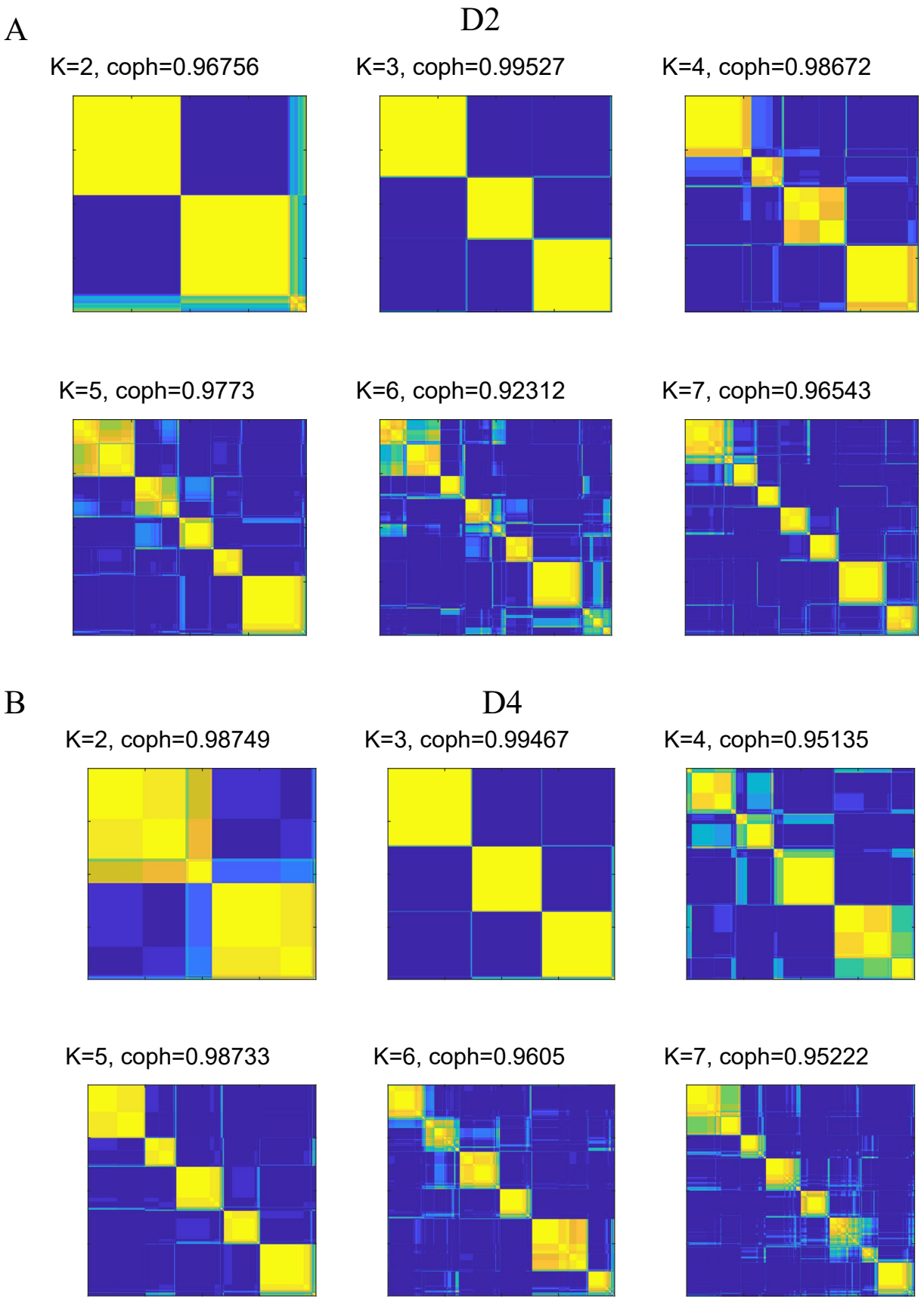

Fig. S4. Clustering stability of NMF clustering in RA day 2 and day 4 TRS matrix for k=2 to 7. Genes are hierarchically clustered by using distances derived from 50 times consensus clustering matrix entries, colored from 0 (deep blue, samples are never in the same cluster) to 1 (dark yellow, samples are always in the same cluster).

Supplementary Figure S5

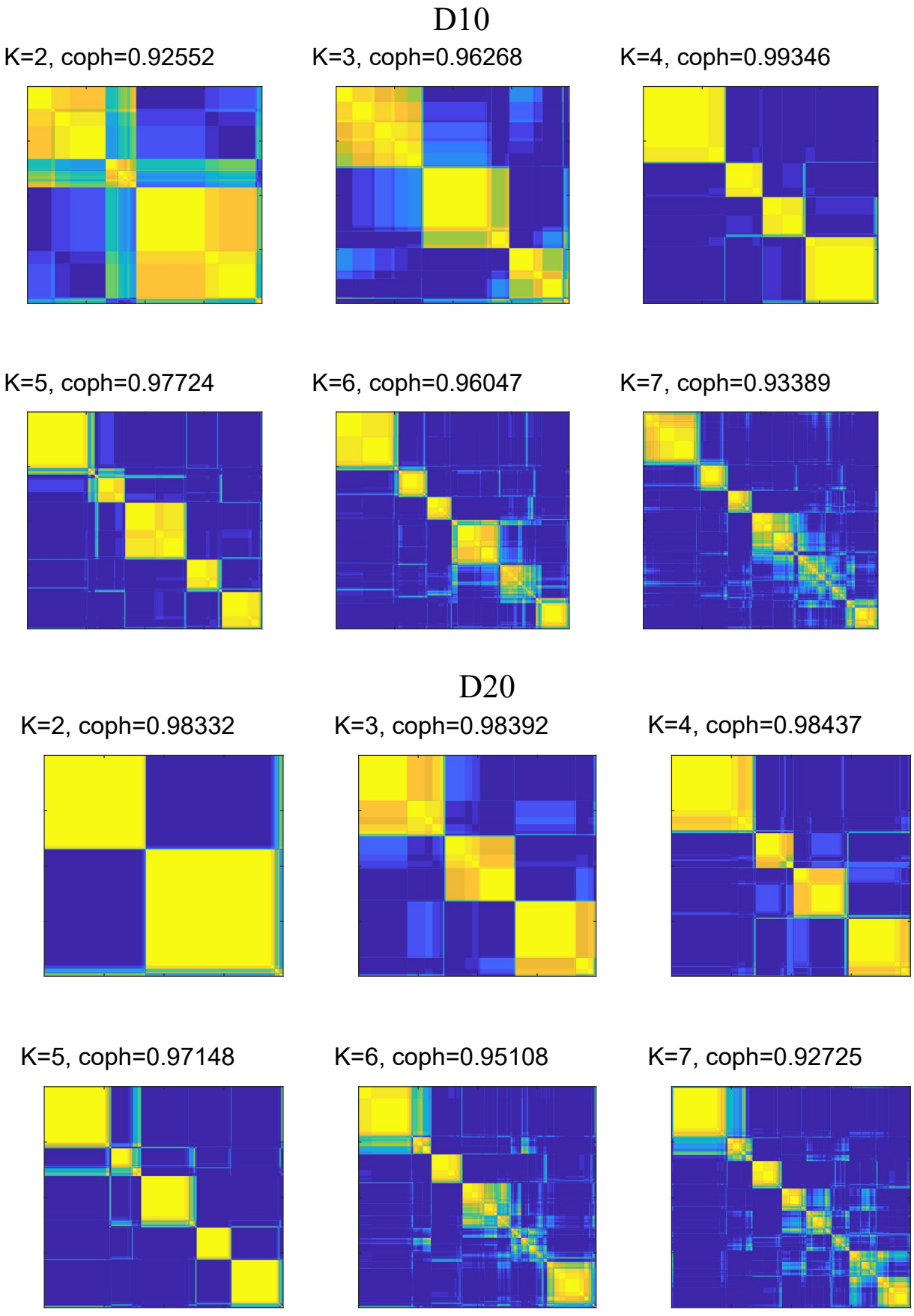

Fig. S5. Clustering stability of NMF clustering in RA day 10 and day 20 TRS matrix for k=2 to 7. Genes are hierarchically clustered by using distances derived from 50 times consensus clustering matrix entries, colored from 0 (deep blue, samples are never in the same cluster) to 1 (dark yellow, samples are always in the same cluster).

Supplementary Figure S6

Module	GO terms	D2	D4	D10	D20
Module 1	axon guidance	7.53, 5.58E-09	9.27, 368E-12	6.67, 9.80E-08	6.67, 9.80E-08
	anterior/posterior pattern specification	14.57, 1.25E-22	10.9, 1.60E-15	10.32, 2.39E-14	7.74, 1.14E-09
	neuron migration	6.14, 4.78E-06	7.69, 3.92E-08	6.03, 6.20E-06	6.9, 5.02E-07
Module 2	epithelium development	4.73, 2.18E-11	4.38,7.03E-10	2.29, 2.97E-03	2.29, 2.97E-03
	cardiovascular system development	3.67, 4.68E-05	3.7, 4.26E-05	3.23, 3.11E-04	1.16, 7.45E-01
	digestion	0.00, 1.00E+00	0.00, 1.00E+00	0.00, 1.00E+00	6.82, 3.51E-06
Module 3	stem cell population maintenance	6.36, 3.33E-06	5.41, 4.20E-05	2.73, 2.30E-02	2.73, 2.24E-02
	neural tube closure	7.14,7.82E-07	7.14, 8.37E-07	4.12, 7.39E-03	7.21, 6.80E-07
	response to retinoic acid	5.81, 3.81E-05	4.65, 5.53E-04	2.35, 5.04E-02	3.53, 5.66E-03
Module 4	glial cell differentiation			6.15, 1.89E-06	8.40, 1.05E-09
	ensheathment of neurons			2.02, 8.82E-02	6.00, 1.45E-05
	oligodendrocyte differentiation			5.13, 2.31E-04	10.13, 1.18E-09

Fig. S6. GO analysis on modules at each time points. Top 100 specifically expressed genes of each module at each time points are selected for GO enrichment analysis. The first number represents fold enrichment and the second number represents enrichment p-values.

Supplementary Figure S7

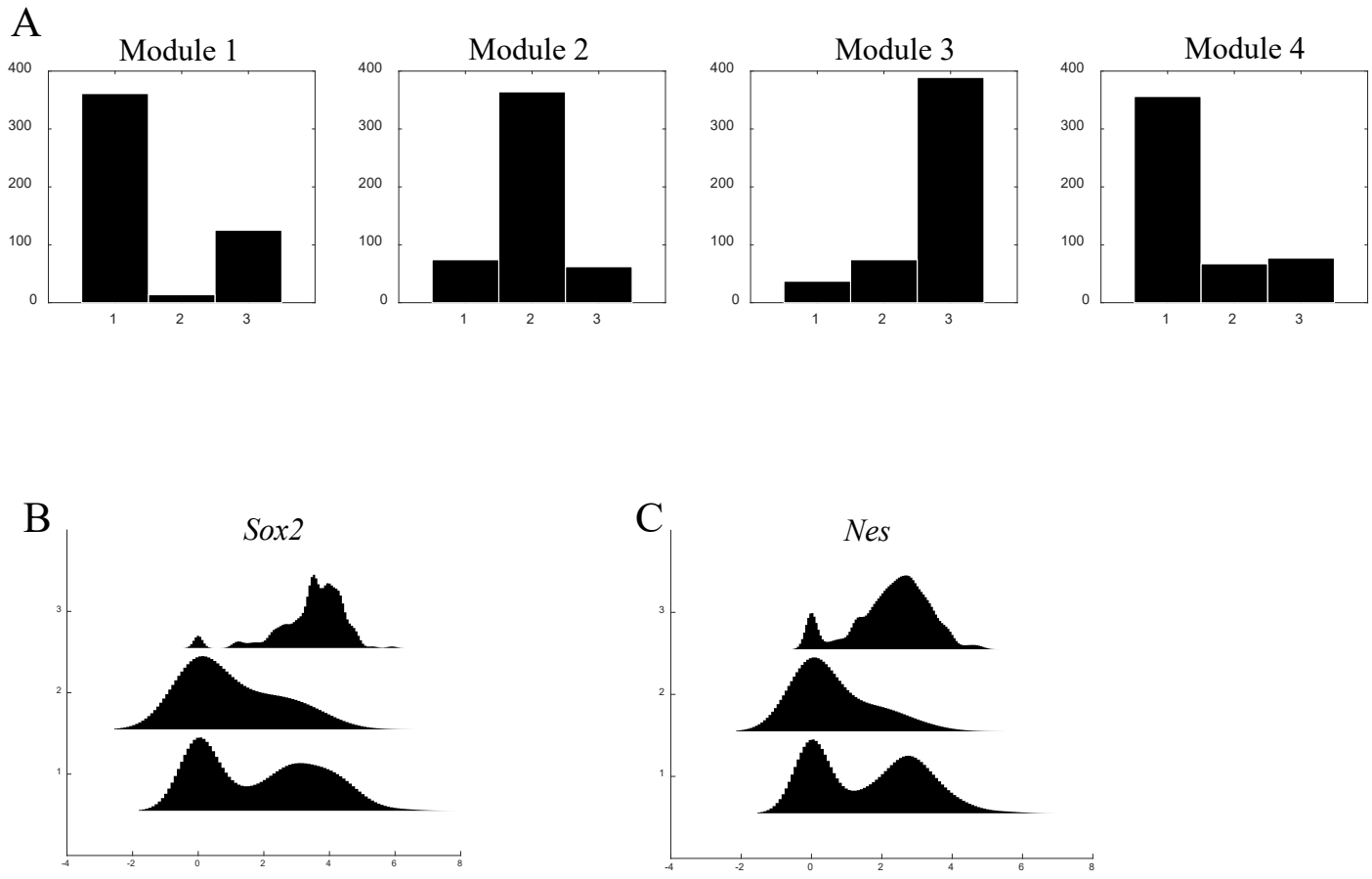

Fig. S7. (A) Histogram of RAd10 top 500 module specific genes' maximum expressed subpopulations from RAd4 scRNA-seq data. (B-C) Distribution of Neural stem cell marker genes Sox2 and Nes expression on day 4 scRNA-seq data.

Supplementary Figure S8

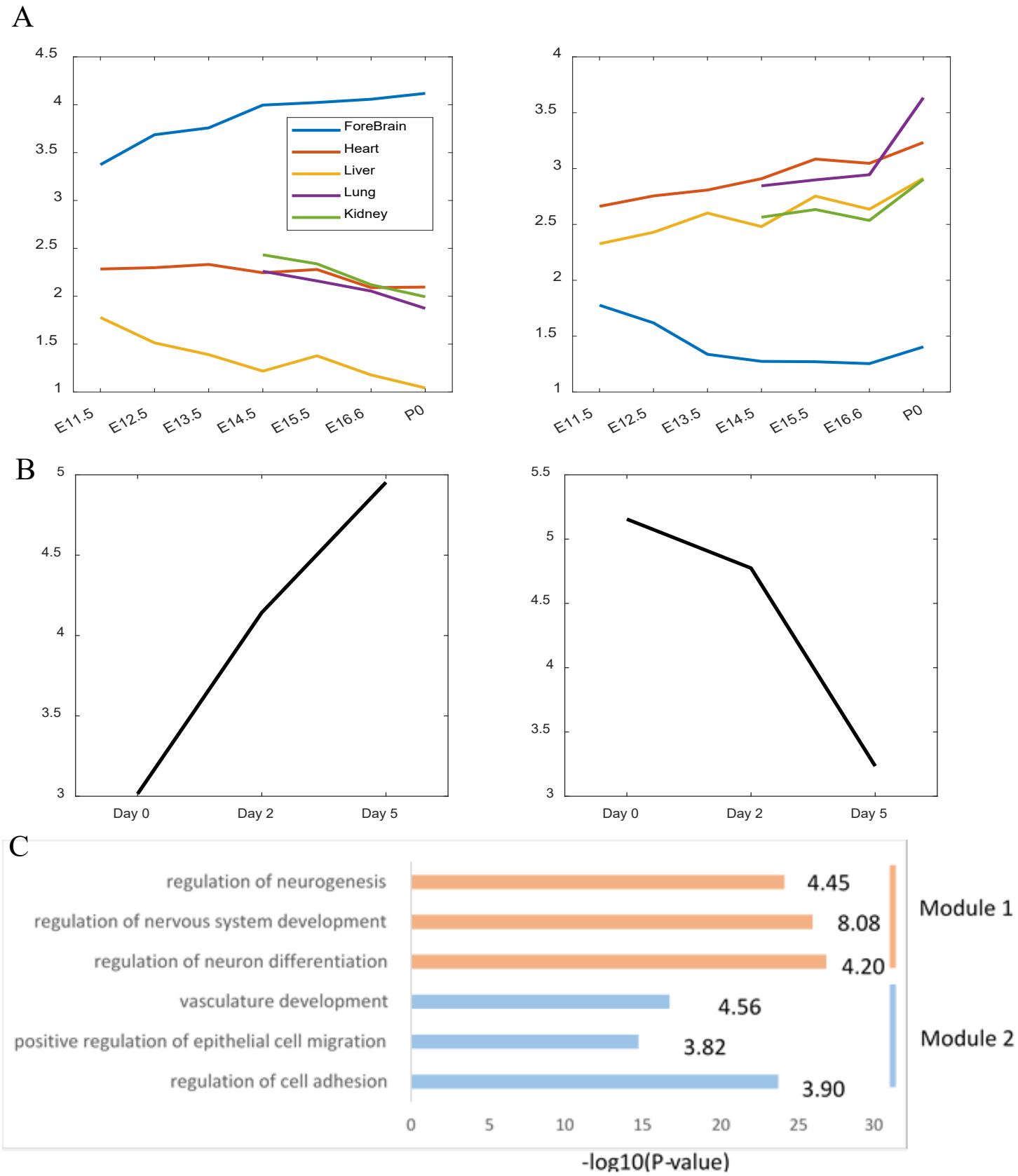

Fig. S8. TimeReg analysis on direct reprogramming from fibroblast to neuron. (A) Mean expression pattern of genes from two different modules of day 2 on 7 tissues' developmental stage. (B) Mean expression pattern of genes from two different modules of day 2 on reprogramming time course. (C) Enriched GO terms in the top 500 module specific genes of two modules. The horizontal axis is $-\log_{10}$(p-value) and the number behind the bar represents fold enrichment.

Supplementary Figure S9

A

Name	Sequence (5'-3')	Reference
OCT4	GATCCCC**GAAGGATGTGGTTCGAGTA**TTCAAGAGA**TACTCGAACCACATCCTTC**TTTTTA AGCTTAAAAAGAAGGATGTGGTTCGAGTATCTCTTGAATACTCGAACCACATCCTTCGGG	Ng
SOX2	GATCCCC**GAAGGAGCACCCGGATTAT**TTCAAGAGA**ATAATCCGGGTGCTCCTTC**TTTTTA AGCTTAAAAAGAAGGAGCACCCGGATTATTCTCTTGAAATAATCCGGGTGCTCCTTCGGG	Ng
NANOG	GATCCCC**GAACTATTCTTGCTTACAA**TTCAAGAGA**TTGTAAGCAAGAATAGTTC**TTTTTA AGCTTAAAAAGAACTATTCTTGCTTACAATCTCTTGAATTGTAAGCAAGAATAGTTCGGG	Tcf3
ESRRB	GATCCCC**GATTCGATGTACATTGAGA**TTCAAGAGA**TCTCAATGTACATCGAATC**TTTTTC TCGAGAAAAAGATTCGATGTACATTGAGATCTCTTGAATCTCAATGTACATCGAATCGGG	Ng
STAT3	GATCCCC**GAGTCACATGCCACGTTGG**TTCAAGAGA**CCAACGTGGCATGTGACTC**TTTTTA AGCTTAAAAAGAGTCACATGCCACGTTGGTCTCTTGAACCAACGTGGCATGTGACTCGGG	Moroni

B

Gene	Primer sequence (5'-3')	Source
OCT4	QT00109186	Qiagen
NANOG	QT01743679	Qiagen
STAT3	QT00148750	Qiagen
β-actin	QT01136772	Qiagen
SOX2	GCACATGAACGGCTGGAGCAACG TGCTGCGAGTAGGACATGCTGTAGG	Ng et al
ESRRB	AACCATTCAAGGCAACATCG CAGCCGTCGCTTGTACTTCT	This study

Fig. S9. List of shRNA constructs**. (A)** Letter in bold indicates the target sequences in both sense and antisense orientation. Underline letter shows hairpin sequence. The restriction enzyme sites use either BglII/HindIII or BglII/XhoI.. (B) qRT-PCR primers used in this work.