Supplemental Material

Replication Timing Networks: a novel class of gene regulatory networks

Juan Carlos Rivera-Mulia[1], Sebo Kim[2], Haitham Gabr[2], Abhijit Chakraborty[3], Ferhat Ay[3,4], Tamer Kahveci[2,*] and David M. Gilbert[5,6,*]

[1]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota Medical School, Minneapolis, MN 55455, USA.
[2]Department of Computer and Information Sciences and Engineering, University of Florida, Gainesville, Florida 32611, USA.
[3]La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, 92037, CA, USA.
[4]School of Medicine, University of California San Diego, La Jolla, CA, 92093, USA.
[5]Department of Biological Science, Florida State University, Tallahassee, FL, 32306-4295, USA.
[6]Center for Genomics and Personalized Medicine, Florida State University, Tallahassee, FL, USA.

Table of Contents

		Correlated RT networks $\binom{nodes}{edges}$					
		Correlation Threshold					
Germ layer	Degree count	0.70	0.75	0.80	0.85	0.90	0.95
All germ layers	>5	4,559 170,803	3,836 80,042	2,677 29,362	1,049 6,881	209 956	17 63
	>10	4,153 154,953	3,225 71,805	1,836 23,549	529 4,687	82 450	0 0
	>20	3,481 139,000	2,349 57,829	983 15,376	230 2,517	17 38	0 0
Ectoderm	>5	3,016 769,041	3,016 635,422	3,012 498,790	2,999 359,360	2,967 220,357	2,767 87,735
	>10	3,015 766,369	3,014 632,465	3,009 495,090	2,982 355,775	2,925 216,518	2,563 84,332
	>20	3,015 759,362	3,010 625,941	2,993 487,308	2,958 346,667	2,823 207,365	2,160 76,814
Mesoderm	>5	3,052 676,055	3,031 526,840	2,995 376,499	2,870 233,470	2,566 108,811	1,613 22,272
	>10	3,042 674,596	3,002 525,315	2,924 374,179	2,737 229,937	2,309 105,737	1,170 19,097
	>20	2,998 671,116	2,936 521,756	2,809 369,728	2,519 222,169	1,984 99,264	747 1,4169
Endoderm	>5	2,889 490,081	2,868 368,667	2,792 253,599	2,606 150,795	2,119 66,995	991 13,928
	>10	2,866 487,435	2,825 366,063	2,696 250,360	2,427 147,223	1,831 63,395	683 12,085
	>20	2,822 481,569	2,716 360,043	2,529 242,581	2,110 138,417	1,416 56,509	438 9,632

Supplemental Table S1. Size of correlated RT networks at distinct thresholds.

Correlated RT networks were constructed at distinct correlation and degree count thresholds and number of nodes and edges were quantified.

	Correlated RT networks $\binom{nodes}{edges}$
ESCs→Pancreas	$\binom{1{,}163}{283{,}668}$
ESCs→DE→Pancreas	$\binom{1{,}831}{221{,}658}$
ESCs→DE→Prim.Gut→Pancreas	$\binom{2{,}067}{157{,}580}$
ESCs→DE→Prim.Gut→Post. Foregut→Pancreas	$\binom{2{,}000}{132{,}295}$

Supplemental Table S2. Correlated RT networks using distinct differentiation time points.

Correlated RT networks were constructed for pancreas differentiation. To test the effect of the number of differentiation stages we constructed networks with all available time points (5) as well as removing intermediate stages to obtain networks wit 2, 3, 4, and 5 time points. Correlation and degree count thresholds were fixed at 0.90 and 20 respectively.

	ChIP-seq peaks at TSS		*p-value*	Enrichment
	Positive	Negative		
***FOXA2* in Liver**	89	113	0.0777	1.361287
***FOXA1* in Pancreas**	153	71	3.9136×10^{-5}	2.189831

Supplemental Table S3. Overlap significance and enrichment of transcription factor binding in directed RT-networks.

ChIP-seq data for *FOXA2* transcription factor in liver tissue were downloaded from ENCODE data portal (https://www.encodeproject.org/). ChIP-seq data for *FOXA1* transcription factor in pancreas tissue were collected from Diaferia et al. 2016 (GSE64557). Peak calling against the respective input was performed using the MACS2 program with parameters: "-g hs –q 0.05". All significant transcription factor peaks (FDR < 0.05) were retained for downstream analysis. Individual transcription factor peaks were mapped to the annotated hg19 transcription start sites (+/- 20Kb) using "bedtools map" function (Quinlan and Hall 2010) with default parameters. Overlap significance and enrichment of transcription factor binding in RT-network specific genes was measured using Fisher's exact test by comparing against a similar number of random set of non-RT-network genes.

	all differentiation pathways	**ectoderm**	**mesoderm**	**endoderm**
RT edges	96	853	623	204
TRN edges	3303	843	865	292
Common edges	15	108	71	9
RT network-specific edges	81	745	552	195
TRN-specific edges	3288	735	794	283
Hypergeometric *p-value*	0.04502	4.619e-12	2.815e-07	0.32358

Supplemental Table S4. Overlap analysis of RT and TRNs interaction edges.

RT networks were constructed for matching cell types in the TRNs (Neph, et al., 2013) and common and unique interaction edges were identified. Only genes within the TRNs were used (475 transcription factors). Hypergeometric test was performed to test the overlap significance (*p-values* are shown). Ectoderm cell types = neural crest, mesenchymal stem cells and neural precursor cells. Mesoderm cell types = lateral plate mesoderm, splanchnic mesoderm, mesothelium and smooth muscle. Endoderm = definitive endoderm, immature hepatic, hepatoblast, liver (hepatocytes), primitive gut, posterior foregut and pancreas (pancreatic endoderm).

Cell type	GO biological process	# genes	Fold Enrichment	*p-value*
ESC	somatic stem cell population maintenance	7	25.03	1.43E-04
	stem cell population maintenance	9	15.5	7.31E-05
DE	cell fate commitment of primary germ layer	5	38.27	2.22E-03
	formation of primary germ layer	7	13.16	1.07E-02
	endodermal cell differentiation	5	26.79	1.26E-02
	endoderm formation	6	25.72	1.37E-03
	endoderm development	6	17.38	1.31E-02
Liver	lipid homeostasis	7	13.74	8.11E-03
	liver development	8	13.44	1.55E-03
	hepaticobiliary system development	8	13.13	1.85E-03
Pancreas	pancreatic A cell differentiation	3	> 100	2.99E-02
	endocrine pancreas development	10	54.95	5.06E-11
	endocrine system development	11	19.32	1.54E-07
	pancreas development	14	41.67	6.64E-15
	enteroendocrine cell differentiation	4	45.12	1.94E-02
	glandular epithelial cell differentiation	5	28.2	9.84E-03
Mesothel	mesodermal cell fate specification	1	30.31	3.25E-02
	mesoderm formation	2	6.33	4.03E-02
	mesoderm morphogenesis	2	6.15	4.24E-02
	mesoderm development	3	5.17	2.08E-02
Smooth Muscle	metanephric smooth muscle tissue	1	> 100	9.20E-03
	kidney smooth muscle tissue development	1	> 100	9.20E-03
	smooth muscle tissue development	2	21.65	3.98E-03
	muscle tissue development	6	4.53	2.23E-03
	myoblast fate commitment	1	43.3	2.28E-02
	muscle structure development	7	3.34	5.22E-03
NPC	generation of neurons	22	3.3	5.06E-03
	neurogenesis	22	3.08	1.54E-02
NC	neural crest cell migration	5	27.82	1.05E-02
	neural crest cell differentiation	6	18.17	1.02E-02
MSC	positive regulation of mesenchymal cell	3	22.5	3.48E-04
	regulation of mesenchymal cell proliferation	3	17.5	7.19E-04
	mesenchymal cell development	4	13.13	2.68E-04

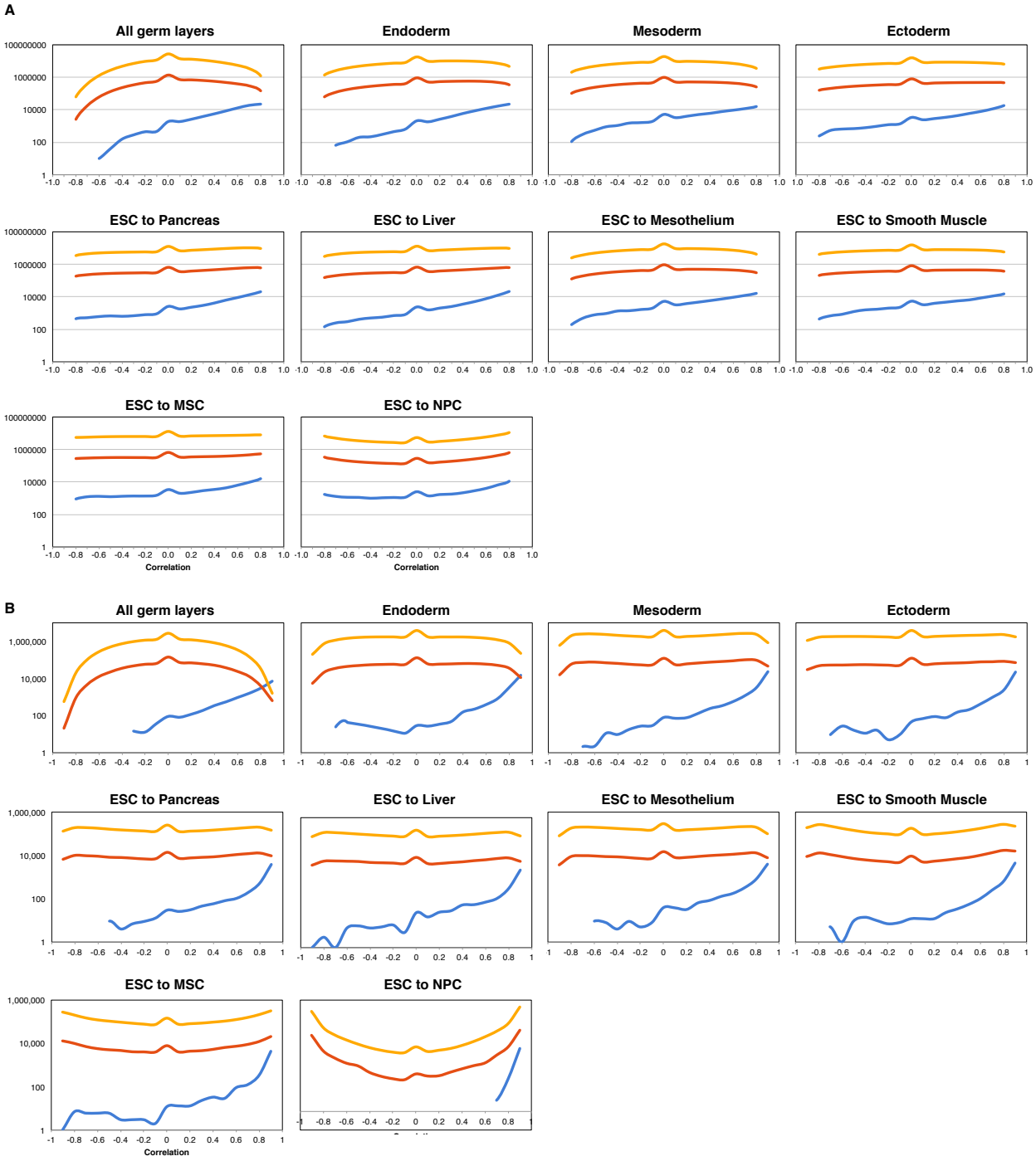
Supplemental Table S5. GO analysis of co-expressed genes in each cell type.

Co-expressed genes were identified by weighted correlation network analysis (Langfelder and Horvath, 2008) and ontology analysis (Ashburner et al., 2000; The Gene Ontology Consortium, 2015) using the top 100 genes was performed for each cell type.

RT-Network	Transcription factor (and combinations)	*p-value*	Enrichment
Pancreas	FOXA1	9.71E-04	1.637152
	PDX1	2.68E-03	1.557779
	FOXA1+PDX1	6.40E-03	1.511523
Liver	FOXA1	3.46E-05	1.617318
	FOXA1+FOXA2	9.81E-06	1.698888
	FOXA1+FOXA2+NR2F2	1.86E-03	1.466759
	FOXA1+FOXA2+NR2F2+HNF4A	2.89E-03	1.441313
	FOXA1+FOXA2+NR2F2+HNF4A+HNF4G	1.63E-02	1.370845
	FOXA2	1.53E-05	1.633169
	FOXA2+NR2F2	1.20E-03	1.456907
	FOXA2+NR2F2+HNF4A	1.75E-03	1.438051
	FOXA2+NR2F2+HNF4A+HNF4G	9.00E-03	1.390195
	NR2F2	6.26E-05	1.562558
	NR2F2+HNF4A	1.25E-04	1.531746
	NR2F2+HNF4A+HNF4G	1.97E-02	1.314695
	HNF4A	5.11E-05	1.642944
	HNF4A+HNF4G	3.78E-02	1.252503
	HNF4G	3.32E-02	1.261569

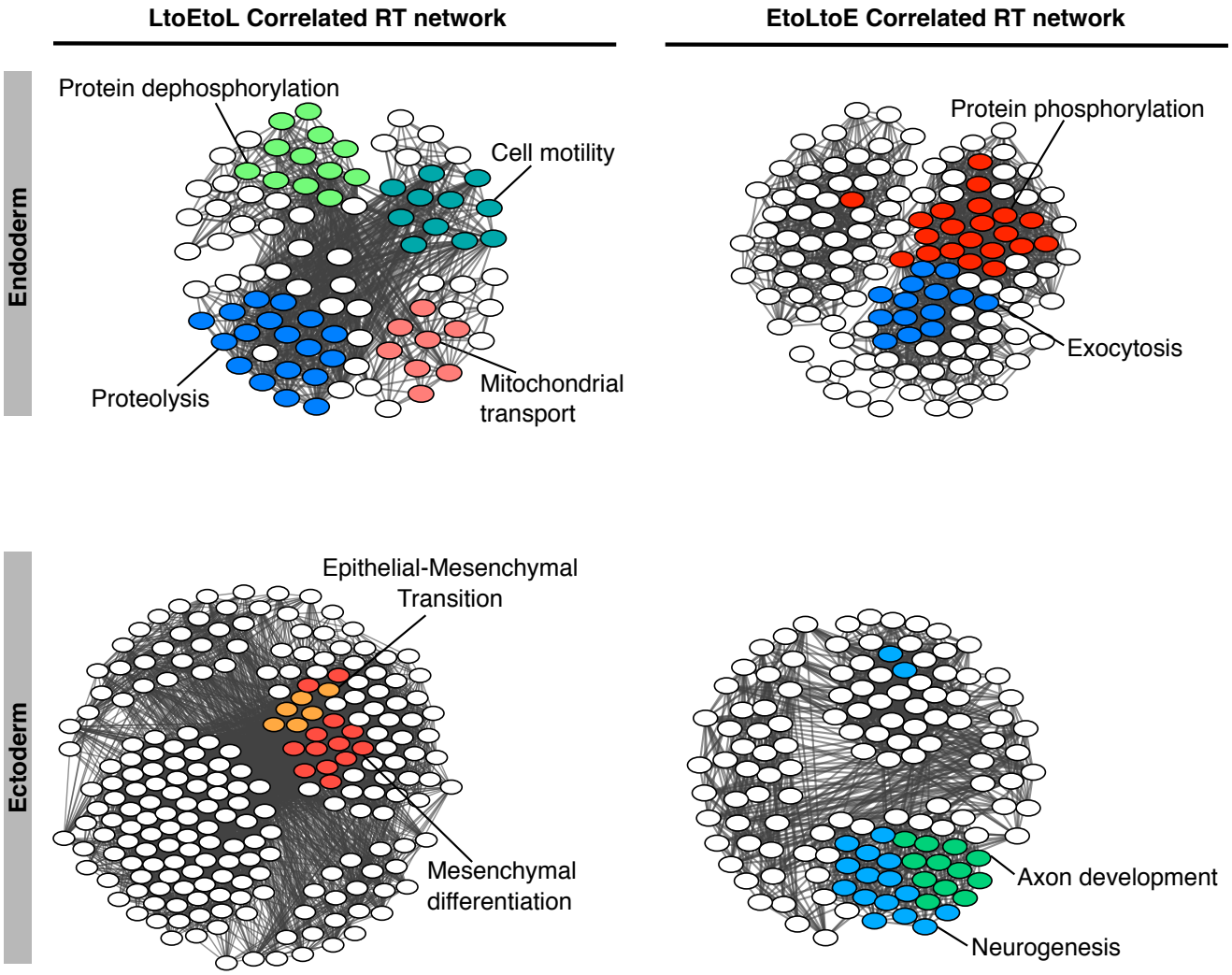
Supplemental Table S6. Overlap significance and enrichment of transcription factor binding in bipartite RT-networks.

ChIP-seq data for *FOXA1*, *FOXA2*, *NR2F2*, *HNF4A*, and *HNF4G* transcription factors in liver tissue were downloaded from ENCODE data portal (www.encodeproject.org/). ChIP-seq data for *FOXA1* transcription factor in pancreas tissue were collected from Diaferia et al. 2016 (GSE64557) and PDX1 specific ChIP-seq data in pancreas tissue was downloaded from Wang et al. 2018 (GSE106949). Aligned reads on hg19 genome assembly were generated using bowtie2 alignment program. Specific peak calling against respective input was performed using the MACS2 program. All significant transcription factor peaks (FDR < 0.05) were retained for downstream analysis. Individual transcription factor peaks were mapped to the annotated hg19 transcription start sites (+/- 20Kb) using “bedtools map” function with default parameters. Overlap significance and enrichment of transcription factor binding in RT-network specific genes was measured using Fisher’s exact test by comparing against a similar number of random set of non-RT-network genes.

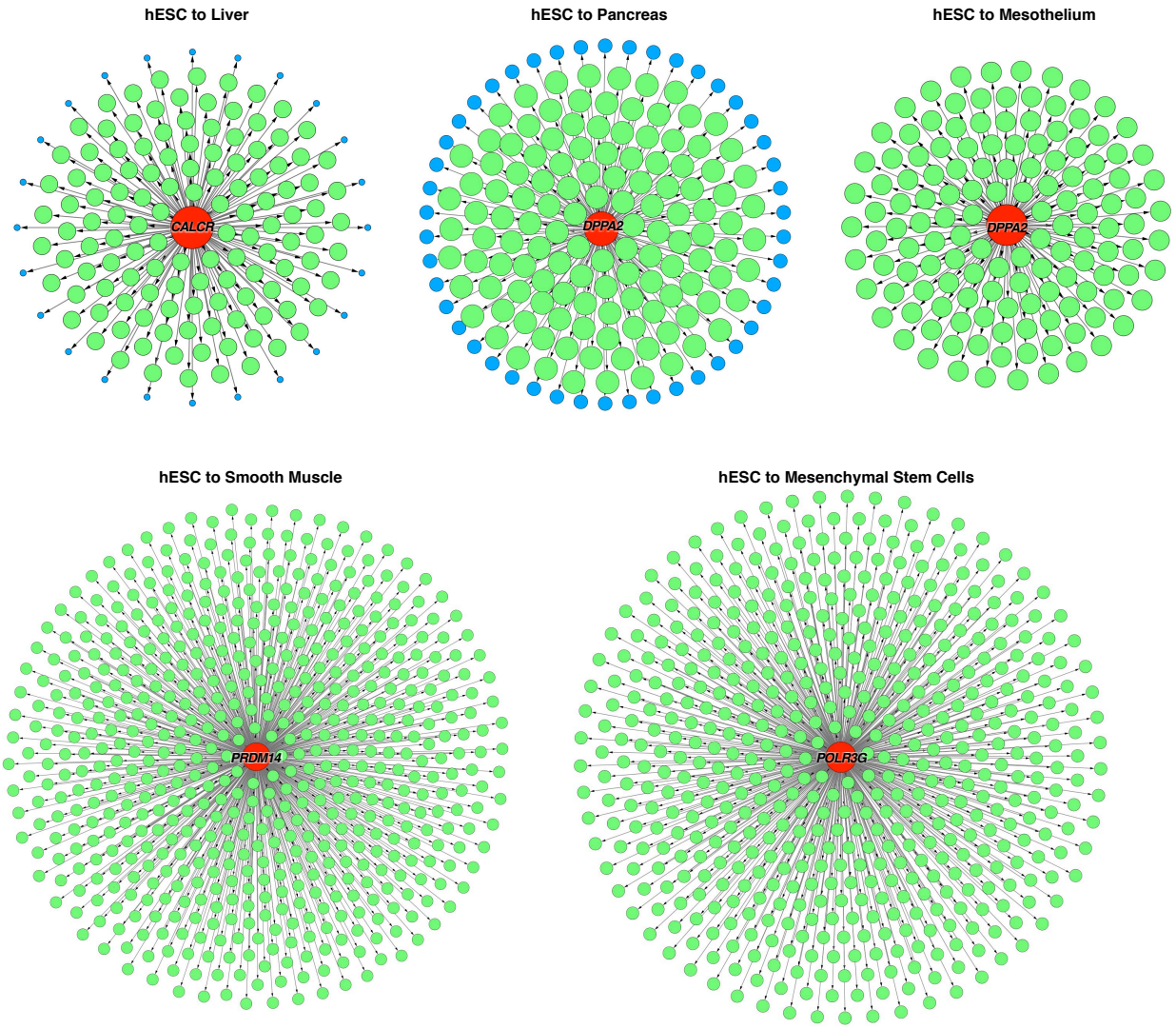


Supplemental Figure S1. RT correlation per gene pairs.

Number of gene pairs as function of RT correlation for distinct categories of gene pairs: co-located close (within 500kb), co-located distant (separated by > 500kb) and not co-located (from different chromosomes). All gene pairs were computed in (A) and gene pairs between genes that change RT significantly within each differentiation pathway (B) are shown.

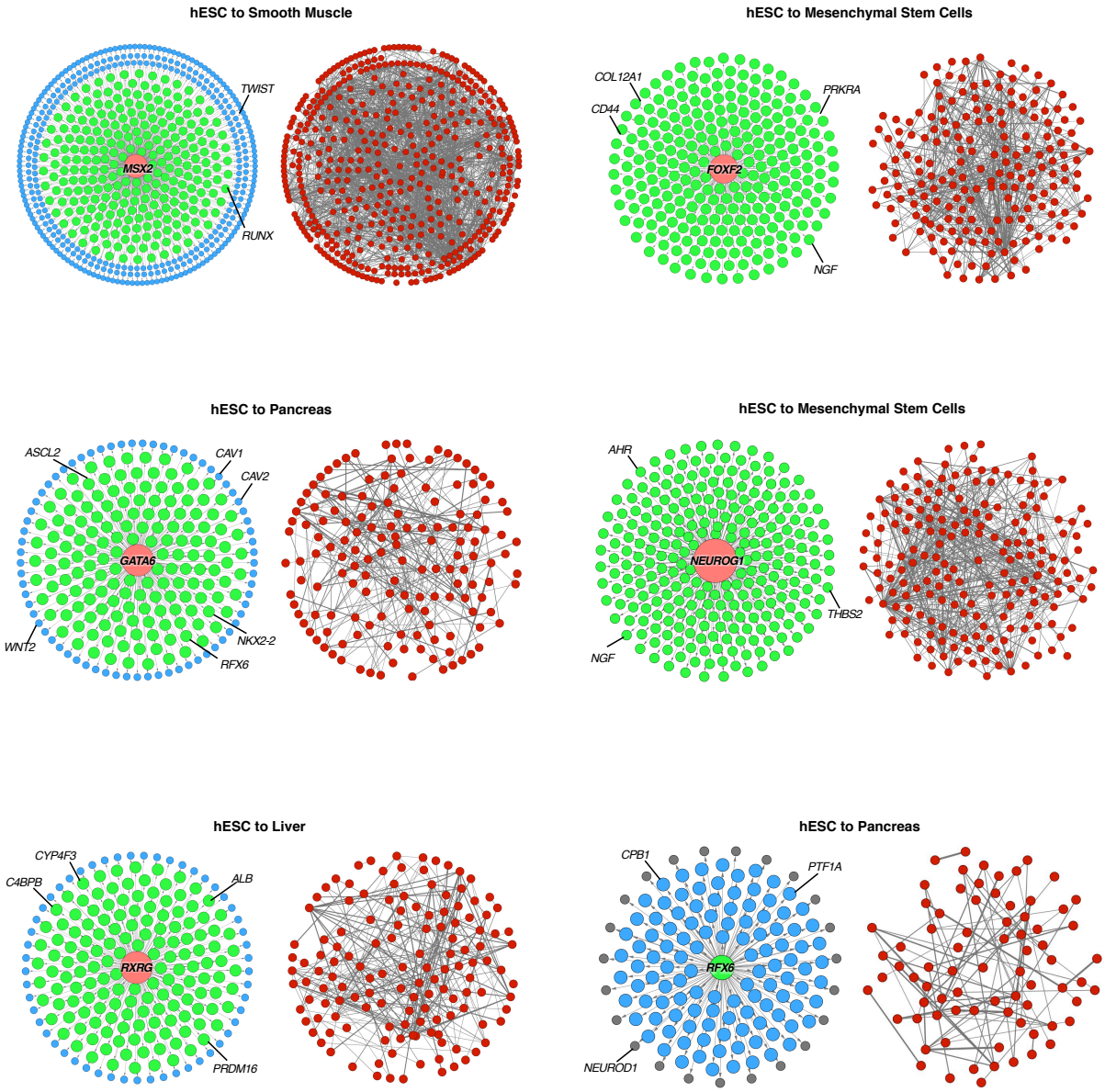


Supplemental Figure S2. Correlated RT networks for LtoEtoL and EtoLtoE changes.
Correlated RT networks for LtoEtoL and EtoLtoE changes were constructed for endodermal and ectodermal cell types. Interaction edges between gene pairs were established for correlated nodes, RT networks were displayed as 2D maps in Cytoscape (Shannon et al. 2003). Highly interconnected sub-network communities were annotated with functional ontology terms using SAFE algorithm (Baryshnikova 2016) and displayed in distinct colors.



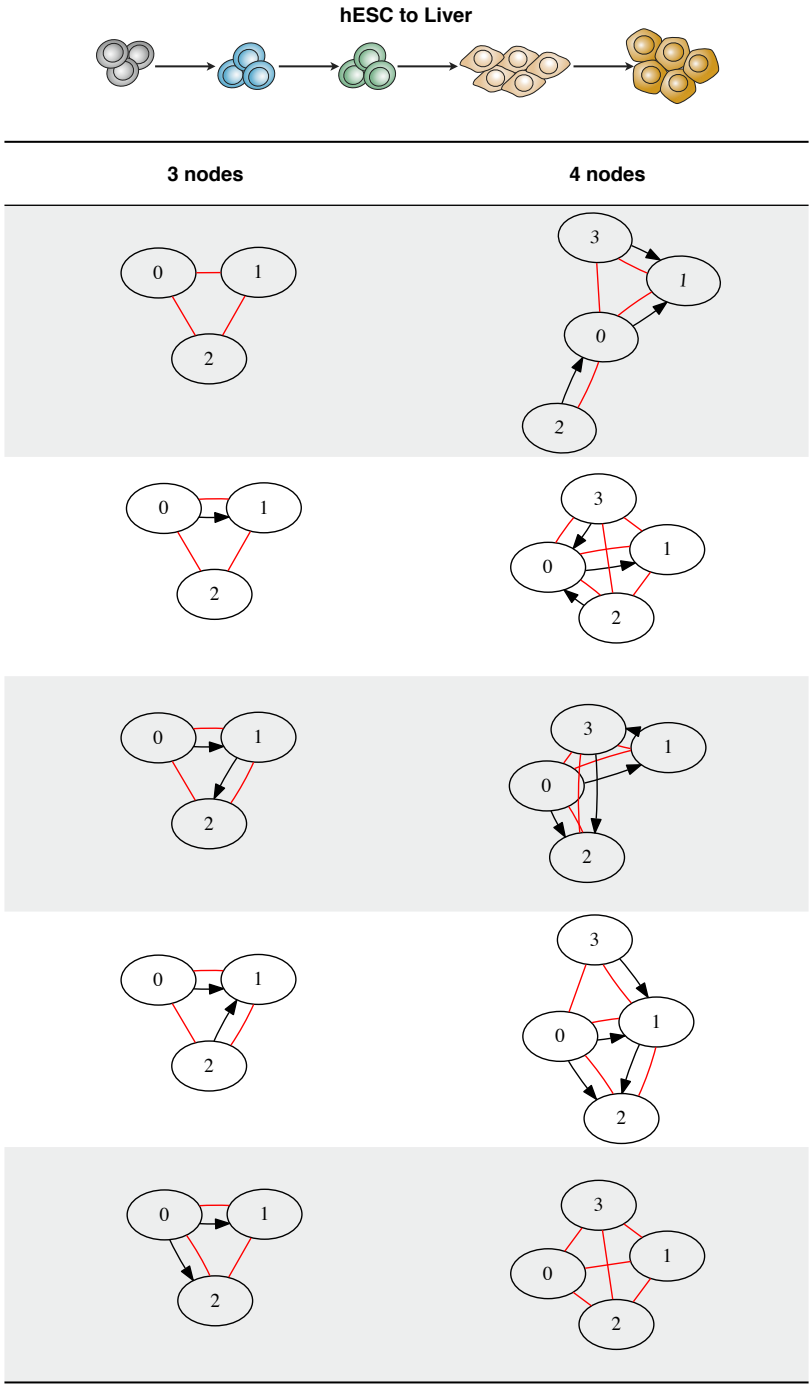
Supplemental Figure S3. Early to Late Directional RT networks.

Exemplary Early to Late directional RT networks for known pluripotency genes of each differentiation pathway are shown. Known genes involved in regulation of pluripotency were selected as source nodes and genes that change Early to Late in subsequent differentiation stages were identified. 2D maps of directional RT networks were then visualized in in Cytoscape (Shannon et al., 2003).



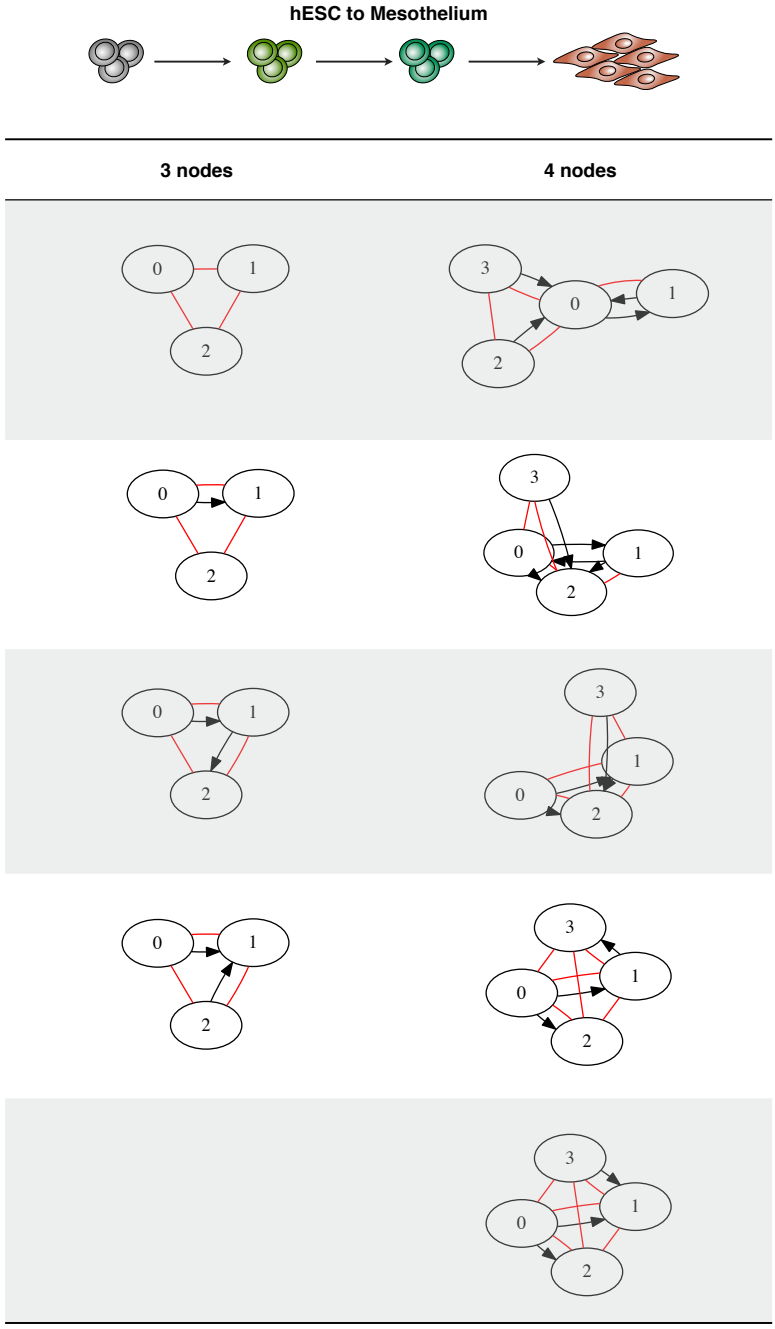
Supplemental Figure S4. Late to Early Directional RT networks.

Exemplary Late to Early directional RT networks for known key regulators of each differentiation pathway are shown. Known TFs involved in regulation cell differentiation for each pathway were selected as source nodes and potential downstream targets were identified based on the temporal RT changes. 2D maps of directional RT networks were then visualized in Cytoscape (Shannon et al., 2003). Known cell type-specific genes that distinguish each lineage were found among the downstream genes in directed RT networks (exemplary genes are highlighted in each directed RT network). Protein-protein interactions (PPINs) are depicted for each directional RT network. PPINs were obtained from the STRING database (Szklarczyk et al. 2019) and nodes were positioned mirroring the layout of the directional RT networks. Edge thickness in the PPINs indicates the strength of the data supporting each interaction.

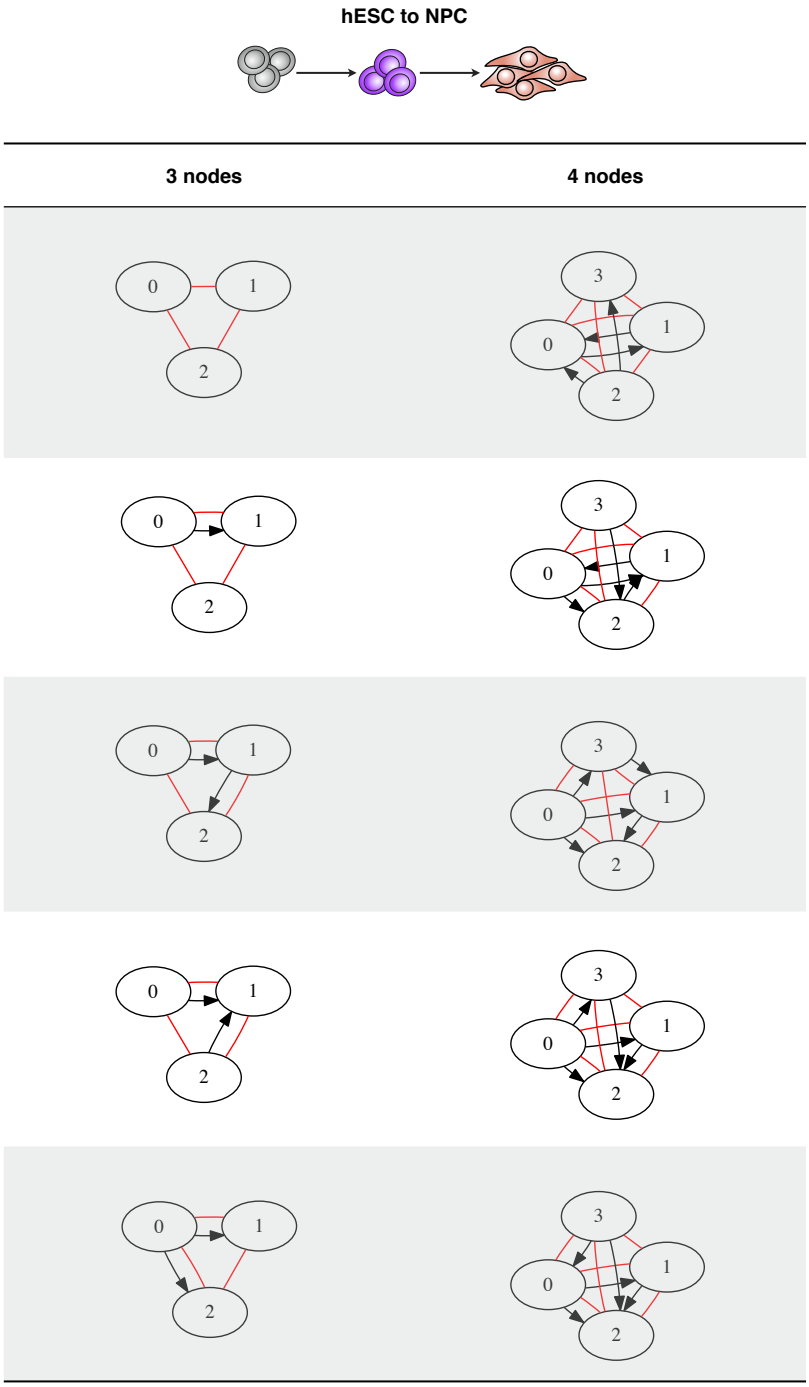


Supplemental Figure S5. Top 5 motifs identified in liver RT networks.

All possible motifs composed were computed and the most enriched motifs were identified. Statistical significance of each motif pattern was calculated by comparison to randomized networks (Baiser et al., 2015; Elhesha and Kahveci, 2016; Milo et al., 2002). Shown are the 5 most enriched motifs in liver RT networks. RT edges are shorn in red (undirected edges) and TRN edges are shown in black (directed edges).



Supplemental Figure S6. Top 5 motifs identified in mesothelium RT networks.
All possible motifs composed by 2-4 nodes were computed and the most enriched motifs were identified. Statistical significance of each motif pattern was calculated by comparison to randomized networks (Baiser et al., 2015; Elhesha and Kahveci, 2016; Milo et al., 2002). Shown are the 5 most enriched motifs in mesothelium RT networks. RT edges are shorn in red (undirected edges) and TRN edges are shown in black (directed edges).



Supplemental Figure S7. Top 5 motifs identified in NPC RT networks.

All possible motifs composed by 2-4 nodes were computed and the most enriched motifs were identified. Statistical significance of each motif pattern was calculated by comparison to randomized networks (Baiser et al., 2015; Elhesha and Kahveci, 2016; Milo et al., 2002). Shown are the 5 most enriched motifs in NPC RT networks. RT edges are shorn in red (undirected edges) and TRN edges are shown in black (directed edges).