Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supporting Information for

Revealing Complex Function, Process and Pathway Interactions with High-throughput Expression and Biological Annotation Data

Nitesh Kumar Singh[1], Mathias Ernst[1], Volkmar Liebscher[2], Georg Fuellen[1,*] and Leila Taher[1,Ŧ,*]

[1]Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock University Medical Center, Rostock, Germany.

[2]Institute for Mathematics and Informatics, Ernst Moritz Arndt University of Greifswald, Greifswald, Germany.

[Ŧ] Current address: Department of Biology, University of Erlangen-Nuremberg, Germany.

* Corresponding authors:

Prof. Dr. Leila Taher, Department of Biology, University of Erlangen-Nuremberg, Staudtstr. 5, 91058 Erlangen, Germany. E-mail: leila.taher@fau.de

Prof. Dr. Georg Fuellen, Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock University Medical Center, Ernst-Heydemann-Str. 8, 18057 Rostock, Germany. E-mail: fuellen@uni-rostock.de.

This file contains:

- Supplementary Text
- Supplementary Figures S1-S9
- Supplementary Tables S1-S7

Supplementary Text

Differentiation of mESCs into embryoid bodies

By construction, the expression profiles of genes within the same sub-clusters were not only more highly correlated than expected by chance (P-value < 0.01, Figs. S1C and S2), but also more than those of genes within the same clusters (medians 0.8 and 0.1 respectively, P-value = 2.0×10^{-9}, Wilcoxon Rank-Sum test). Additionally, the sub-clusters preserve the biological relationships between the genes in the clusters. Thus, we found that the average intra-cluster and intra-sub-cluster STRING connectivity scores (see Methods) were significantly greater than expected by chance (P-values < 0.01, Figs. S1B and D), but not significantly different from one another.

The number of genes shared by sub-clusters connected by edges in the top network was significantly greater than that observed for the bottom network (P-value = 2.1×10^{-23}, Wilcoxon Rank-Sum test, see Fig. S6). This hints at the mutual interaction between the biological functions, processes and/or pathways represented by these sub-clusters.

The number of sub-clusters within each cluster is correlated with the number of genes in the cluster (Pearson correlation coefficient = 0.97). The largest sub-cluster (SC 7.3, see Table S2) consists of 57 genes, while the smallest sub-clusters included only 12 genes (SC 9.1, SC 13.3 and SC 14.1, see Table S2). The maximum overlap between any two sub-clusters contained 32 genes, and was observed between a sub-cluster associated with "embryonic organ development" (SC 4.1) and a sub-cluster associated with "embryonic development" (SC 7.3).

Embryonic liver development vs adult liver regeneration

We applied our approach to a second microarray dataset, GSE6998 which corresponds to an mRNA expression profiling experiment comparing the developing embryonic liver and the regenerating adult liver in mice [1]. We considered liver gene expression at 10.5, 11.5, 12.5, 13.5, 14.5, and 16.5 days post-conception, as well at 1, 2, 6, 12, 18, 24, 30, 48, and 72 hours after partial hepatectomy in adult mice. Affymetrix probes were mapped to 19,808 transcripts in the Mouse Genome Informatics database (http://www.informatics.jax.org/, [2]). As per experimental design, the dataset was divided into 2 datasets, one for embryonic liver development (*development* dataset), and one for adult liver regeneration

(*regeneration* dataset). Differentially expressed genes were identified between pairs of time points by applying the Empirical Bayes t-statistics of the R package "Limma" [3]. P-values were corrected for multiple testing using the false discovery rate (FDR, [4]). We selected genes with FDR≤0.01, restricting the final datasets to 3,000 genes using the aforementioned ranking aggregation method. The final development and regeneration datasets contained 3,000 and 2,401 genes respectively.

Adult liver has a unique ability to renew and repair after injury. To compare the transcriptional program of the regenerating liver with that of the developing liver, Otu et al. generated and analyzed microarray data of adult liver after partial hepatectomy as well as of the developing embryonic liver [1]. We applied a procedure similar to that described above to both datasets (see Materials and Methods for more details). For the adult liver regeneration dataset we obtained 12 clusters that were further divided into 28 sub-clusters (Table S6). The largest cluster (C6, 387 genes) was associated with "nucleotide binding", suggesting a general importance for transcriptional activity in liver regeneration. The second largest cluster (C1) and the largest sub-cluster (SC 1.5, 67 genes) were associated with "cell cycle", whose regulation is necessarily relevant to tissue repair. In turn, the embryonic liver development dataset generated 24 clusters, which were further divided into 53 sub-clusters (Table S7). The largest cluster (C4, 696 genes) was associated with "metal ion binding", and the largest sub-cluster (SC 3.1, 84 genes) with "embryonic development". Control of metal homeostasis has been shown to be essential throughout development as early as at the blastocyst stage [5]. Furthermore, in agreement with the developmental nature of the data, we found 7 clusters associated with "developmental process".

As described for previous datasets, we then constructed two co-expression networks of sub-clusters, based on the absolute Pearson correlation coefficient between any two pairs of sub-clusters in the liver regeneration and in the liver development datasets, respectively. As before, we further defined two *top* and two *bottom* networks (comprising the 10% of the edges in the co-expression networks with the highest and the lowest absolute Pearson correlation coefficients respectively, see Materials and Methods). We showed that edges with high absolute Pearson correlation coefficients are more likely represent biologically meaningful interactions than those with low absolute Pearson correlation coefficients (see Figs. S7 and S8). The corresponding top networks were used for further analysis. The top network for the adult liver regeneration dataset (*liver regeneration* network) comprised 38 edges and

25 out of 28 total sub-clusters (Fig. S9A). The network consisted of many disconnected sub-networks, with the largest sub-network involving 11 sub-clusters. These 11 sub-clusters were associated with normal metabolism, biosynthesis of small molecules and amino acids, and drug metabolism. Drug metabolism is presumably linked to the fact that the mice in this study had been given an analgesic drug following hepatectomy, but could also be attributed to the re-establishment of adequate liver function over time. Additionally, as expected considering the annotation of the largest clusters, many sub-clusters were associated with “nucleotide binding” and “cell cycle”. However, none of the 7 sub-clusters associated with ’cell cycle’ were part of the largest sub-network. Instead, “cell cycle” sub-clusters formed separate sub-networks, and were mainly connected with sub-clusters associated with “nucleotide binding”. This suggests a strong interaction between “cell cycle” and “nucleotide binding”, and a specific transcriptional control of the cell cycle in the regenerating liver. Re-entry of hepatocytes into cell cycle and hepatocyte proliferation to reconstitute lost liver mass is well documented [6, 7]. Furthermore, our results underscore a strong role for transcriptional regulation of biosynthetic processes (represented by SC 6.2 and SCs 3.2 and 8.2, respectively). Unlike the sub-clusters describing the differentiation of mESCs into embryoid bodies, the sub-clusters of the liver regeneration dataset did not exhibit monotonically increasing or decreasing expression profiles (data not shown). Instead, for this dataset, the highest expression levels were observed between 12 hr and 18 hr. Similar findings were reported by Hsieh et al. [8]. The top network for the developing embryonic liver (*liver development* network) comprised 138 edges and 51 sub-clusters out of 53 sub-clusters (Fig. S9B). The two sub-clusters excluded from the network were related to “metal ion binding”; the fact that the network comprised 5 other sub-clusters with similar annotation indicates that such molecular function is indeed relevant, but regulated at multiple levels and, possibly, organized in a redundant configuration. Furthermore, the liver development network presented a relatively high median degree of 6 (P-value < 0.0001, computed by observing how many out of 10,000 random networks have median degree 6 or higher), suggesting mutual interactions between all biological functions, processes, and/or pathways involved. Similar to the network describing early differentiation of mESCs into embryoid bodies (Fig. 3), sub-clusters in the liver development network exhibiting increasing and decreasing expression profiles were mostly separated into disconnected sub-networks. Twenty-six sub-clusters exhibited expression profiles decreasing over time, while 27 sub-clusters showed opposite trends. The

sub-cluster with the highest degree (10, SC 10.1, Fig. S9B, and Table S7) in the liver development network was related to “blood vessel development”; the development of hepatic vasculature has been shown to play an integral part in the generation of the liver bud [9]. Along with this sub-cluster, the 3 sub-clusters with the next highest degree (9), all putative hubs, exhibited expression patterns increasing over time. These observations indicate that gene activation has a major role in liver development. Moreover, while the highest-degree sub-clusters with decreasing expression profiles were found to be related to development and cell motion (Fig. S9B, top network), those with increasing expression profiles were associated with development and metabolism (Fig. S9B, bottom network). We hypothesize that this reflects the fact that the liver becomes functional only after its cells have attained the proper spatial organization.

We did not find any one-to-one mapping between the liver regeneration dataset and the liver development dataset, indicating that the biological functions, processes, and/or pathways involved in liver regeneration and liver development are fundamentally different. Otu et al. [1] reached similar conclusions.

Sub-cluster overlap

Let $C_1 \in \{G_1^1, G_1^2, G_1^3, \ldots, G_1^{N_1}\}$ and $C_2 \in \{G_2^1, G_2^2, G_2^3, \ldots, G_2^{N_2}\}$ be two sub-clusters containing N_1 and N_2 genes, respectively, connected by an edge in a network. We defined the overlap between C_1 and C_2 as:

$$Overlap(C_1, C_2) = \frac{2 \cdot |C_1 \cap C_2|}{N_1 + N_2}$$

where $C_1 \cap C_2$ is the number of genes in C_1 that are also members of C_2. We hypothesize that interactions between biological functions, processes, and/or pathways involve a relatively large number of common genes. Therefore, we expect the sub-clusters associated with edges in the top network to exhibit a significantly higher overlap than the sub-clusters associated with edges in the bottom network. The overlap between two sub-clusters, however, is not independent from the absolute Pearson correlation coefficient between the corresponding expression profiles. Hence, we do not strictly consider the sub-cluster overlap as validation.

Supplementary Figures

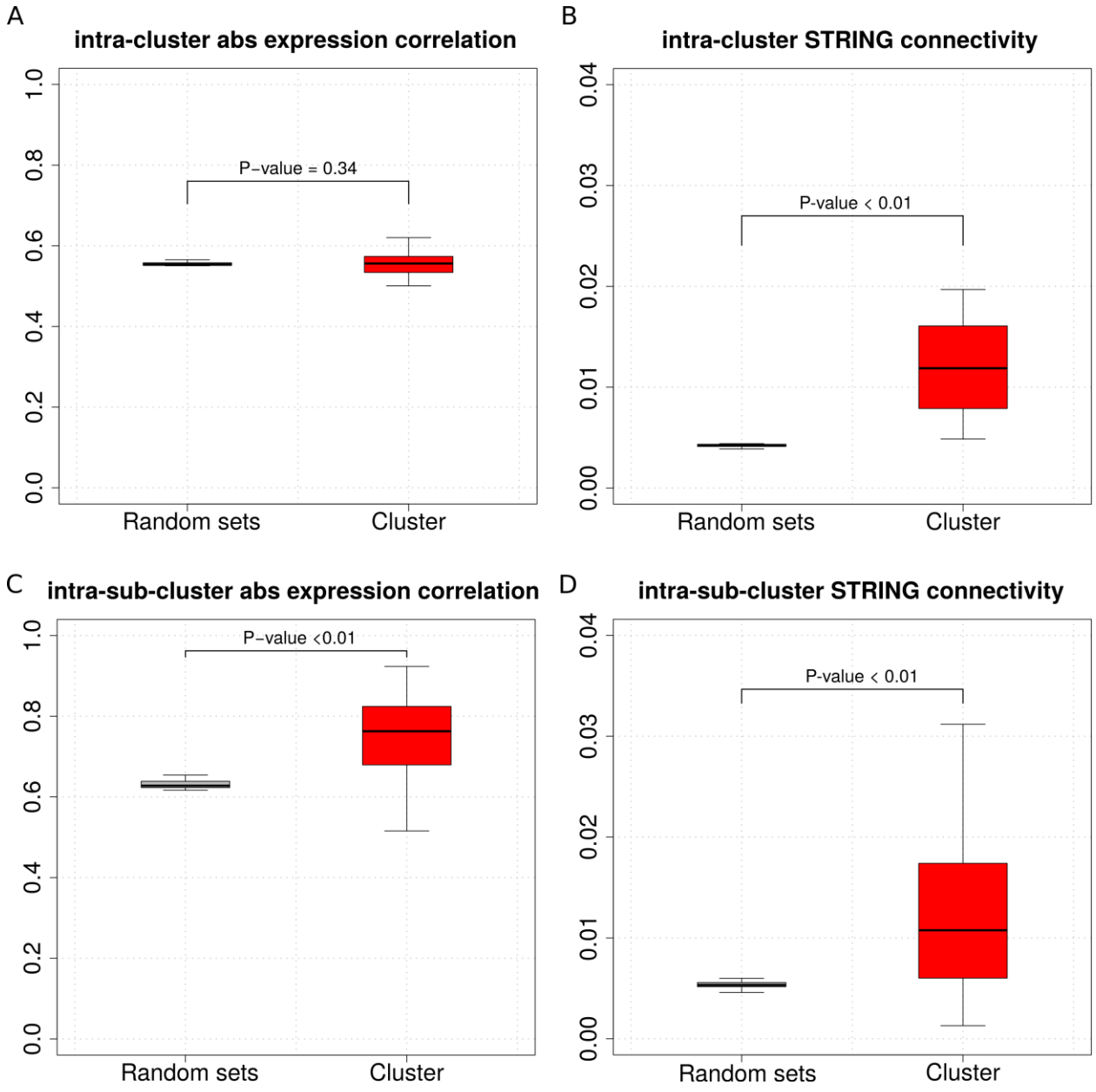

Fig. S1. Cluster and sub-cluster validation for the differentiation of mESCs into embryoid bodies. The boxes represent the interquartile range; the medians are indicated by thick lines; the whiskers extend to the most extreme data points which are no more than 1.5 times of the interquartile range of the data. P-values were determined empirically by sampling 100 random gene sets matching the sizes of the clusters from the set of all genes represented in the 16 clusters (or 51 sub-clusters). A) Average pairwise absolute Pearson correlation coefficient

between the expression profiles of genes in the same cluster, as compared to random gene sets. B) Average intra-cluster STRING connectivity between genes in the same cluster, as compared to random gene sets. C) Average pairwise absolute Pearson correlation coefficient between the expression profiles of genes in the same sub-cluster, as compared to random gene sets. D) Average intra-sub-cluster STRING connectivity between genes in the same sub-cluster, as compared to random gene sets.

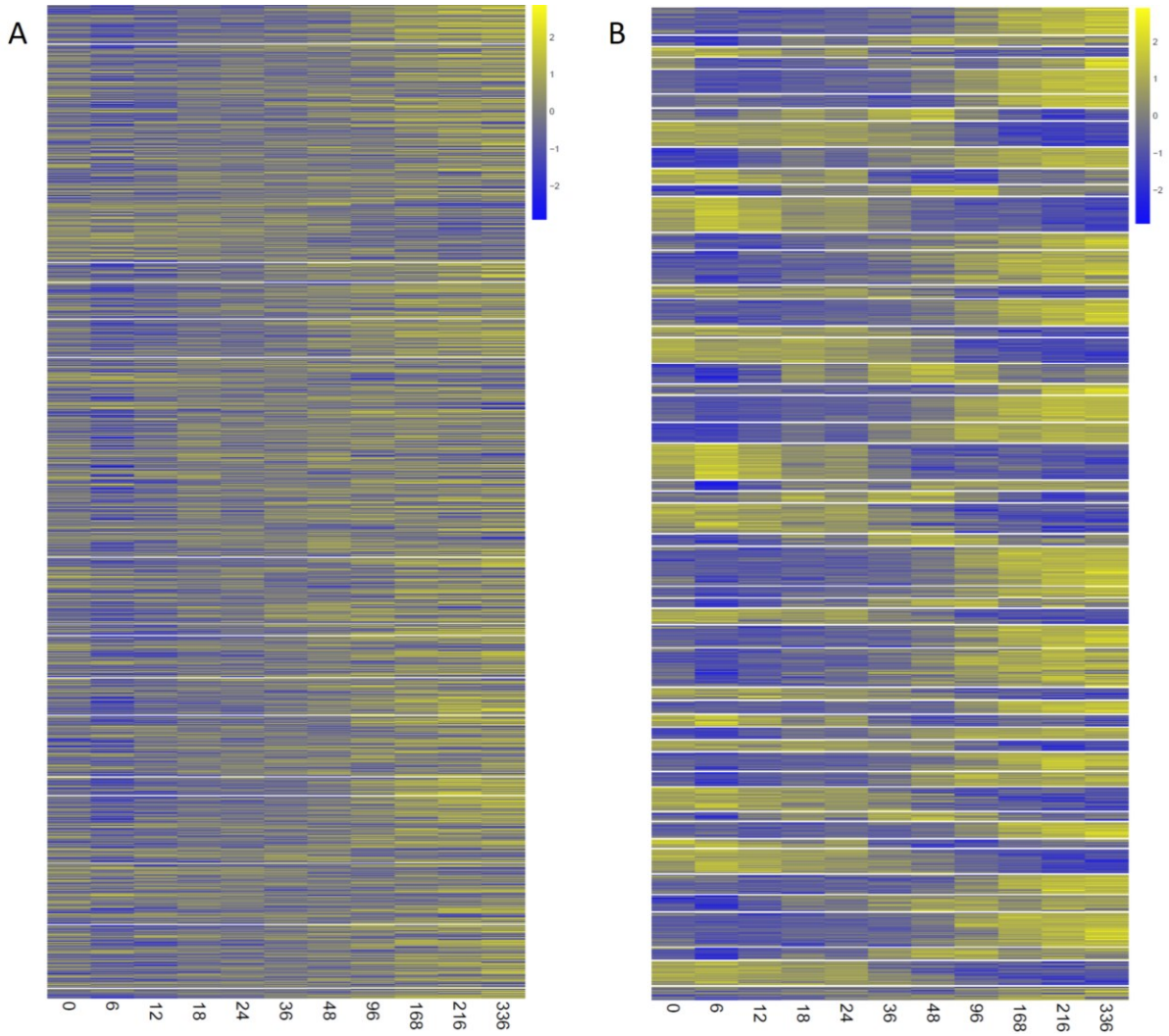

Fig. S2. Heatmap of (A) 16 clusters and (B) 51 sub-clusters obtained for the microarray dataset describing the differentiation of mESCs into embryoid bodies. Clusters and sub-clusters are separated by white bars. Log_2 transformed gene expression values are represented by colors in the

heatmap. The color range of the heatmap is shown by the scale bar. Gene expression values are centered and scaled across time.

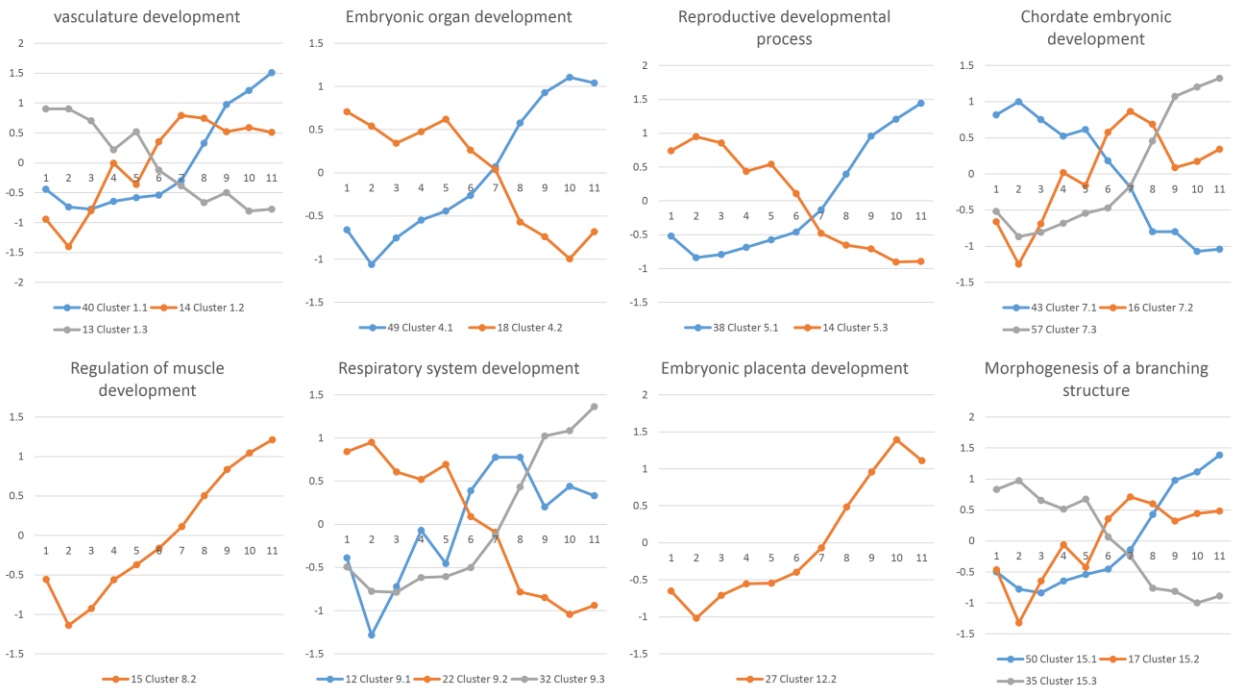

Fig. S3. Plot of expression profiles of sub-clusters belonging to *developmental* clusters in the dataset describing the differentiation of mESCs into embryoid bodies.

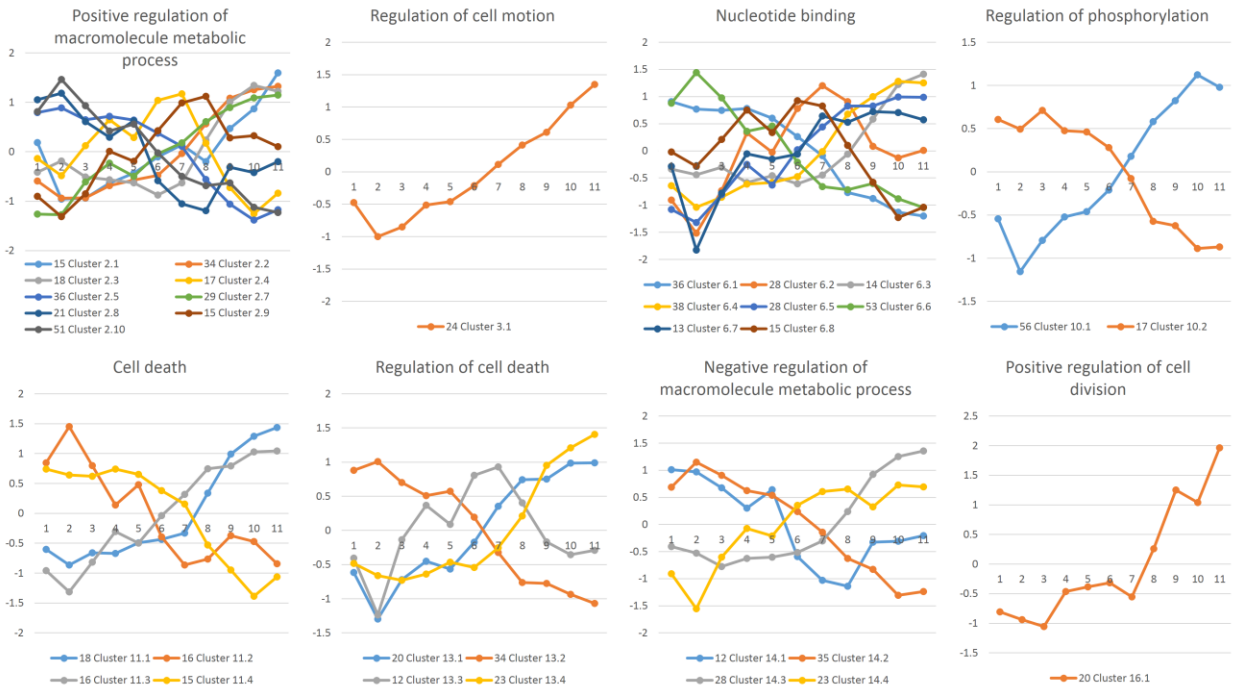

Fig. S4. Plot of expression profiles of sub-clusters related to *non-developmental* clusters in the dataset describing the differentiation of mESCs into embryoid bodies.

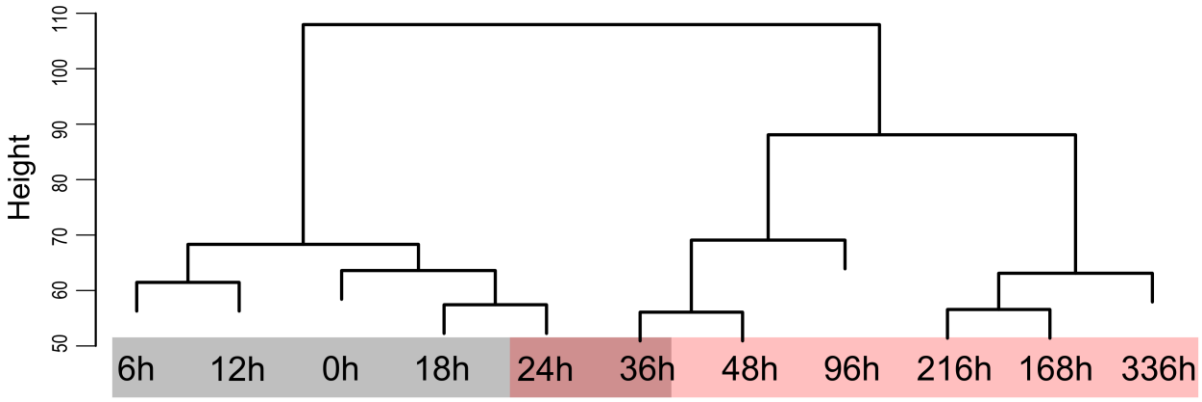

Fig. S5. Dendrogram showing the similarity between time points as mESCs differentiate into embryoid bodies in terms of their gene expression profiles. Similarity between time points was quantified using Euclidean distance. Hierarchical clustering was computed using complete linkage. The time points included in the early and late differentiation datasets are indicated by the color of the background, in grey and red respectively.

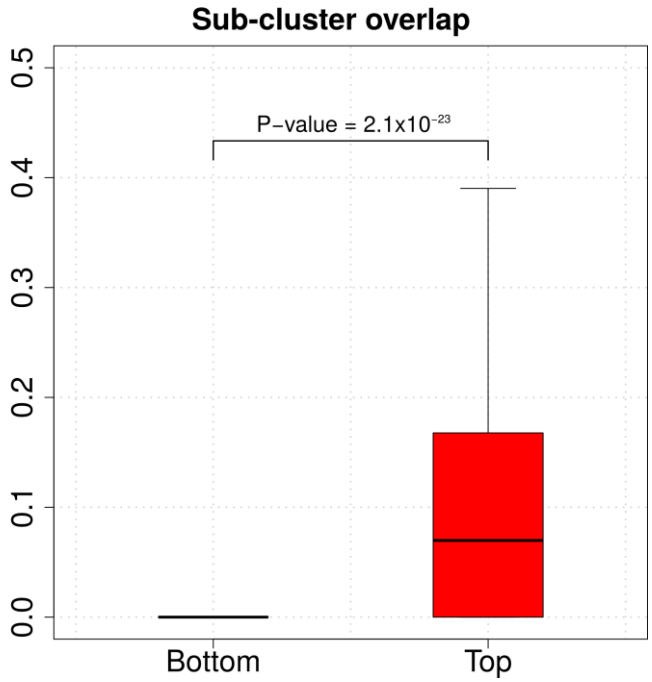

Fig. S6. Comparison of the Average sub-cluster overlap for sub-clusters connected by edges in the top network as compared to the bottom network describing the differentiation of mESCs into embryoid bodies. The boxes represent the interquartile range; the medians are indicated by thick lines; the whiskers extend to the most extreme data points which are no more than 1.5 times interquartile range of the data. P-values were computed using the Wilcoxon Rank-Sum test.

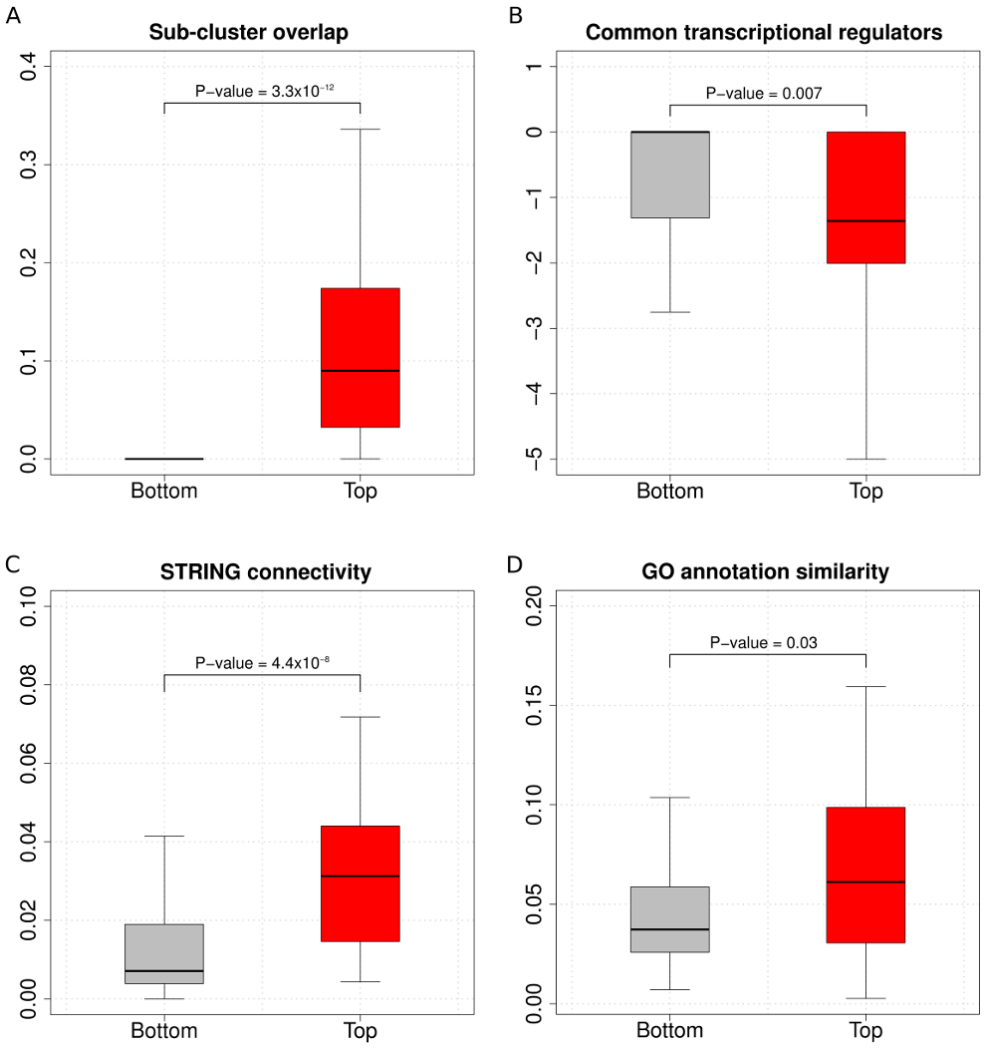

Fig. S7. Comparison of the top and bottom networks describing adult mouse liver regeneration after partial hepatectomy. See Fig. S6 for details regarding the box plots. A) Average sub-cluster overlap for sub-clusters connected by edges in the top network as compared to the bottom network. B) Average significance ($\log_{10}$ P-value) of shared transcriptional regulators binding to the promoters of genes in sub-clusters connected by edges in the top network as compared to the bottom network. C) Average STRING connectivity score between sub-clusters connected by edges in the top network as compared to the bottom network. D) Average GO annotation similarity score between sub-clusters connected by edges in top network as compared to the bottom network.

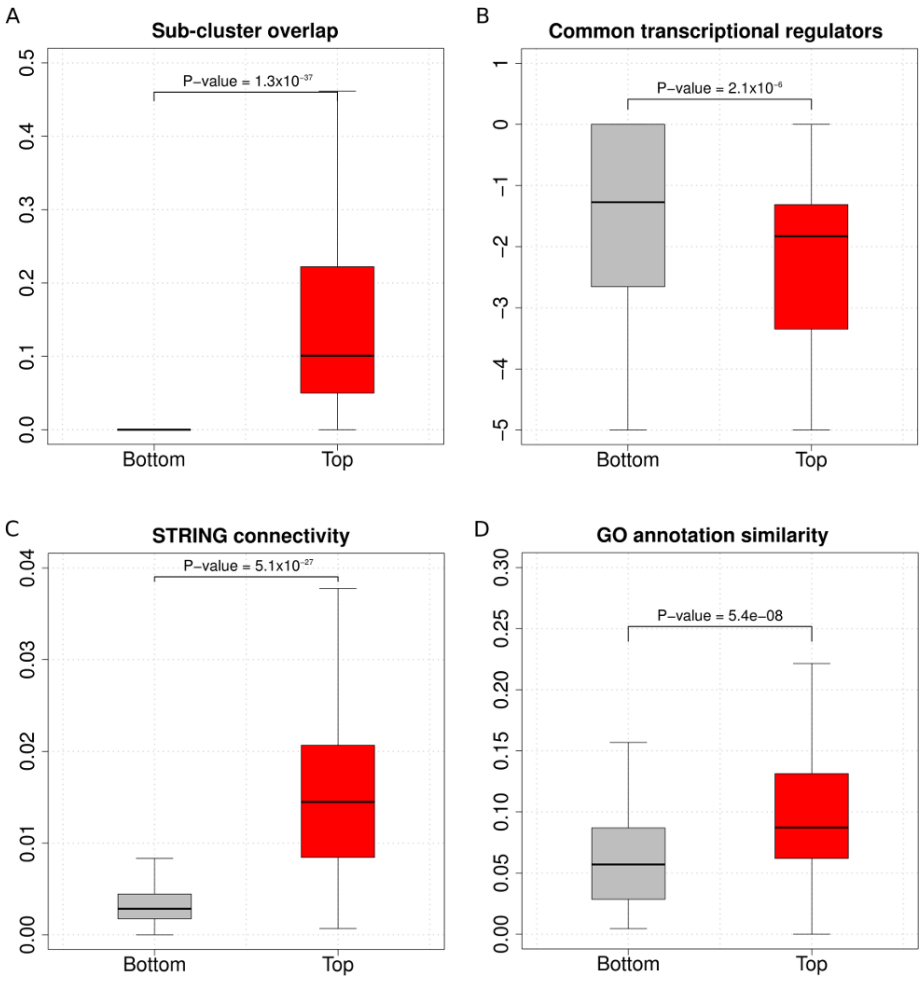

Fig. S8. Comparison of the top and bottom networks describing mouse embryonic liver development. See Fig. S6 for details regarding the box plots. A) Average sub-cluster overlap for sub-clusters connected by edges in the top network as compared to the bottom network. B) Average significance ($\log_{10}$ P-value) of shared transcriptional regulators binding to the promoters of genes in sub-clusters connected by edges in the top network as compared to the bottom network. C) Average STRING connectivity score between sub-clusters connected by edges in the top network as compared to the bottom network. D) Average GO annotation similarity score between sub-clusters connected by edges in top network as compared to the bottom network.

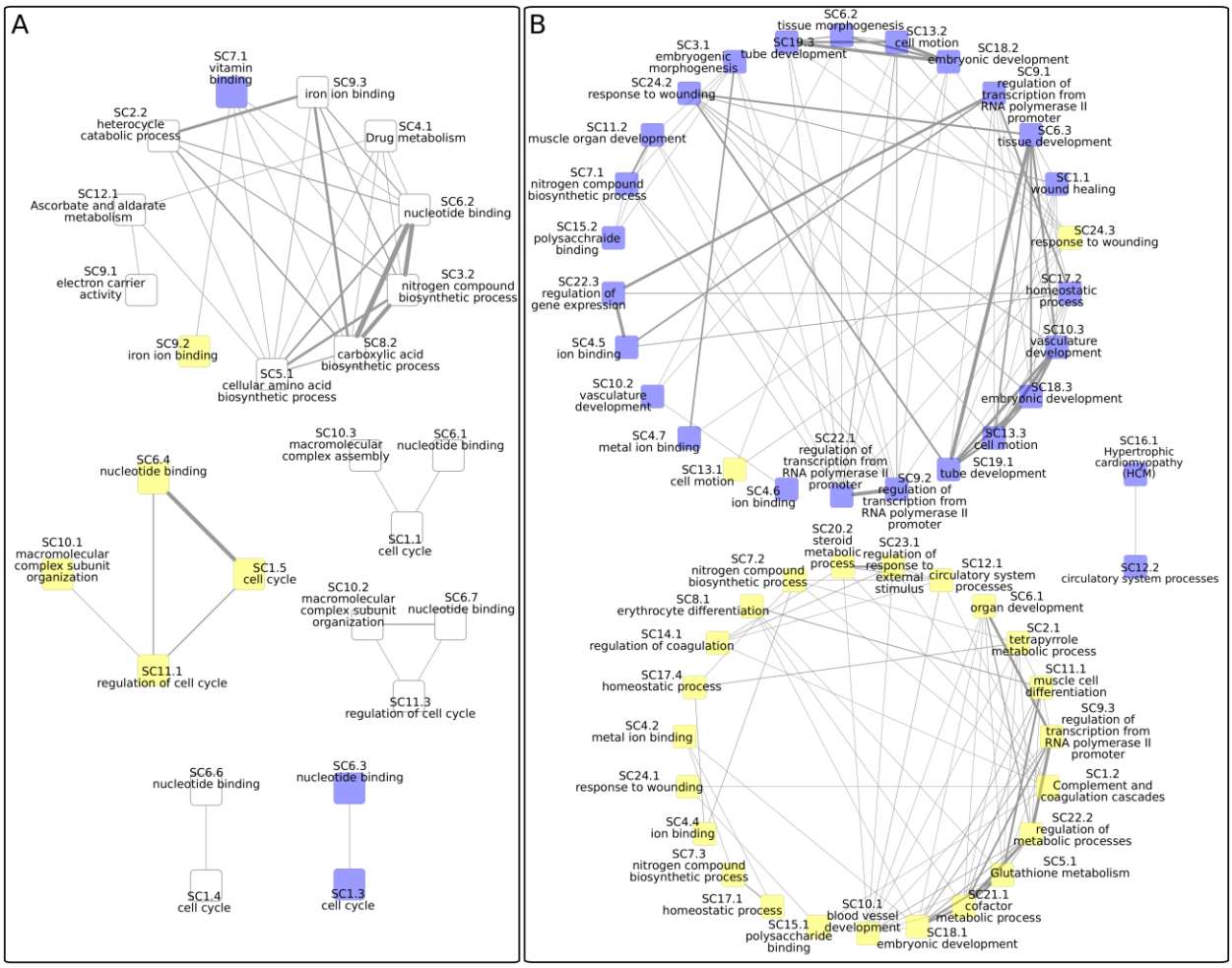

Fig. S9. Top networks describing the mouse adult liver regeneration and mouse embryonic liver development. Node labels include the numbering and first annotation term associated with the sub-clusters (see first and fourth columns in Tables S6 and S7). Nodes are colored in yellow if the $\log_2$ fold-change between the last and the first time point is greater than 1, blue if the $\log_2$ fold-change between the last and the first time point is less than -1, and white otherwise. The thickness of an edge corresponds to Pearson correlation coefficient between expression profiles of connected sub-clusters. A) Liver regeneration network, constructed based on time points 1 hour to 72 hours after partial hepatectomy in adult mice. B) Liver development network, constructed based on time points 10.5 days to 16.5 days post-conception in mice.

Supplementary Tables

Cluster	Size	Annotation
C1*	85	vasculature development; blood vessel development; blood vessel morphogenesis
C2	460	positive regulation of macromolecule metabolic process; regulation of transcription from RNA polymerase II promoter; positive regulation of gene expression
C3	39	regulation of cell motion; regulation of cell migration
C4*	74	embryonic organ development
C5*	79	reproductive developmental process
C6	420	nucleotide binding
C7*	138	chordate embryonic development; embryonic development ending in birth or egg hatching; in utero embryonic development
C8*	22	regulation of muscle development; regulation of striated muscle tissue development
C9*	86	tube development; respiratory system development
C10	77	regulation of phosphorylation
C11	127	death; cell death; programmed cell death
C12*	37	embryonic placenta development; placenta development
C13	140	regulation of cell death; regulation of apoptosis; regulation of programmed cell death
C14	127	negative regulation of macromolecule metabolic process; negative regulation of RNA metabolic process; negative regulation of transcription
C15*	132	tube development; morphogenesis of a branching structure; tube morphogenesis
C16	20	positive regulation of cell division; regulation of cell division

Table S1. Annotation for the clusters identified in the dataset describing the differentiation of mESCs into embryoid bodies. Only the three most significantly associated categories were included as annotation for each cluster. Clusters related to developmental processes are marked with an asterisk.

Sub-cluster	Size	Enrichment against cluster	Enrichment against genome	Degree
SC 1.1*	40		vasculature development; blood vessel development	8
SC 1.2*	14	blood vessel morphogenesis; blood vessel development	vasculature development; blood vessel development	1
SC 1.3*	13		vasculature development; blood vessel morphogenesis	2
SC 2.1	15	regulation of cell differentiation		0
SC 2.2	34	transmembrane receptor protein tyrosine kinase signaling pathway; positive regulation of transcription from RNA polymerase II promoter	regulation of macromolecule metabolic process; regulation of metabolic process	13
SC 2.3	18		regulation of transcription; DNA-dependent	0
SC 2.4	17		DNA binding; nucleic acid binding	0
SC 2.5	36		DNA binding; nucleic acid binding	12

SC 2.7	29		regulation of primary metabolic process; regulation of macromolecule metabolic process	2
SC 2.8	21		DNA binding; regulation of transcription	1
SC 2.9	15		DNA binding; regulation of gene expression	0
SC 2.10	51		DNA binding; nucleic acid binding	5
SC 3.1	24	regulation of cell motion; regulation of localization	regulation of cell motion; regulation of cell migration	2
SC 4.1*	49	embryonic organ development; embryonic placenta development	embryonic organ development; embryonic development	10
SC 4.2*	18	embryonic organ development; embryonic development	embryonic organ development; embryonic development	2
SC 5.1*	38	reproductive developmental process; reproductive process	reproductive developmental process; reproductive process	11
SC 5.3*	14		reproductive developmental process; reproductive process	7
SC 6.1	36		nucleotide binding; purine nucleotide binding	6
SC 6.2	28		nucleotide binding; ribonucleotide binding	0
SC 6.3	14		nucleotide binding; purine nucleotide binding	0
SC 6.4	38		nucleotide binding; purine nucleotide binding	14
SC 6.5	28		nucleotide binding; purine ribonucleotide binding	7
SC 6.6	53	nucleic acid binding; RNA binding	nucleotide binding; ribonucleotide binding	1
SC 6.7	13		purine nucleotide binding; nucleotide binding	0
SC 6.8	15		nucleotide binding; purine ribonucleotide binding	0
SC 7.1*	43	embryonic development ending in birth or egg hatching; chordate embryonic development	embryonic development; chordate embryonic development	12
SC 7.2*	16	embryonic organ development; embryonic development	embryonic development; embryonic organ development	0
SC 7.3*	57	embryonic development ending in birth or egg hatching; embryonic development	embryonic development; chordate embryonic development	11
SC 8.2*	15	regulation of striated muscle tissue development; regulation of muscle development	regulation of striated muscle tissue development; regulation of muscle development	6
SC 9.1*	12	tube morphogenesis; tube development	tube development; tube morphogenesis	1

SC 9.2*	22	tube development; organ development	tube development; organ development	5
SC 9.3*	32	epithelium development; branching morphogenesis of a tube	tube development; tube morphogenesis	10
SC 10.1	56		regulation of phosphorylation; regulation of phosphorus metabolic process	7
SC 10.2	17	regulation of phosphorus metabolic process; regulation of phosphate metabolic process	regulation of phosphorylation; regulation of phosphate metabolic process	4
SC 11.1	18	intracellular signaling cascade; programmed cell death	death; cell death	10
SC 11.2	16	apoptosis; programmed cell death	death; apoptosis	2
SC 11.3	16		apoptosis; programmed cell death	9
SC 11.4	15		death; apoptosis	3
SC 12.2*	27	placenta development; embryonic placenta development	placenta development; embryonic placenta development	10
SC 13.1	20		regulation of apoptosis; regulation of programmed cell death	8
SC 13.2	34	regulation of apoptosis; regulation of programmed cell death	regulation of apoptosis; regulation of programmed cell death	10
SC 13.3	12		regulation of apoptosis; regulation of programmed cell death	0
SC 13.4	23	negative regulation of apoptosis; negative regulation of cell death	regulation of apoptosis; regulation of programmed cell death	7
SC 14.1	12		negative regulation of macromolecule metabolic process; negative regulation of metabolic process	1
SC 14.2	35	negative regulation of macromolecule metabolic process; negative regulation of metabolic process	negative regulation of macromolecule metabolic process; negative regulation of metabolic process	10
SC 14.3	28		negative regulation of cellular metabolic process; negative regulation of macromolecule metabolic process	7
SC 14.4	23	negative regulation of gene expression; negative regulation of macromolecule metabolic process	negative regulation of metabolic process; negative regulation of macromolecule metabolic process	0
SC 15.1*	50	tube development; tube morphogenesis	organ development; tissue development	10
SC 15.2*	17	epithelium development; gland morphogenesis	tube development; epithelium development	2
SC 15.3*	35		organ development; tissue	7

			development	
SC 16.1	20	regulation of cell division; positive regulation of cell division	regulation of cell division; positive regulation of cell division	0

Table S2. Sub-clusters identified in the dataset describing the differentiation of mESCs into embryoid bodies. Only the two most significantly associated categories were included as annotation for each sub-cluster. Sub-clusters with profiles increasing over time (defined as those for which the $\log_2$ fold-change between the last and the first time point in the series is greater than 1) are highlighted in yellow; Sub-clusters with profiles decreasing over time (defined as those for which the $\log_2$ fold-change between the last and the first time point in the series is less than -1) are highlighted in blue. The column *size* indicates the number of genes included in each sub-cluster; *Enrichment against cluster* contains the functional categories enriched among the genes in the sub-cluster when compared to the cluster from which the sub-cluster arose (see Methods), while *Enrichment against genome* contains the functional categories enriched among the genes in the sub-cluster when compared to the entire mouse genome (see Methods); the column *degree* shows the number of sub-clusters interacting with each sub-cluster in the top network (constructed with the 10% of the pairs of sub-clusters with the highest absolute Pearson correlation coefficients, see Methods). The sub-clusters absent in the top network have degree 0. Sub-clusters related to developmental processes are marked with an asterisk.

Sub-cluster	Size	Enrichment against cluster	Enrichment against genome	Degree
SC 1.1	48	regulation of endopeptidase activity; regulation of peptidase activity	death; programmed cell death	4
SC 1.2	62	apoptosis; cell death	cell death; death	2
SC 2.1	59	negative regulation of cellular metabolic process; negative regulation of macromolecule metabolic process	negative regulation of macromolecule metabolic process; negative regulation of metabolic process	1
SC 2.2	44		negative regulation of macromolecule metabolic process; negative regulation of metabolic process	1
SC 3.1	21	mRNA processing; mRNA metabolic process	RNA processing; RNA metabolic process	1
SC 3.2	85	RNA processing; RNA metabolic process	RNA processing; RNA metabolic process	5
SC 4.1	33	transcription factor binding; nervous system development	transcription factor binding; transcription cofactor activity	5
SC 4.2	37	transcription factor binding; transcription cofactor activity	transcription factor binding; transcription cofactor activity	1
SC 5.1	46	ribonucleoprotein complex biogenesis; RNA processing	ribonucleoprotein complex biogenesis; ribosome biogenesis	3

SC 6.1	121	nucleotide binding	nucleotide binding; purine nucleotide binding	3
SC 6.2	16		nucleotide binding; purine nucleotide binding	1
SC 6.3	14		nucleotide binding; purine nucleotide binding	0
SC 7.1	22		localization; protein localization	0
SC 7.2	16		macromolecule localization; protein localization	0
SC 7.3	60		protein localization; macromolecule localization	5
SC 8.1	16		posttranscriptional regulation of gene expression; regulation of translation	1
SC 8.2	21	posttranscriptional regulation of gene expression; RNA binding	posttranscriptional regulation of gene expression; regulation of translation	0
SC 9.1	15		protein catabolic process; macromolecule catabolic process	0
SC 10.1	52	cell cycle phase; interphase of mitotic cell cycle	cell cycle; cell cycle process	3
SC 10.2	80	cell cycle; cell cycle process	cell cycle; cell cycle process	2

Table S3. Sub-clusters identified in the *early differentiation* dataset describing the early differentiation of mESCs into embryoid bodies. See Table S2 for further details.

Sub-cluster	Size	Enrichment against cluster	Enrichment against genome	Degree
SC 1.1	68	blood vessel development; labyrinthine layer development	chordate embryonic development; embryonic development ending in birth or egg hatching	7
SC 1.2	52	embryonic development ending in birth or egg hatching; chordate embryonic development	chordate embryonic development; embryonic development ending in birth or egg hatching	13
SC 2.1	23		negative regulation of macromolecule metabolic process; negative regulation of metabolic process	4
SC 2.3	13		negative regulation of macromolecule biosynthetic process; negative regulation of cellular biosynthetic process	0
SC 3.1	21	cell cycle; cell cycle process	cell cycle; cell cycle process	1

SC 3.2	46	cell cycle; cell cycle process	cell cycle; cell cycle process	7
SC 3.3	41	nucleobase; nucleoside	cell cycle; cell cycle process	2
SC 4.2	18		translation factor activity; nucleic acid binding	1
SC 5.1	24	placenta development; embryonic placenta development	placenta development; embryonic placenta development	5
SC 6.1	13		macromolecular complex subunit organization; macromolecular complex assembly	1
SC 6.2	16	macromolecular complex assembly; macromolecular complex subunit organization	macromolecular complex subunit organization; cellular macromolecular complex subunit organization	5
SC 6.3	27	protein complex assembly; protein complex biogenesis	macromolecular complex subunit organization; macromolecular complex assembly	8
SC 7.1	32	tube morphogenesis; epithelial tube morphogenesis	tube development; organ development	3
SC 7.2	50	tube development; lung development	tube development; tube morphogenesis	10
SC 8.1	19		regulation of programmed cell death; regulation of cell death	1
SC 8.2	57	regulation of cell death; regulation of apoptosis	regulation of cell death; regulation of programmed cell death	13
SC 8.3	12	regulation of cell death; regulation of programmed cell death	regulation of apoptosis; regulation of programmed cell death	1
SC 9.1	41	transcription factor binding; negative regulation of transcription from RNA polymerase II promoter	transcription factor binding; transcription cofactor activity	3
SC 9.2	37	transcription factor binding; transcription cofactor activity	transcription factor binding; transcription cofactor activity	12
SC 10.1	61	ATP binding; adenyl ribonucleotide binding	nucleotide binding; purine nucleotide binding	11
SC 10.2	40	nucleotide binding; adenyl nucleotide binding	nucleotide binding; purine nucleotide binding	12
SC 10.3	33		nucleotide binding; purine nucleotide binding	1
SC 10.4	38		nucleotide binding; purine ribonucleotide binding	0
SC 10.5	38		nucleotide binding; adenyl nucleotide binding	0
SC 11.1	18		nucleobase; nucleoside	0
SC 12.2	25	epithelial cell differentiation involved in prostate gland development; cell maturation	morphogenesis of a branching structure; branching morphogenesis of a tube	9
SC 13.1	19		DNA replication; DNA replication	1
SC 14.2	21	one-carbon metabolic process; biopolymer methylation	one-carbon metabolic process; methylation	3
SC 15.1	27		RNA processing; RNA metabolic	1

			process	
SC 15.2	11		RNA metabolic process; mRNA metabolic process	9
SC 15.3	22		RNA processing; RNA metabolic process	1
SC 16.1	28		Glutathione metabolism; transferase activity	0
SC 17.1	31		nucleobase; nucleoside	0
SC 18.1	19	reproductive developmental process; reproduction	reproductive developmental process; reproductive process	3
SC 18.2	43	reproductive developmental process; reproductive process	reproductive developmental process; reproductive process	11
SC 19.1	53	blood vessel development; vasculature development	vasculature development; blood vessel development	12
SC 19.2	17	vasculature development; blood vessel development	vasculature development; blood vessel development	4
SC 20.1	54	ribosome biogenesis; ribonucleoprotein complex biogenesis	ribonucleoprotein complex biogenesis; ribosome biogenesis	0
SC 21.1	23		macromolecule localization; protein localization	1
SC 21.2	29		protein localization; macromolecule localization	4
SC 21.3	26		protein localization; protein transport	3
SC 21.4	36		protein localization; macromolecule localization	3
SC 22.1	35	DNA metabolic process; DNA replication	DNA metabolic process; DNA replication	1
SC 22.2	30	negative regulation of DNA recombination; regulation of DNA recombination	DNA metabolic process; cellular response to stress	1
SC 22.3	43	regulation of metabolic process; DNA metabolic process	cellular response to stress; cellular response to stimulus	10

Table S4. Sub-clusters identified in the *late differentiation* dataset describing the late differentiation of mESCs into embryoid bodies. See Table S2 for further details.

Early differentiation Sub-cluster	Late differentiation Sub-cluster	Mapping character	Early differentiation Sub-cluster Annotation	Late differentiation Sub-cluster Annotation
SC 1.1	SC 8.3	A	death; programmed cell death	regulation of apoptosis; regulation of programmed cell death
SC 4.1	SC 9.2	B	transcription factor binding; transcription cofactor activity	transcription factor binding; transcription cofactor activity
SC 4.2	SC 9.1	C	transcription factor binding; transcription cofactor activity	transcription factor binding; transcription cofactor activity
SC 5.1	SC 20.1	D	ribonucleoprotein complex biogenesis; ribosome	ribonucleoprotein complex biogenesis; ribosome

			biogenesis	biogenesis
SC 6.1	SC 10.5	E	nucleotide binding; purine nucleotide binding	nucleotide binding; adenyl nucleotide binding
SC 6.3	SC 10.3	F	nucleotide binding; purine nucleotide binding	nucleotide binding; purine nucleotide binding
SC 7.1	SC 21.4	G	localization; protein localization	protein localization; macromolecule localization
SC 7.2	SC 21.1	H	macromolecule localization; protein localization	macromolecule localization; protein localization
SC 7.3	SC 21.3	I	protein localization; macromolecule localization	protein localization; protein transport
SC 10.1	SC 3.3	J	cell cycle; cell cycle process	cell cycle; cell cycle process
SC 10.2	SC 3.2	K	cell cycle; cell cycle process	cell cycle; cell cycle process

Table S5. Mapping of sub-clusters from *early differentiation* dataset to sub-clusters from *late differentiation* dataset. The mapping character is a common character to a pair of mapped sub-clusters that will be used to identify mapped sub-clusters in the networks. After visual inspection we discarded the mapping between SC 6.2 ("nucleotide binding") in the *early differentiation* dataset and SC 3.1 ("cell cycle") in the *late differentiation* dataset.

Sub-cluster	Size	Enrichment against cluster	Enrichment against genome	Degree
SC 1.1	19		cell cycle; cell cycle process	2
SC 1.2	11		cell cycle phase; cell cycle	0
SC 1.3	13		cell cycle	1
SC 1.4	13		cell cycle	1
SC 1.5	67		cell cycle; cell cycle process	2
SC 2.2	13		heterocycle catabolic process; heterocycle metabolic process	5
SC 3.1	16		nitrogen compound biosynthetic process; nucleobase	0
SC 3.2	41	nitrogen compound biosynthetic process; cellular biosynthetic process	nitrogen compound biosynthetic process; amine biosynthetic process	7
SC 4.1	17		Drug metabolism; Metabolism of xenobiotics by cytochrome P	5
SC 5.1	12		cellular amino acid biosynthetic process; amine biosynthetic process	8
SC 6.1	27	mitochondrion organization; nucleotide binding	nucleotide binding; purine nucleotide binding	1
SC 6.2	49		nucleotide binding; purine nucleotide binding	7
SC 6.3	27		nucleotide binding; purine nucleotide binding	1
SC 6.4	52		nucleotide binding; purine nucleotide binding	2
SC 6.5	19		nucleotide binding; RNA binding	0

SC 6.6	31	RNA binding; nucleic acid binding	nucleotide binding; ribonucleotide binding	1
SC 6.7	40		nucleotide binding; purine nucleotide binding	2
SC 7.1	35		vitamin binding; vitamin B binding	5
SC 8.2	25		carboxylic acid biosynthetic process; organic acid biosynthetic process	7
SC 9.1	17		electron carrier activity; oxidoreductase activity	1
SC 9.2	15		iron ion binding; electron carrier activity	1
SC 9.3	28	iron ion binding; transition metal ion binding	iron ion binding; oxidoreductase activity	5
SC 10.1	16	cellular macromolecular complex subunit organization; macromolecular complex subunit organization	macromolecular complex subunit organization; cellular macromolecular complex subunit organization	1
SC 10.2	25		macromolecular complex subunit organization; macromolecular complex assembly	2
SC 10.3	18	macromolecular complex assembly; macromolecular complex subunit organization	macromolecular complex assembly; macromolecular complex subunit organization	1
SC 11.1	15	regulation of cell cycle; regulation of mitotic cell cycle	regulation of cell cycle; regulation of cell cycle process	3
SC 11.3	12		regulation of cell cycle	2
SC 12.1	12		Ascorbate and aldarate metabolism; Pentose and glucuronate interconversions	3

Table S6. Sub-clusters identified in the mouse *liver regeneration* dataset describing the regeneration of mouse liver. See Table S2 for further details.

Sub-cluster	Size	Enrichment against cluster	Enrichment against genome	Degree
SC 1.1	16	system development; anatomical structure development	wound healing; response to wounding	7
SC 1.2	42		Complement and coagulation cascades; response to wounding	7
SC 2.1	25		tetrapyrrole metabolic process; porphyrin metabolic process	7
SC 3.1	84		embryonic morphogenesis; embryonic development	6
SC 4.1	24		metal ion binding; cation binding	0
SC 4.2	53	transition metal ion binding; ion binding	metal ion binding; cation binding	3
SC 4.3	30		metal ion binding; cation binding	0
SC 4.4	45	proteolysis; catalytic activity	ion binding; metal ion binding	2
SC 4.5	73		ion binding; metal ion binding	3
SC 4.6	38	calcium ion binding; ion binding	ion binding; metal ion binding	1

SC 4.7	67		metal ion binding; cation binding	2
SC 5.1	22		Glutathione metabolism; glutathione transferase activity	9
SC 6.1	17		organ development; system development	6
SC 6.2	21		tissue morphogenesis; morphogenesis of an epithelium	6
SC 6.3	37	tube development; cell differentiation	tissue development; organ development	7
SC 7.1	11		nitrogen compound biosynthetic process; nucleotide biosynthetic process	5
SC 7.2	16		nitrogen compound biosynthetic process; amine biosynthetic process	5
SC 7.3	12		nitrogen compound biosynthetic process; heterocycle metabolic process	2
SC 8.1	16		erythrocyte differentiation; erythrocyte homeostasis	4
SC 9.1	36		regulation of transcription from RNA polymerase II promoter; regulation of transcription	7
SC 9.2	39		regulation of transcription from RNA polymerase II promoter; regulation of transcription	8
SC 9.3	49		regulation of transcription from RNA polymerase II promoter; regulation of transcription	7
SC 10.1	18	apoptosis; positive regulation of apoptosis	blood vessel development; vasculature development	10
SC 10.2	12		vasculature development; blood vessel development	3
SC 10.3	25	blood vessel development; vasculature development	vasculature development; blood vessel development	8
SC 11.1	12	apoptosis; positive regulation of programmed cell death	muscle cell differentiation; cell differentiation	7
SC 11.2	53		muscle organ development; muscle tissue development	5
SC 12.1	18	blood circulation; circulatory system process	circulatory system process; blood circulation	6
SC 12.2	20		circulatory system process; blood circulation	1
SC 13.1	18		cell motion; cell motility	2
SC 13.2	14		cell motion; cellular developmental process	7
SC 13.3	19		cell motion; locomotion	8
SC 14.1	14	regulation of coagulation; regulation of multicellular organismal process	regulation of coagulation; regulation of blood coagulation	4
SC 15.1	19	polysaccharide binding; pattern binding	polysaccharide binding; pattern binding	2
SC 15.2	21		polysaccharide binding; pattern binding	3
SC 16.1	21		Hypertrophic cardiomyopathy (HCM);	1

			Dilated cardiomyopathy	
SC 17.1	26	cellular homeostasis; homeostatic process	homeostatic process; regulation of biological quality	2
SC 17.2	23		homeostatic process; regulation of biological quality	8
SC 17.4	30	homeostatic process; regulation of biological quality	homeostatic process; regulation of biological quality	3
SC 18.1	27	in utero embryonic development; embryonic development ending in birth or egg hatching	embryonic development; embryonic development ending in birth or egg hatching	9
SC 18.2	31		embryonic development; embryonic morphogenesis	7
SC 18.3	39	negative regulation of biological process; transition metal ion binding	embryonic development; anatomical structure morphogenesis	8
SC 19.1	18		tube development; embryonic development	8
SC 19.3	12	tube development; tissue morphogenesis	tube development; tube morphogenesis	6
SC 20.2	43		steroid metabolic process; lipid metabolic process	5
SC 21.1	54		cofactor metabolic process; cofactor biosynthetic process	9
SC 22.1	42		regulation of transcription from RNA polymerase II promoter; regulation of transcription	9
SC 22.2	62		regulation of metabolic process; regulation of macromolecule metabolic process	8
SC 22.3	53	regulation of transcription from RNA polymerase II promoter; regulation of transcription	regulation of gene expression; regulation of macromolecule biosynthetic process	3
SC 23.1	28	regulation of response to external stimulus; regulation of response to stress	regulation of response to external stimulus; regulation of response to stress	5
SC 24.1	28	proteolysis; Complement and coagulation cascades	response to wounding; response to external stimulus	2
SC 24.2	34		response to wounding; response to external stimulus	6
SC 24.3	18		response to wounding; response to external stimulus	7

Table S7. Sub-clusters identified in the mouse *liver development* dataset describing the development of mouse liver. See Table S2 for further details.

References

1. H. H. Otu, K. Naxerova, K. Ho, H. Can, N. Nesbitt, T. A. Libermann and S. J. Karp, *J Biol Chem*, 2007, **282**, 11197-11204.
2. J. A. Blake, J. E. Richardson, M. T. Davisson and J. T. Eppig, *Nucleic Acids Res*, 1997, **25**, 85-91.
3. G. K. Smyth, *Stat Appl Genet Mol Biol*, 2004, **3**, Article3.
4. Y. Benjamini and Y. Hochberg, *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, **57, No. 1**, 289-300.
5. T. Kambe, B. P. Weaver and G. K. Andrews, *genesis*, 2008, **46**, 214-228.
6. U. A. Kohler, S. Kurinna, D. Schwitter, A. Marti, M. Schafer, C. Hellerbrand, T. Speicher and S. Werner, *Hepatology*, 2013.
7. G. K. Michalopoulos, *J. Cell. Physiol.*, 2007, **213**, 286-300.
8. H.-C. Hsieh, Y.-T. Chen, J.-M. Li, T.-Y. Chou, M.-F. Chang, S.-C. Huang, T.-L. Tseng, C.-C. Liu and S.-F. Chen, *Journal of Proteome Research*, 2009, **8**, 1004-1013.
9. R. Zhao and S. A. Duncan, *Hepatology*, 2005, **41**, 956-967.