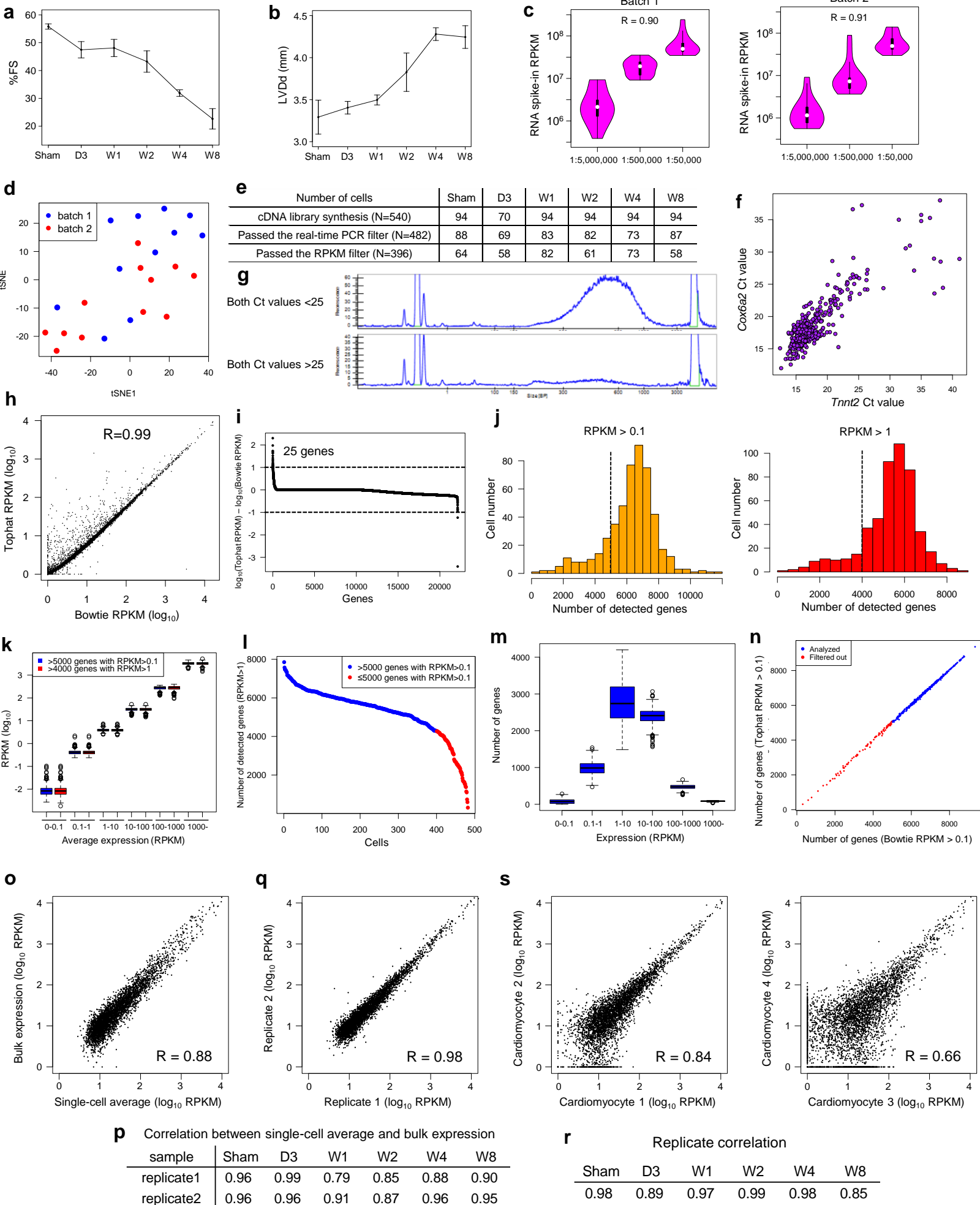


Supplementary Figures

Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure

Nomura et al.

Supplementary Figure 1

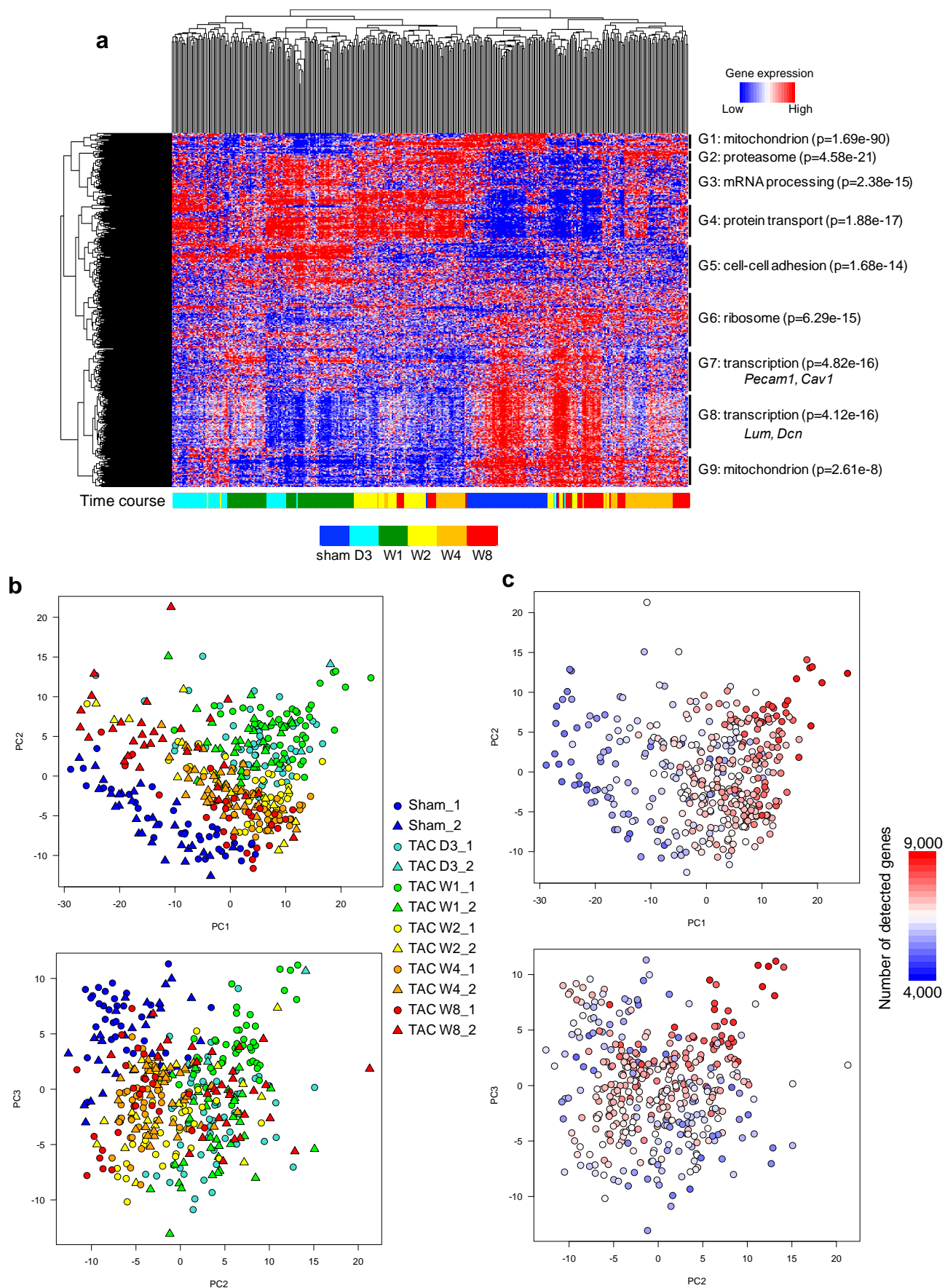


(legend on next page)

Supplementary Figure 1. Single-cardiomyocyte RNA-seq profiles.

- (a,b) Echocardiographic assessment of the mice analyzed in this study. Line graphs show fractional shortening (%FS) (a) and left ventricular diastolic diameter (LVDd) (b). Mean and standard error of the mean are shown ($n = 3-4$ at each time point).
- (c) Violin plot showing the expression levels of RNA spikes (RPKM) at the indicated concentrations ($n = 15$ at each concentration). A good correlation between the RPKM values and concentrations supports the accuracy of our single-cell RNA-seq analysis.
- (d) t-SNE plot using single-cardiomyocyte transcriptomes from normal C57BL/6 mice (RPKM values) in 2 different batches. Cardiomyocytes from different batches could not be classified.
- (e) Table showing the number of samples during the filtering steps.
- (f) Scatter plot showing the Ct values (*Tnnt2* and *Cox6a2*) of single-cardiomyocyte cDNA libraries ($n = 540$).
- (g) Representative results of capillary electrophoresis to assess the efficacy of the reverse transcription and amplification steps.
- (h) Scatter plot showing the correlation between averaged single-cell RPKM values of sham cardiomyocytes mapped by Tophat and Bowtie.
- (i) Plot showing the differences between averaged single-cell RPKM values of sham cardiomyocytes mapped by Tophat and Bowtie. There were 25 genes that were identified to be 10-fold higher in Tophat mapping RPKM values. Gene ontology analysis revealed that ribosomal protein genes were enriched in these genes ($p = 7.9e-6$).
- (j) Bar plots showing the distribution of the number of detected genes with RPKM > 0.1 (left) and RPKM > 1 (right) across all cardiomyocytes ($n = 482$). More than 5,000 genes were detected with RPKM > 0.1 in 82.1% of cardiomyocytes (396/482 cells), and more than 4,000 genes were detected with RPKM > 1 in 86.9% of cardiomyocytes (419/482 cells).
- (k) Boxplot showing the averaged expression levels of all cardiomyocytes ($n = 482$). Genes are categorized by the expression levels. Expression levels were comparable between cardiomyocytes that had more than 5,000 genes with RPKM > 0.1 ($n = 396$) and those that had more than 4,000 genes with RPKM > 1 ($n = 419$). Horizontal lines indicate the medians. Boxes show the 25th–75th percentiles. Whiskers represent the minimum and maximum values.
- (l) Plot showing the number of detected genes (RPKM > 1). Among all of the cardiomyocytes that had more than 5,000 genes with RPKM > 0.1 ($n = 396$), over 4,000 genes were detected with RPKM > 1. Cells are ordered by the number of detected genes.
- (m) Boxplot showing the number of genes in cardiomyocytes that had more than 5,000 genes with RPKM > 0.1 ($n = 396$). Genes are categorized by the expression levels. Horizontal lines indicate the medians. Boxes show the 25th–75th percentiles. Whiskers represent the minimum and maximum values.
- (n) Scatter plot showing the numbers of detected genes (RPKM > 0.1) mapped by Tophat and Bowtie. The difference of mapping tools does not affect the number of detected genes both in cardiomyocytes analyzed in this study ($n = 396$, blue) and in those filtered out through the quality control process ($n = 86$, red).
- (o) Scatter plot showing the correlation between the single-cell average (\log_{10} RPKM) and bulk expression (\log_{10} RPKM) in cardiomyocytes from mice at 4 weeks after pressure overload (replicate 1).
- (p) Table showing the correlations between the single-cell average and bulk expression at each time point.
- (q) Scatter plot showing the correlation between biological replicates of single-cell average transcriptomes from mice at 4 weeks after pressure overload.
- (r) Table showing the correlations between biological replicates of single-cell average transcriptomes at each time point.
- (s) Representative scatter plots showing the correlations of single-cell transcriptomes between cardiomyocytes from the same mice (4 weeks after pressure overload). Relationships between cardiomyocytes in the same clusters (left) and in the different clusters (right) are shown.

Supplementary Figure 2

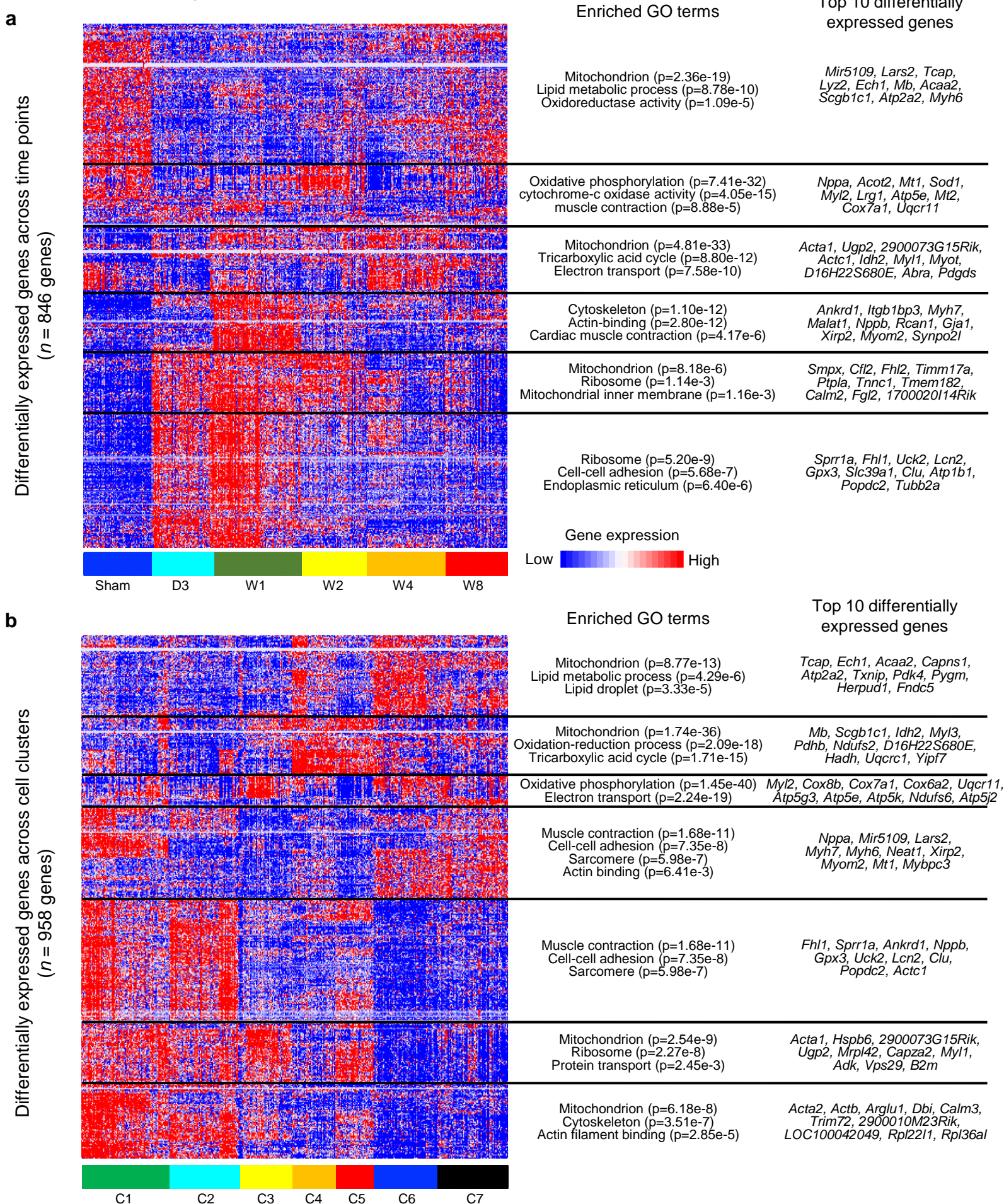


Supplementary Figure 2. Global landscape of single-cardiomyocyte RNA-seq.

(a) Hierarchical clustering of all cardiomyocytes ($n = 396$) using all genes ($n = 10,720$) expressed (RPKM ≥ 5) in at least 1% of cardiomyocytes.

(b,c) PCA plot colored by batches (b) and by the number of detected genes (c).

Supplementary Figure 3

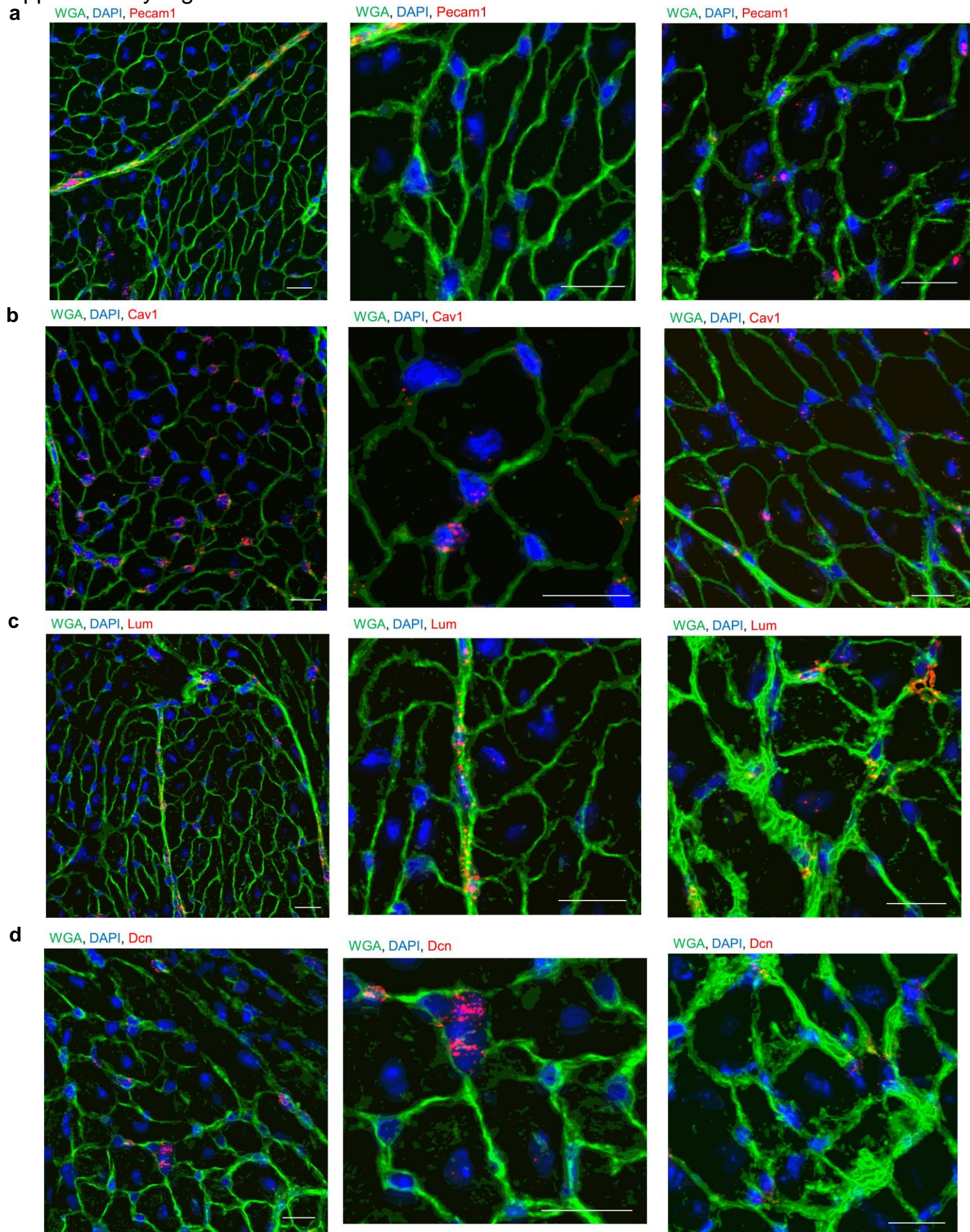


Supplementary Figure 3. Differential expression analysis.

(a,b) Heatmap showing the expression levels of genes differentially expressed across the time points (a) and cell clusters (b).

Top 10 differentially expressed genes are also denoted.

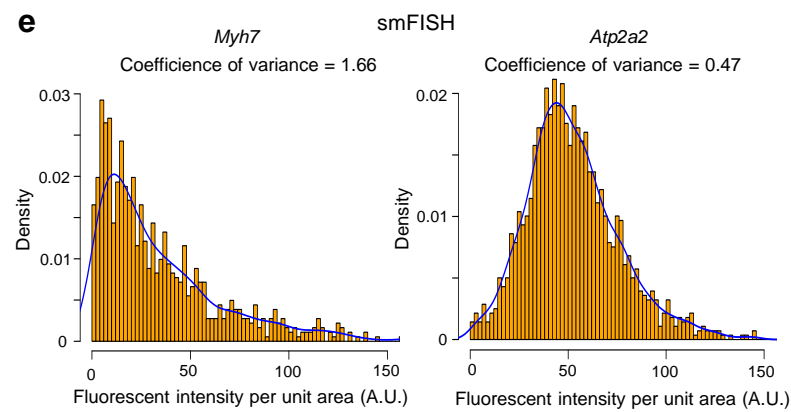
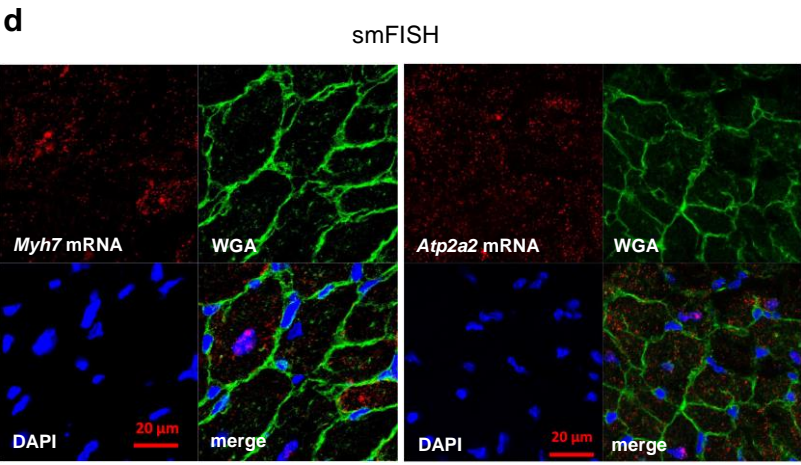
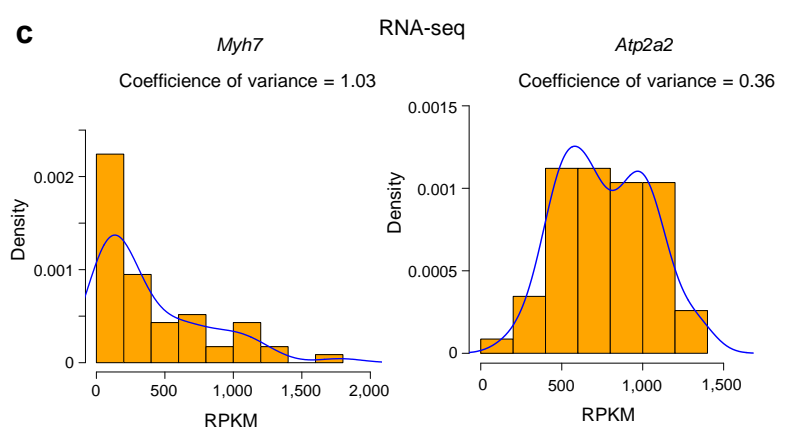
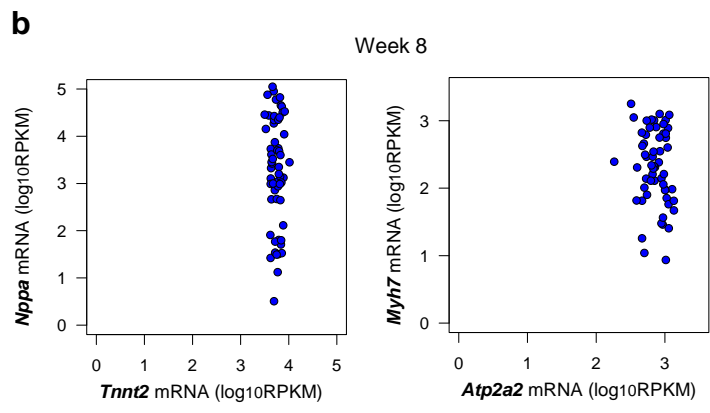
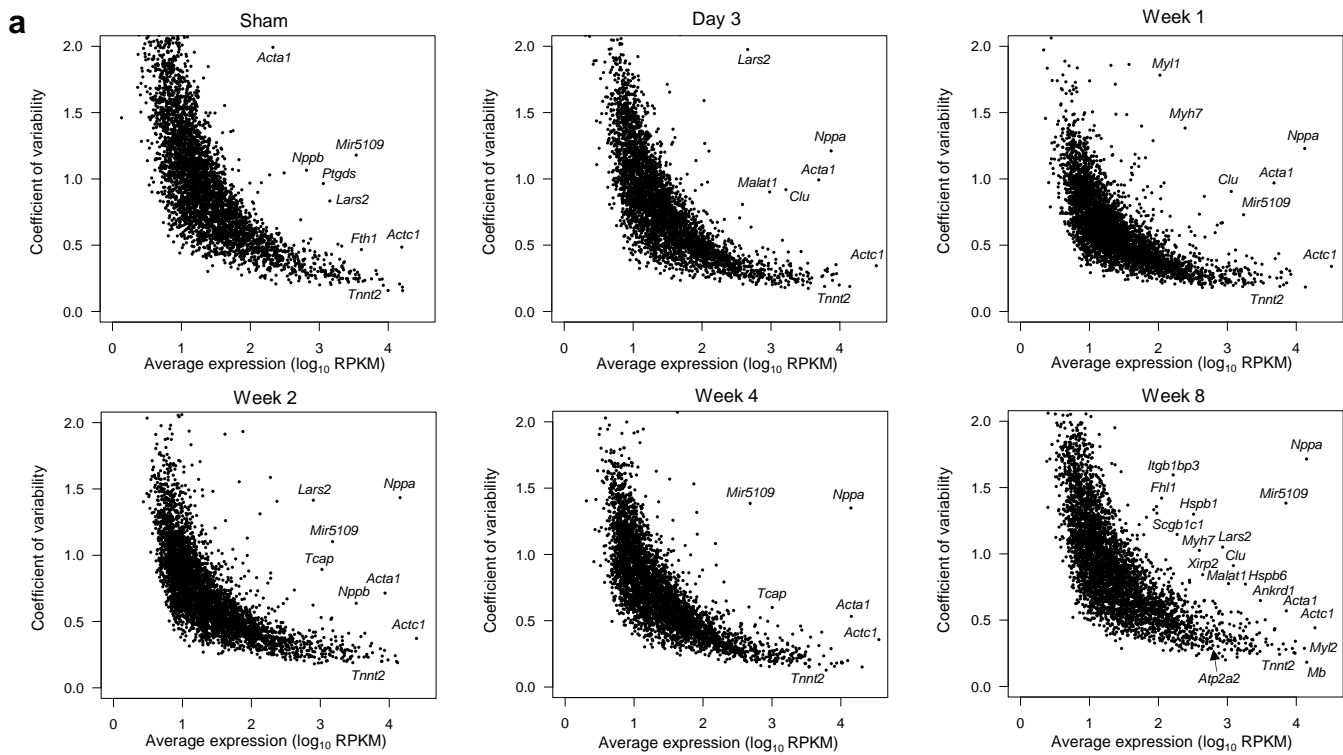
Supplementary Figure 4



Supplementary Figure 4. Single-molecule RNA *in situ* hybridization of endothelial and fibrotic genes.

(a-d) Single-molecule RNA *in situ* hybridization of endothelial and fibrotic genes (a, *Pecam1*; b, *Cav1*; c, *Lum*; d, *Dcn*) in the heart from mice after sham (left and middle) and TAC (right) operation (week 8). Scale bar, 20 μ m.

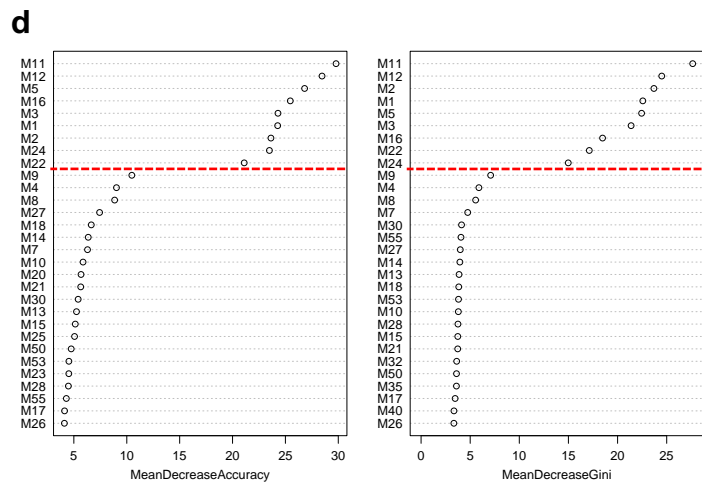
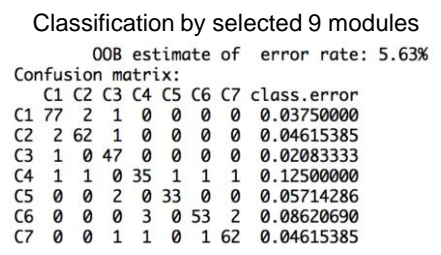
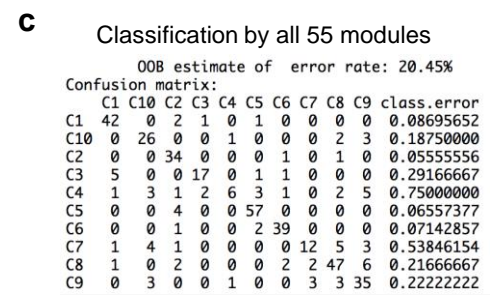
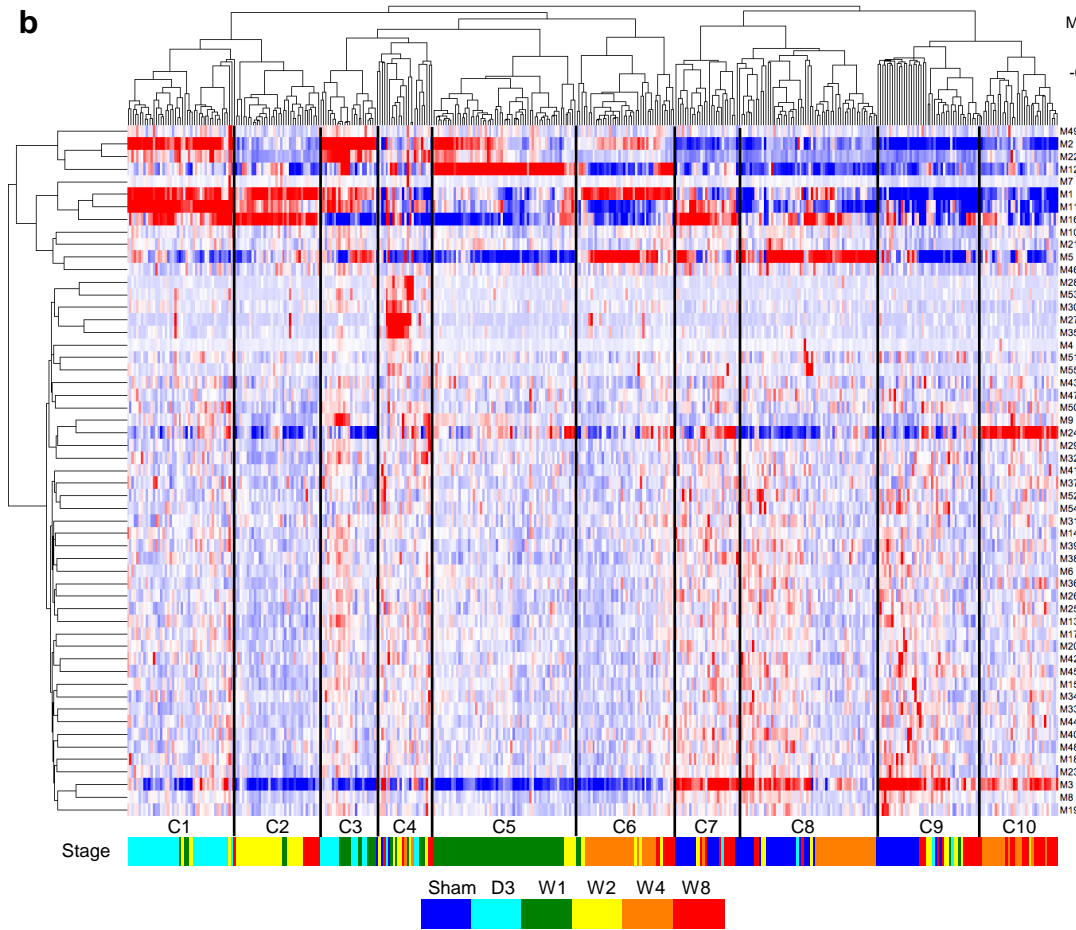
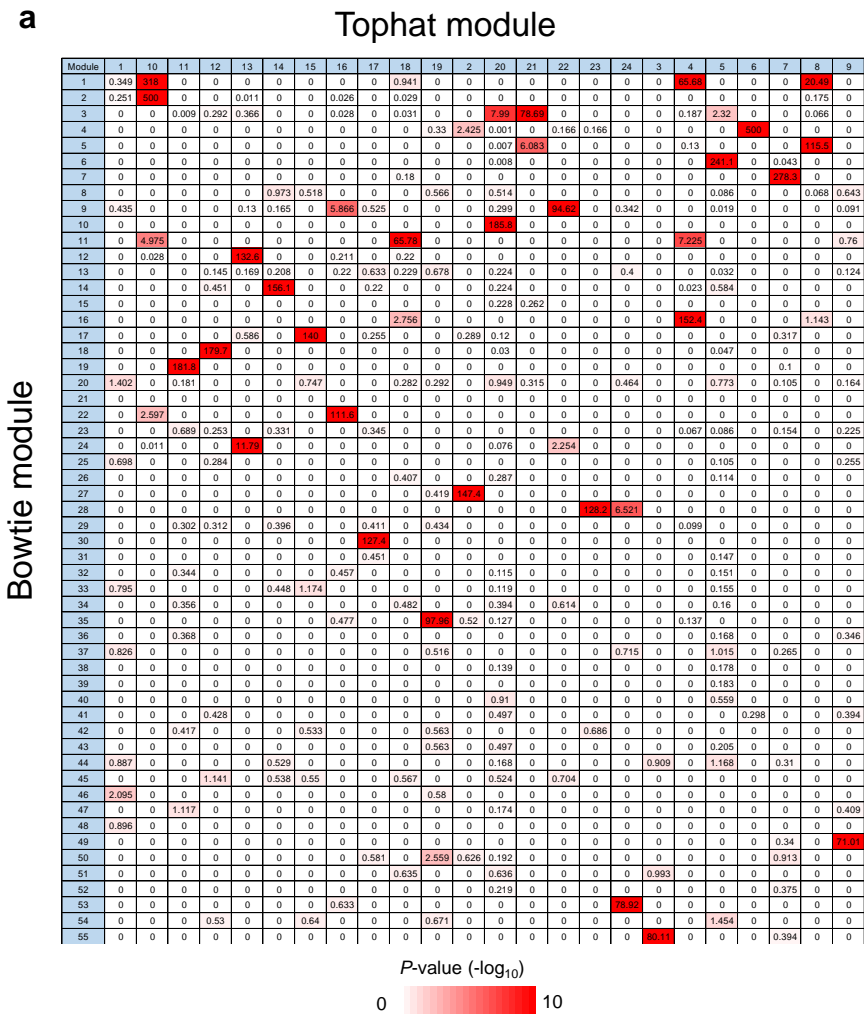
Supplementary Figure 5



Supplementary Figure 5. Variability of gene expression in cardiomyocytes.

(a) Scatter plots showing the variability of gene expression in cardiomyocytes from mice at each time point.
 (b) Scatter plots showing the single-cell expression values of *Nppa* and *Tnnt2* (left) and *Myh7* and *Atp2a2* (right) in cardiomyocytes from mice at 8 weeks after pressure overload.
 (c) Bar plots showing the distribution of cells corresponding to the single-cell expression level (RPKM) detected by single-cardiomyocyte RNA-seq at 8 weeks after pressure overload. Density plots are also shown with blue curves.
 (d) Representative images of smFISH of the heart from mice at 8 weeks after pressure overload. Wheat germ agglutinin (WGA) and DAPI are used as markers of the plasma membrane and nucleus, respectively. Scale bar, 20 μ m.
 (e) Bar plots showing the distribution of cells corresponding to the single-cell fluorescent intensity detected by smFISH of the heart at 8 weeks after pressure overload. Density plots are also shown with blue curves.

Supplementary Figure 6



(legend on next page)

Supplementary Figure 6. Co-expression module-based clustering.

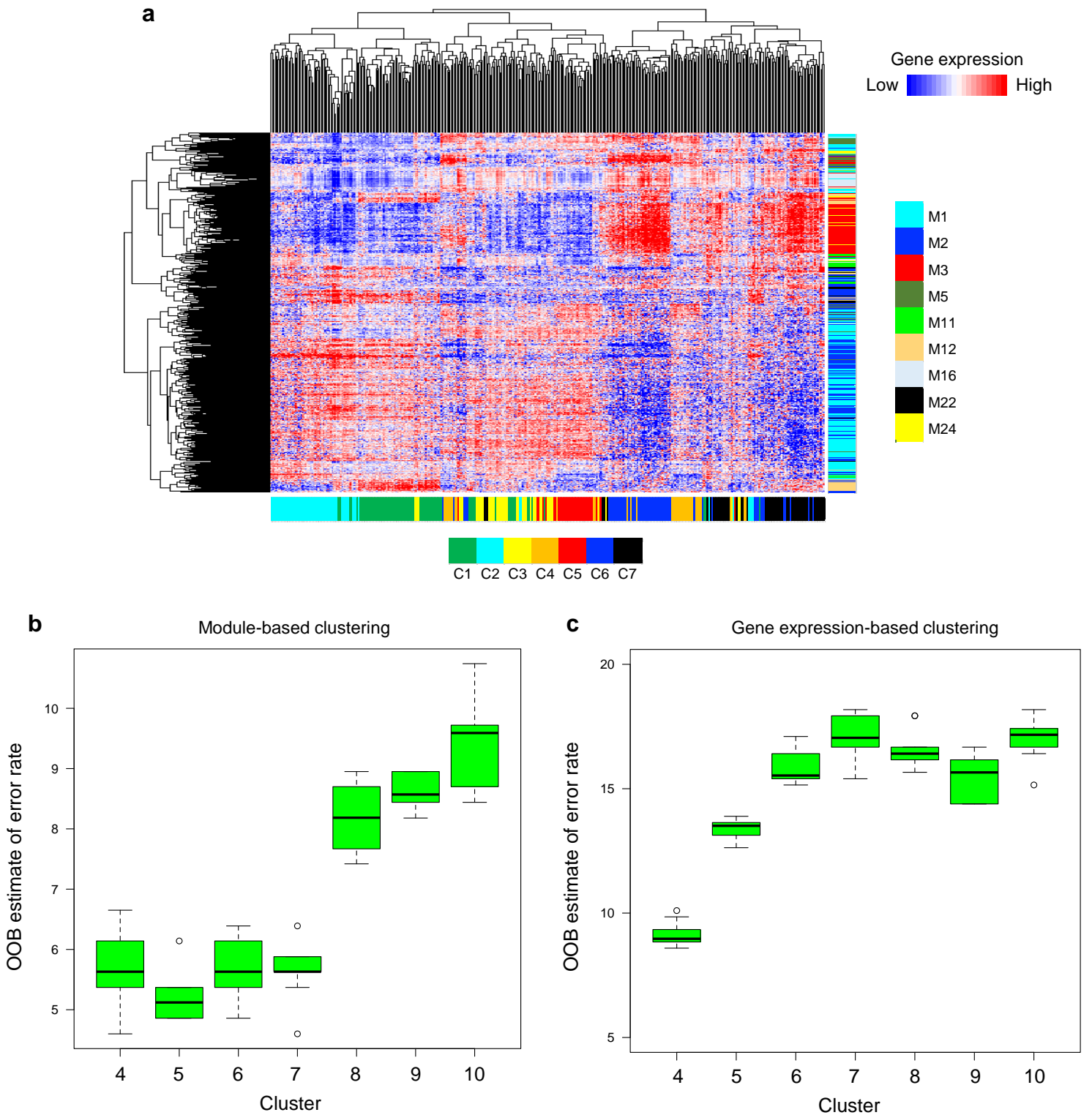
(a) Cross-tabulation table of the modules obtained using expression values from Bowtie (columns) and Tophat (rows). The table is colored by $-\log_{10}(P\text{-value})$, obtained with Fisher's exact test, according to the color legend below the table. The selected 9 modules were recovered in the gene modules obtained using Tophat.

(b) Hierarchical clustering of the expression of all 55 modules detected by weighted gene co-expression network analysis (WGCNA). The colored bar below the heatmap indicates the time when the cardiomyocytes were extracted.

(c) Representative estimated error rates of classifications by all 55 modules (top) and 9 selected modules (bottom).

(d) Mean decrease accuracy and mean decrease Gini (sorted decreasingly from top to bottom) of attributes as assigned by the Random Forest classifier. The red lines indicate the significant thresholds used in this study.

Supplementary Figure 7

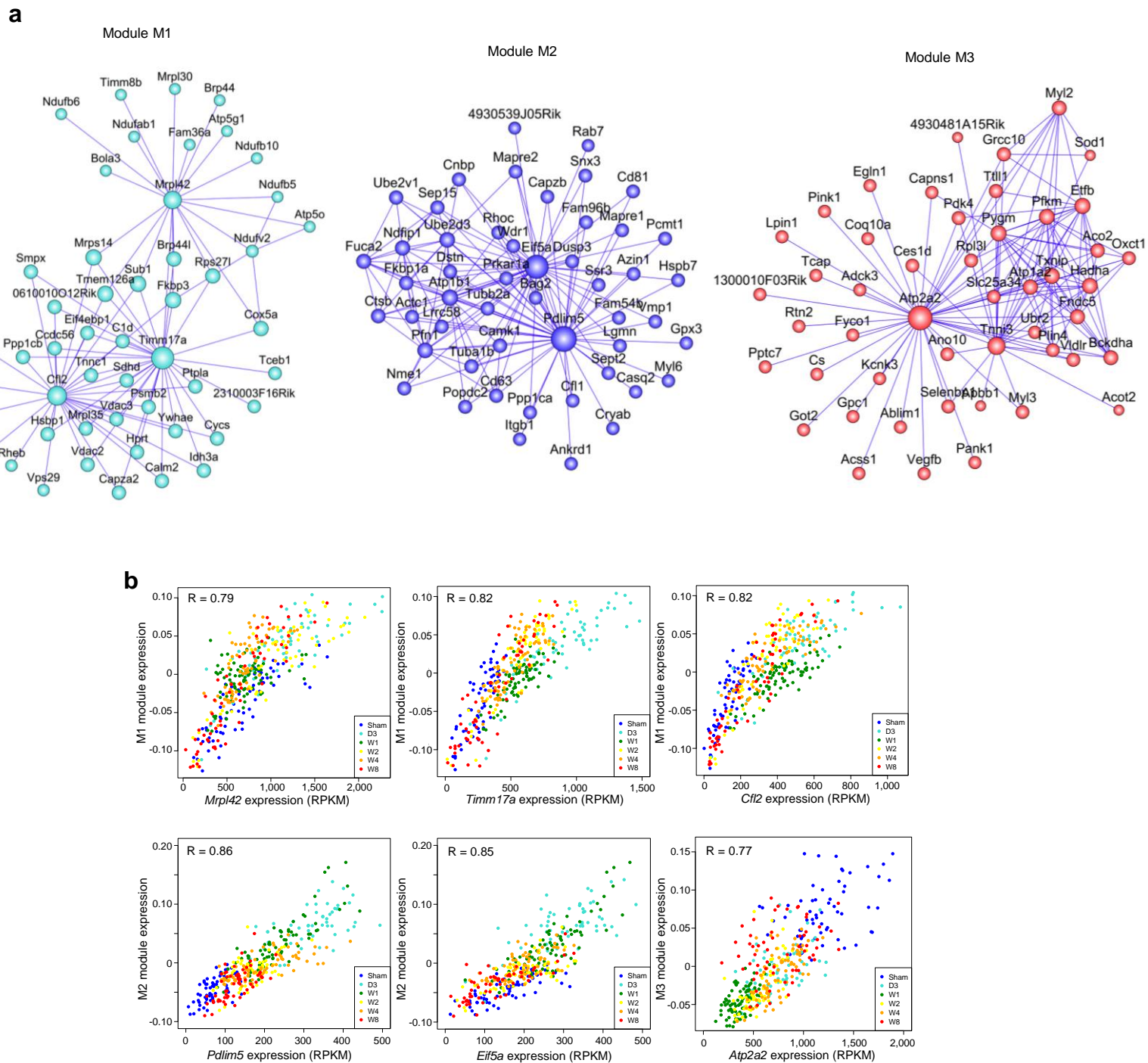


Supplementary Figure 7. Assessment of classification accuracy.

(a) Hierarchical clustering using the expression levels of the genes assigned to the 9 modules (M1, M2, M3, M5, M11, M12, M16, M22, and M24).

(b,c) Boxplot showing changes in the out-of-bag (OOB) error rates at the indicated numbers of clusters in **Figure 1d** (module-based clustering, **b**) and in **Supplementary Figure 7a** (gene expression-based clustering, **c**). Assessment of classification accuracy using Random Forests was performed 10 times. Boxplots show medians (central bars), the 1st to 3rd quartile range (boxes), other data out to 1.5 the interquartile range (whiskers), and outliers (dots).

Supplementary Figure 8



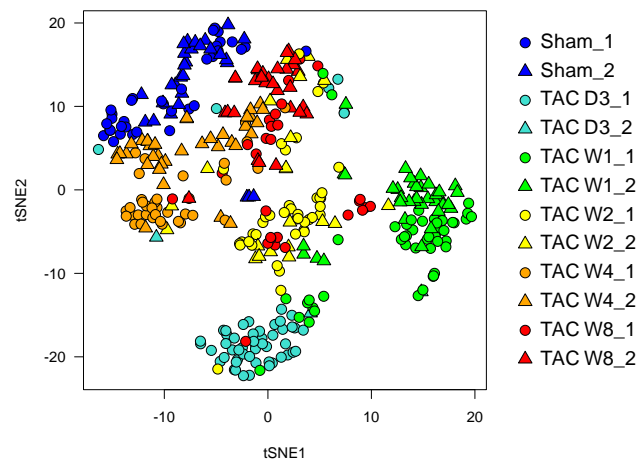
Supplementary Figure 8. Hub gene network analysis.

(a) Hub gene networks of M1, M2, and M3. The size of the dots represents node centrality.

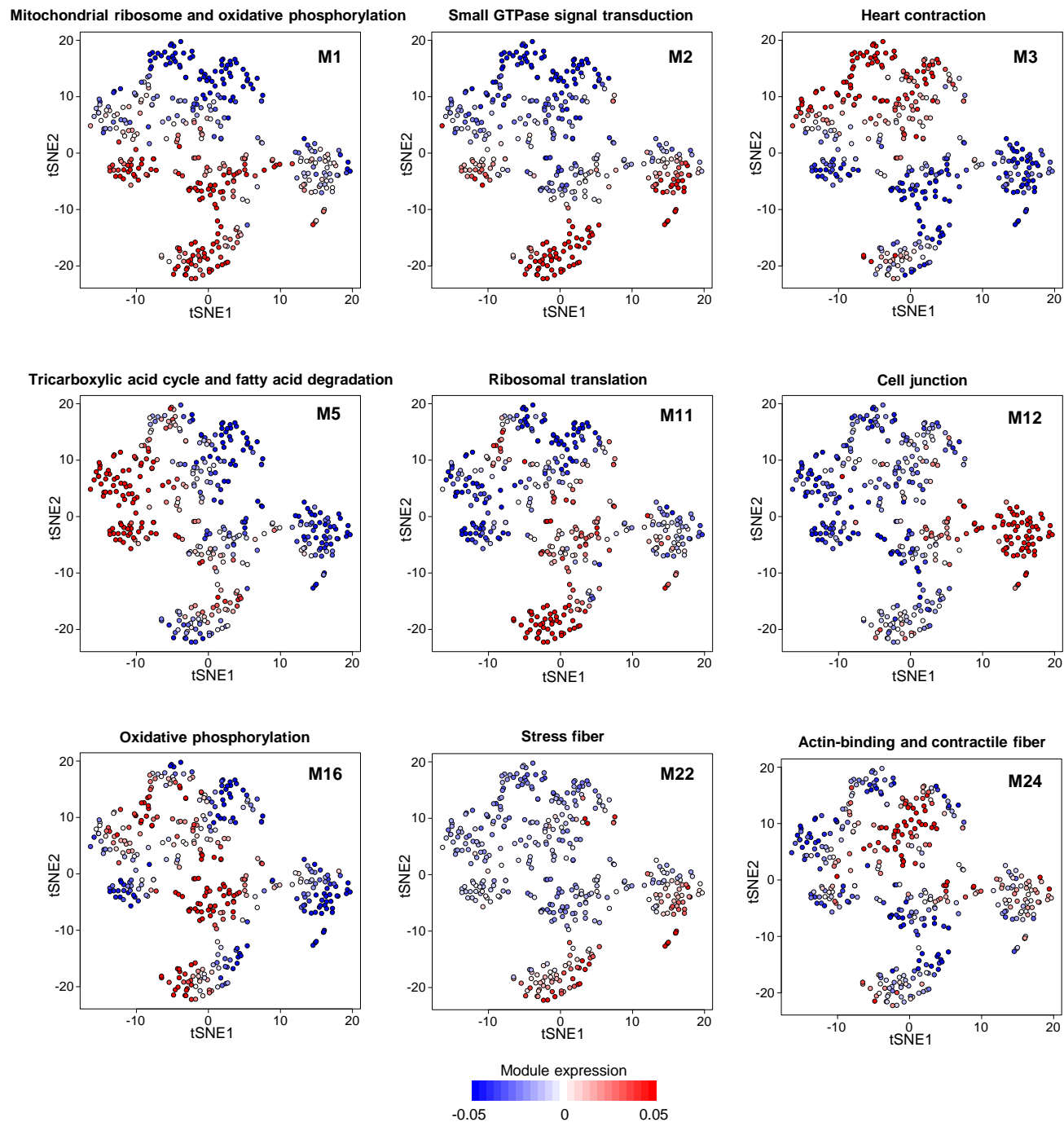
(b) Scatter plot showing the correlations between module expression and the expression of hub genes. Samples are colored according to when the cardiomyocytes were extracted.

Supplementary Figure 9

a



b

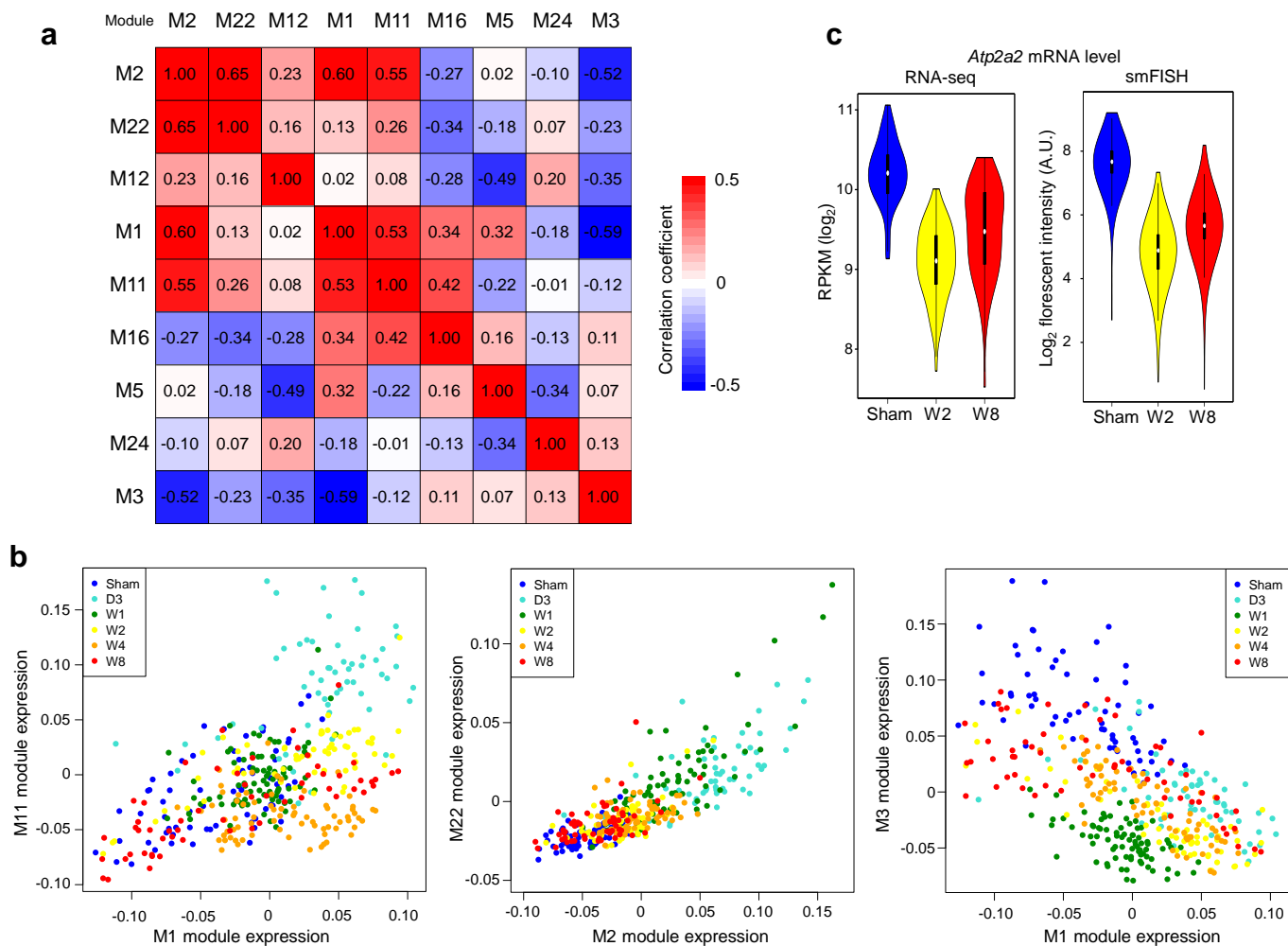


Supplementary Figure 9. Module expression on a t-SNE map.

(a) *t*-distributed stochastic neighbor embedding (t-SNE) plot (Figure 1g) colored by batch.

(b) t-SNE plots colored by the expression levels of each module.

Supplementary Figure 10



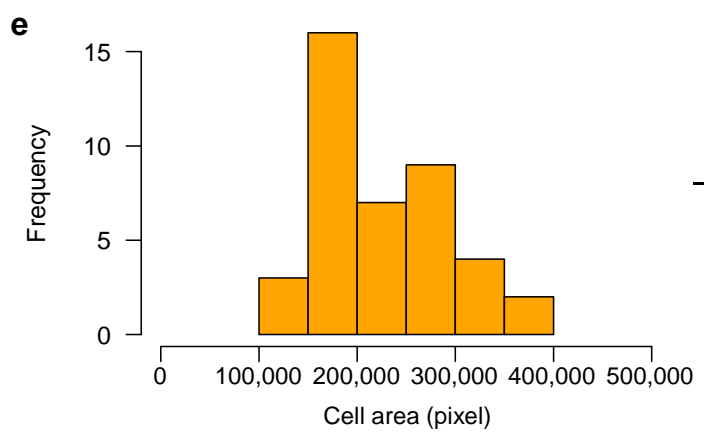
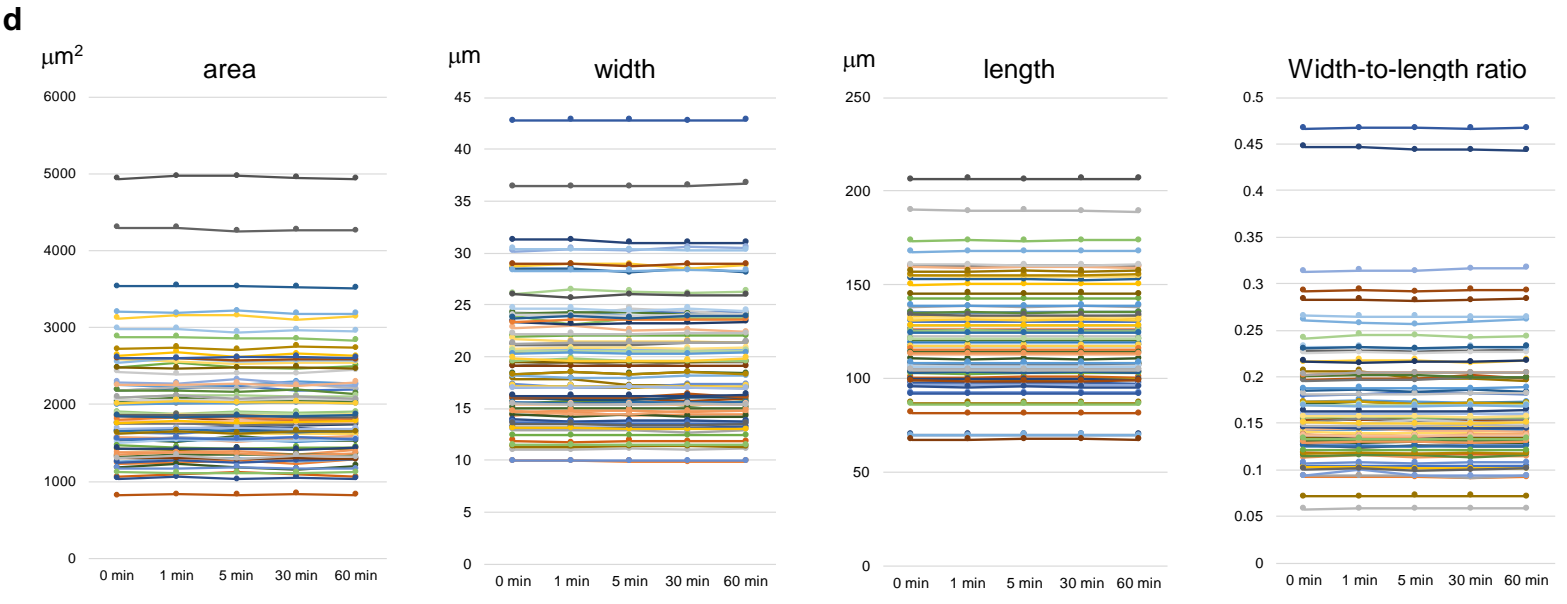
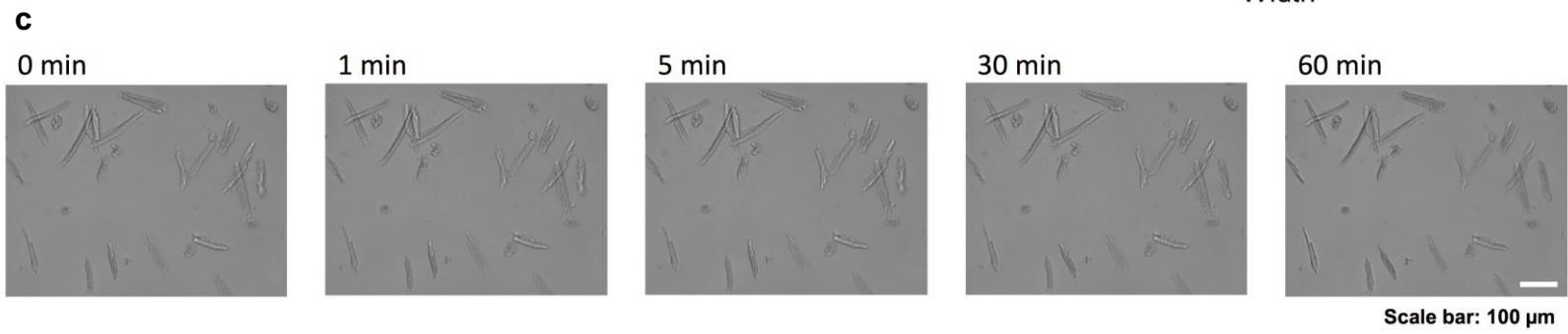
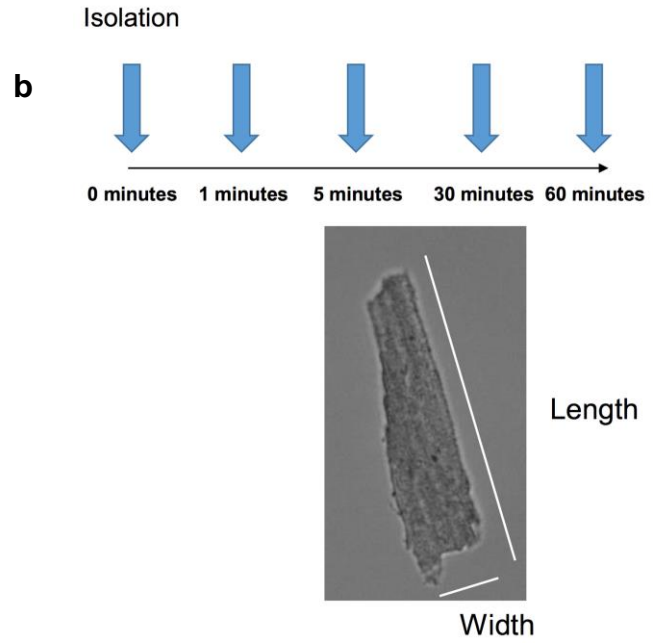
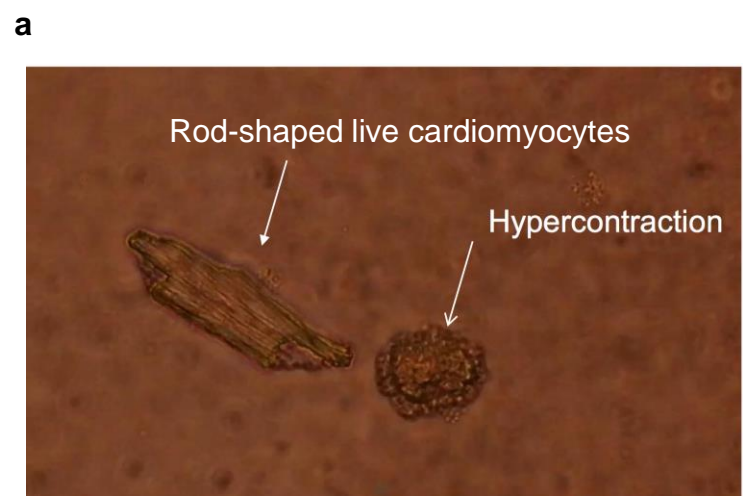
Supplementary Figure 10. Module-to-module correlation analysis.

(a) Pearson correlation matrix showing the correlations between modules.

(b) Scatter plots showing the correlations between modules.

(c) Violin plot showing the dynamics of single-cell *Atp2a2* mRNA expression levels in cardiomyocytes detected by single-cell RNA-seq (left) and smFISH (right).

Supplementary Figure 11



f

Normalized by an endogenous control

GO term	Count	P-value
Mitochondrion (GO:0005739)	72	3.50E-20
Mitochondrial inner membrane (GO:0005743)	22	2.00E-08
Cell redox redox homeostasis (GO:0045454)	9	1.60E-06
Mitochondrial ribosome (GO:0005761)	5	2.50E-04

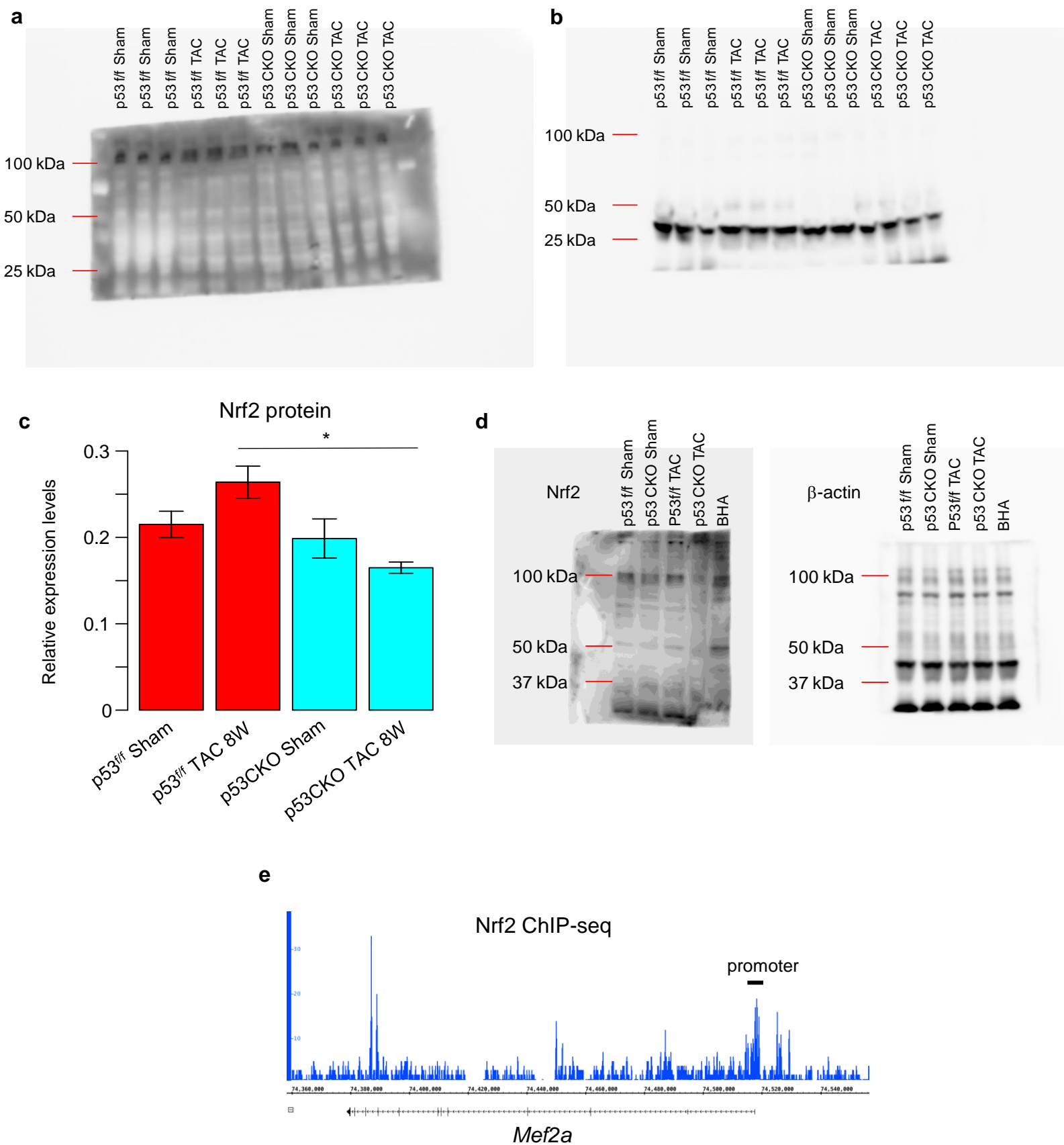
Supplementary Figure 11. Preliminary morphological assessment of isolated cardiomyocytes

- (a) Morphological differences between rod-shaped live cardiomyocytes and cardiomyocytes causing hypercontraction.
- (b) Overview of the assessment of the morphological dynamics of cardiomyocytes after isolation.
- (c) Representative time-lapse images of cardiomyocytes.
- (d) Line graph showing the dynamics of morphological parameters (cell area, width, length, and width-to-length ratio) after isolation.
- (e) Histogram of the area of cells used in **Fig. 3**. Morphological parameters were maintained after isolation.
- (f) List of the most enriched GO terms in the top 300 correlated genes with cell area. Gene expression values were normalized by an endogenous control (*Tnnt2*).

Supplementary Figure 13. Identification and assessment of hypertrophy stage-specific modules.

- (a) Cross-tabulation table of modules obtained using all cardiomyocytes (columns) and modules obtained using cardiomyocytes without the hypertrophy stage (at 1 and 2 weeks after pressure overload) (rows). The table is colored by $-\log_{10}(P\text{-value})$, obtained with Fisher's exact test, according to the color legend below the table. Modules colored yellow are hypertrophic stage-specific (all module enrichment P -values were $>1e-5$).
- (b) Assessment of p21 immunostaining in cardiomyocytes. p21 protein was observed in a small fraction of cardiomyocytes of wild-type mice at 2 weeks after TAC, but not in cardiomyocytes of p21 knockout mice. WGA and DAPI are used as markers of the plasma membrane and nucleus, respectively. Scale bar, 50 μm . The arrows indicate the nuclei of p21-positive cardiomyocytes.
- (c) Representative higher magnification image of p21-positive cardiomyocytes (area of dashed quadrangle in **b**). Scale bar, 20 μm .
- (d) Gene ontology analysis showing the enrichment of genes involved in the oxidative stress response.
- (e) Correlation coefficient between cell area and the expression of genes involved in the oxidative stress response.
- (f) Single-cell quantitative real-time polymerase chain reaction of *Trp53* mRNA (encoding p53) in cardiomyocytes from p53^{flx/flx} (p53^{f/f}), p53CKO, and wild-type mice at 2 weeks after pressure overload.

Supplementary Figure 14



Supplementary Figure 14. p53-Nrf2-Mef2a axis in failing cardiomyocytes.

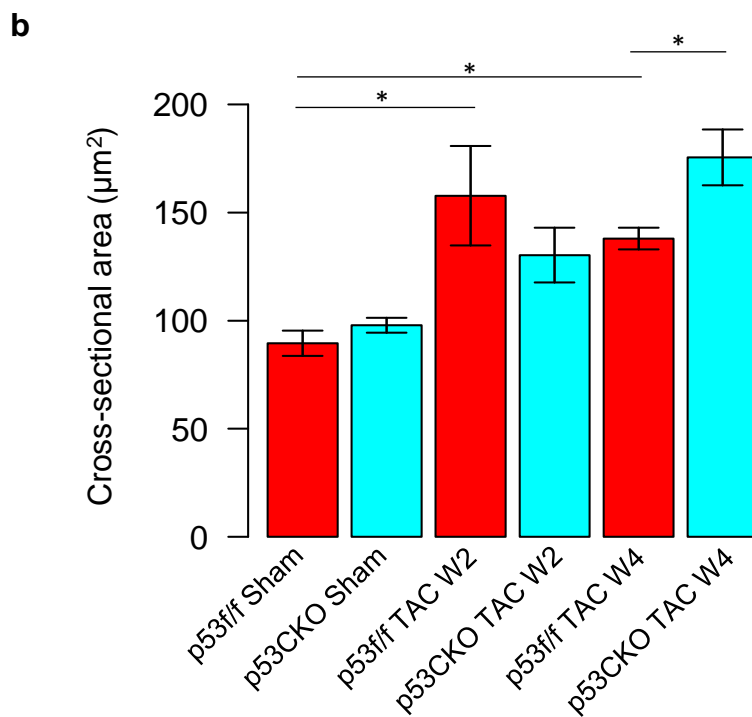
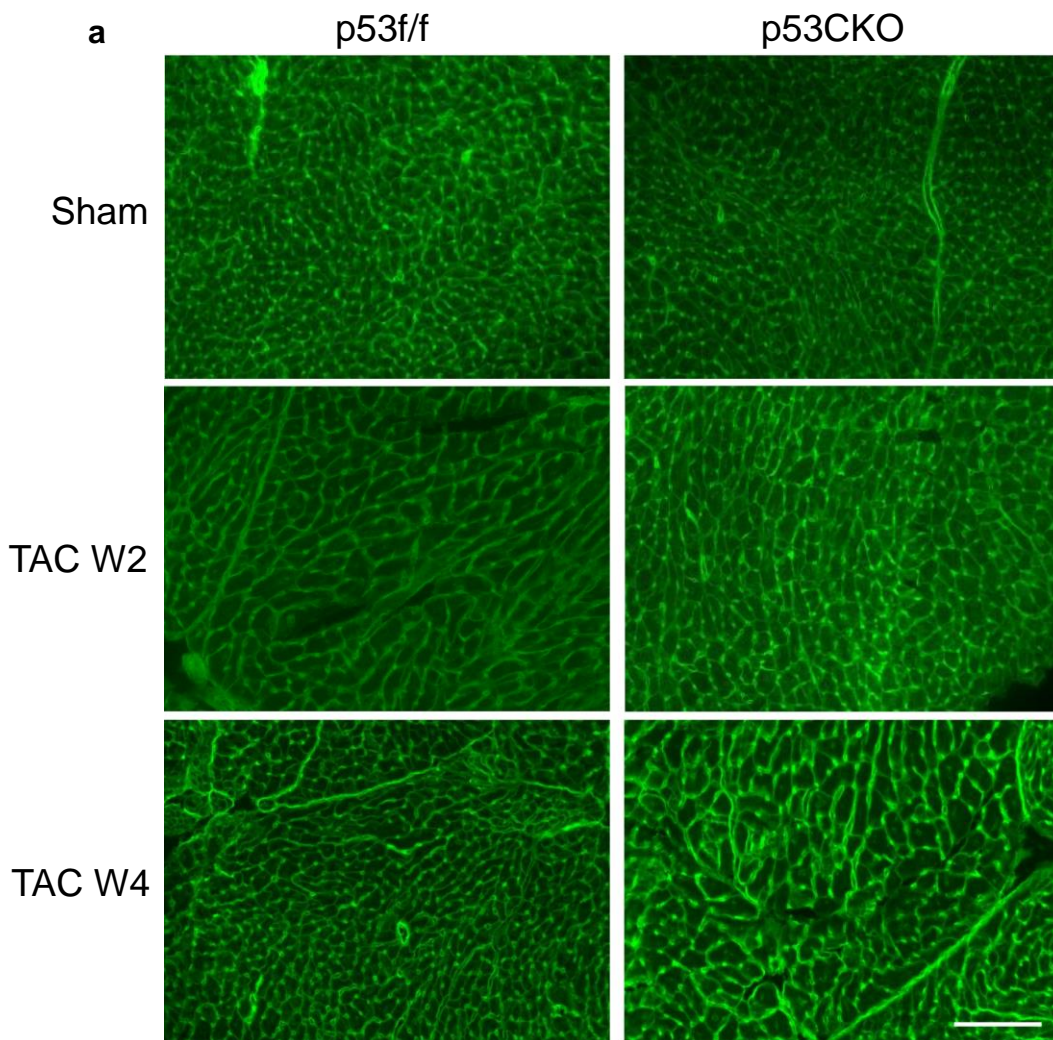
(a,b) Western blot analysis of heart tissues using antibodies against Nrf2 (a) and β -actin (b) ($n = 3$ each).

(c) Densitometric quantification of Nrf2 protein expression levels normalized by β -actin expression levels. Data are represented as mean and standard error of the mean ($n = 3$ each). Asterisks indicate statistical significance ($P < 0.01$, two-tailed unpaired t -test).

(d) Uncropped images of the blots for Fig. 6e.

(e) Public Nrf2 ChIP-seq data of macrophages (GSE72964) showing Nrf2 binding to the *Mef2a* promoter.

Supplementary Figure 15

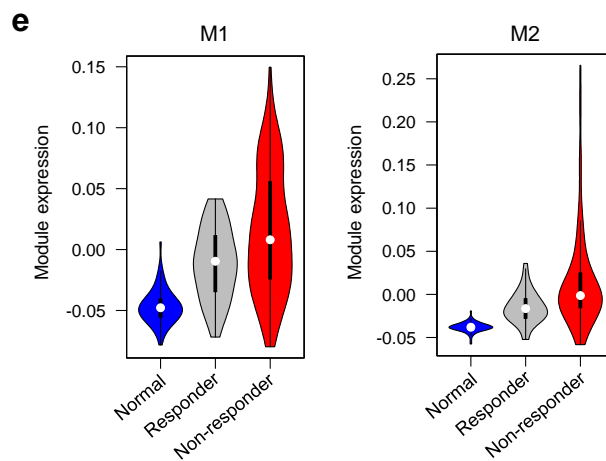
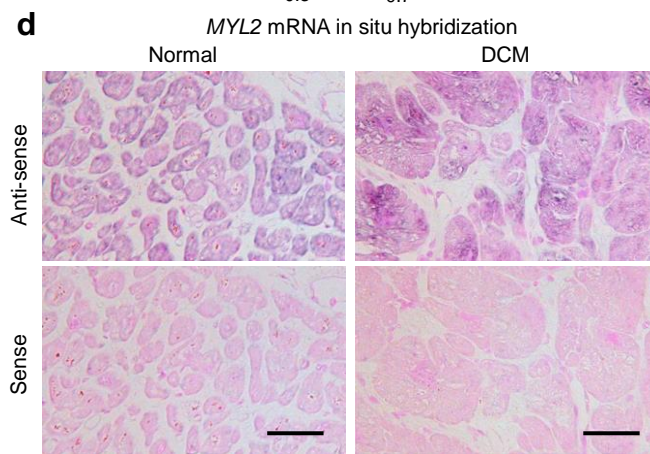
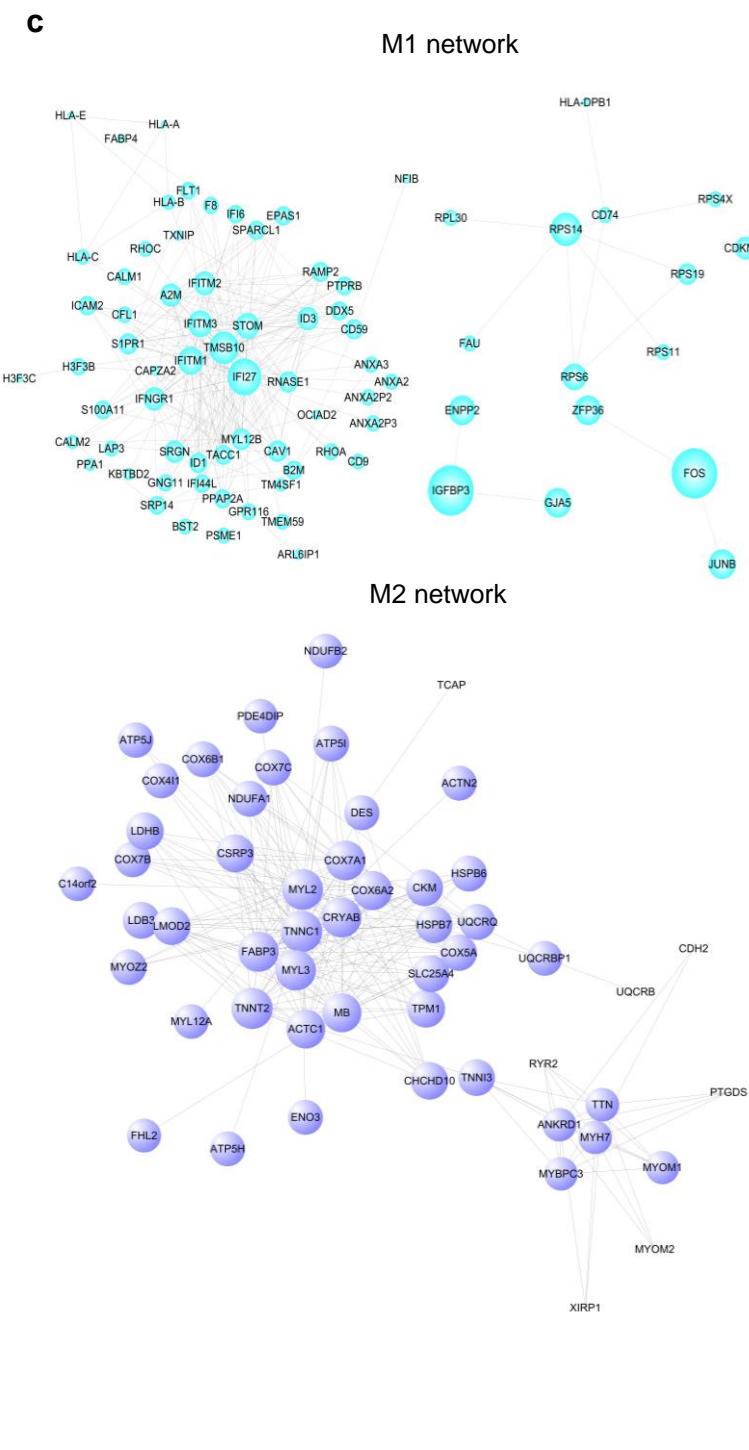
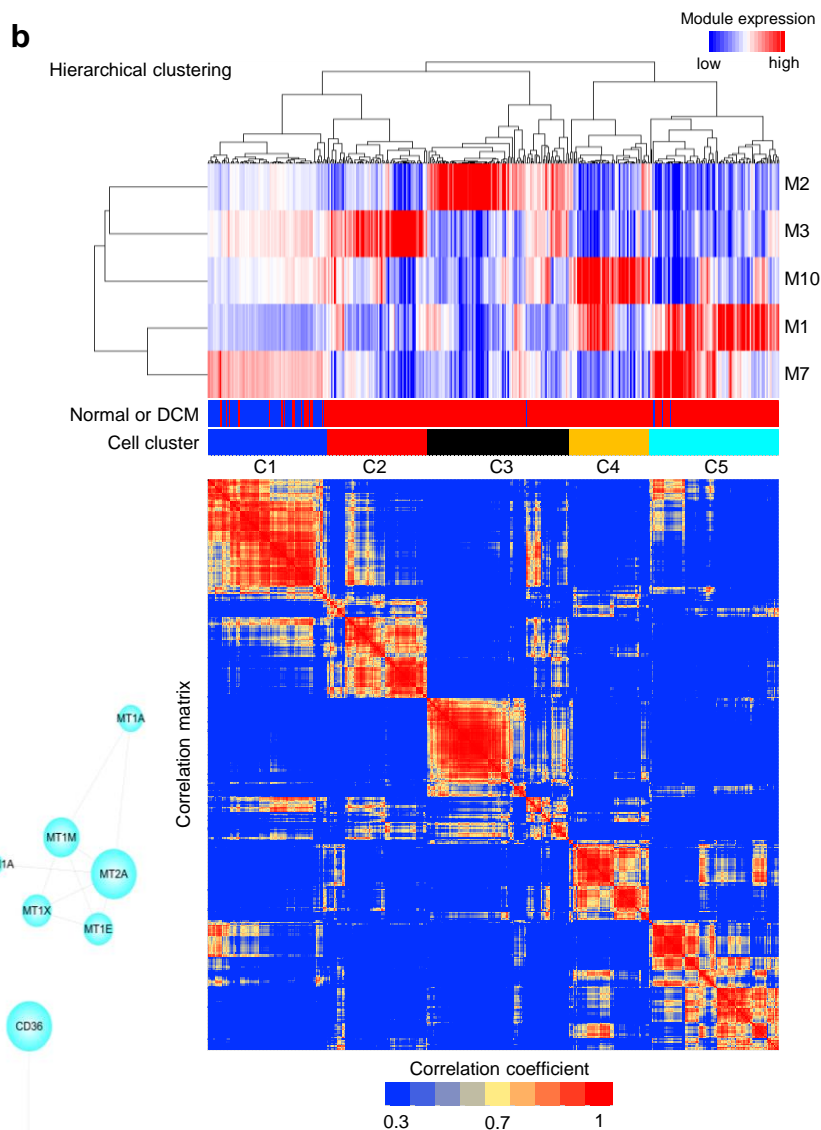
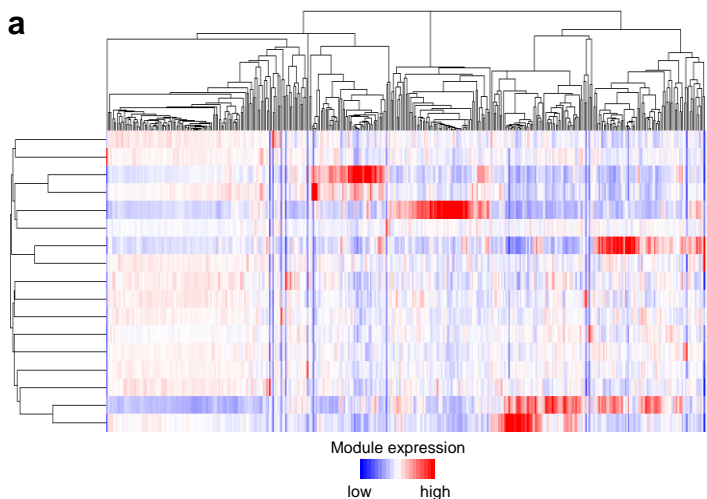


Supplementary Figure 15. Measurement of cardiomyocyte cross-sectional area.

(a) Representative images of wheat germ agglutinin-stained cardiac sections.

(b) Bar plots showing the quantitative measurement of cardiomyocyte cross-sectional area. Data are represented as the mean and standard error of the mean ($n = 6$). Asterisks indicate statistical significance ($P < 0.05$, two-tailed unpaired t -test).

Supplementary Figure 16



Supplementary Figure 16. Distinct gene programs and their pathogenicity in human cardiomyocytes.

(a) Hierarchical clustering of human cardiomyocytes using the expression of all 17 modules detected by WGCNA.

(b) Unsupervised co-expression module-based clustering classifying all cardiomyocytes ($n = 411$) into 5 cell clusters (C1–C5). The colored bar below the heatmap indicates the origin of each cardiomyocyte (normal subjects [blue] or dilated cardiomyopathy [DCM] patients [red]) and the cell clusters. The correlation matrix is also indicated.

(c) The co-expression networks of M1 and M2. The size of the dots represents node centrality.

(d) Representative images of *MYL2* mRNA *in situ* hybridization analysis of the heart from normal subjects and DCM patients. The anti-sense probe specifically hybridized with *MYL2* mRNA, while the sense probe was used as a negative control. Scale bar, 50 μm .

(e) Violin plot showing the distribution of module expression.