Cell Systems, Volume *12*

Supplemental information

Cell states beyond transcriptomics: Integrating structural organization and gene expression in hiPSC-derived cardiomyocytes

Kaytlyn A. Gerbin, Tanya Grancharova, Rory M. Donovan-Maiye, Melissa C. Hendershott, Helen G. Anderson, Jackson M. Brown, Jianxu Chen, Stephanie Q. Dinh, Jamie L. Gehring, Gregory R. Johnson, HyeonWoo Lee, Aditya Nath, Angelique M. Nelson, M. Filip Sluzewski, Matheus P. Viana, Calysta Yan, Rebecca J. Zaunbrecher, Kimberly R. Cordes Metzler, Nathalie Gaudreault, Theo A. Knijnenburg, Susanne M. Rafelski, Julie A. Theriot, and Ruwanthi N. Gunawardane

Supplemental Information Inventory

Cell states beyond transcriptomics: integrating structural organization and gene expression in hiPSC-derived cardiomyocytes

This manuscript contains the following supplemental materials:

Supplemental figures and legends:

Figure S1, related to Figure 1: hiPSC-derived cardiomyocytes as a model system for studying the relationship between transcript abundance, protein abundance, and cellular organization

Figure S2, related to Figure 2: Classification of local alpha-actinin-2 patterns using a machine-learning-based model

Figure S3, related to Figure 3: Global alignment of alpha-actinin-2 organization quantified using Haralick correlation

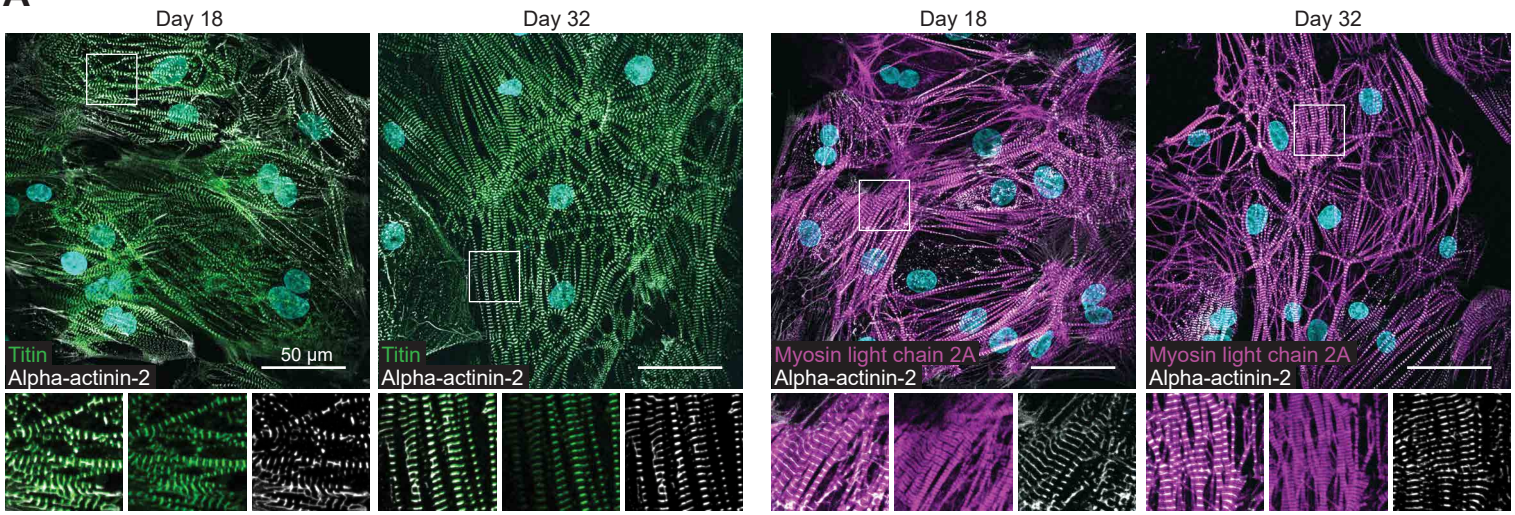
Figure S4, related to Figure 5: A single metric combining multiple cell features can quantify cardiomyocyte organization at scale across datasets and timepoints

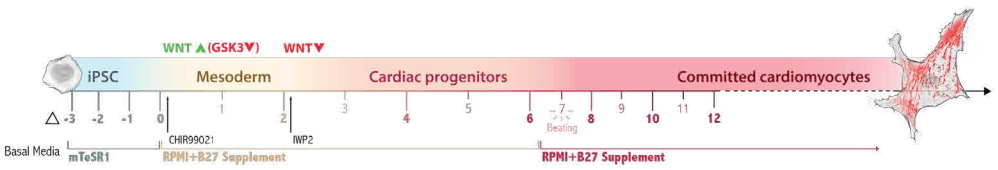
Figure S5, related to Figure 6: Measurement of correlation between cell organization and transcript abundance allows for identification of genes that correlate with cell organization, timepoint, or both

Figure S6, related to Figure 6: Simultaneous measurement of transcript abundance, protein abundance, and organization of alpha-actinin-2 provides experimental context for correlations

Figure S7, related to Figure 4: Incorporating *MYH6* and *MYH7* transcript abundance as features for predicting expert scores has a minimal impact on model performance

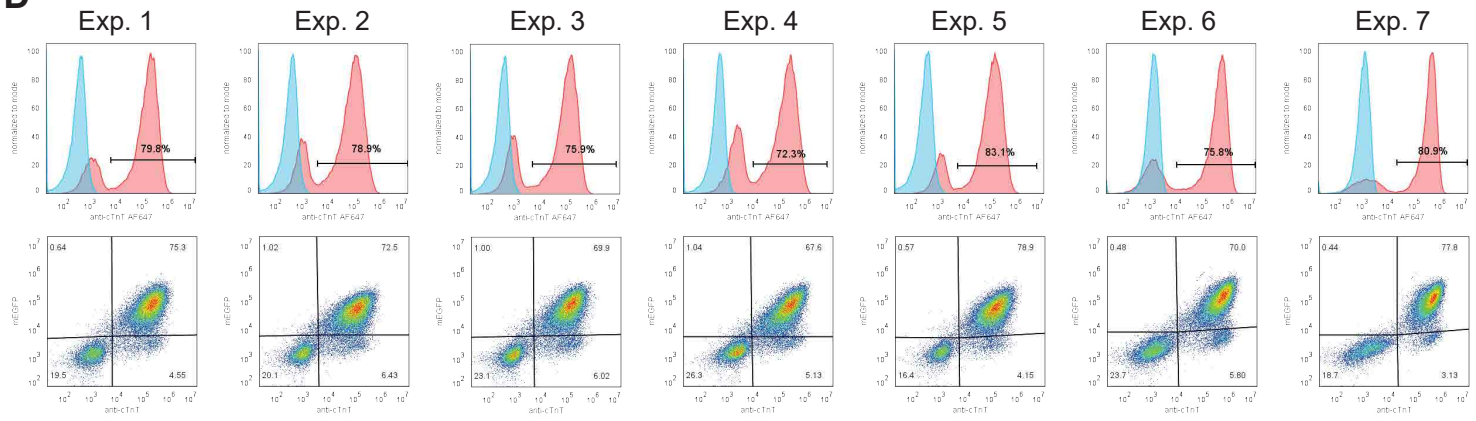
Figure S1

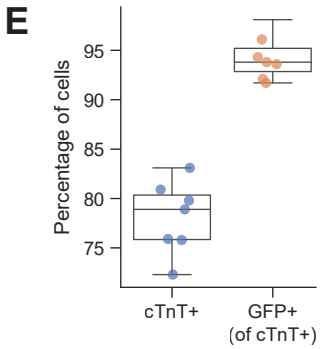


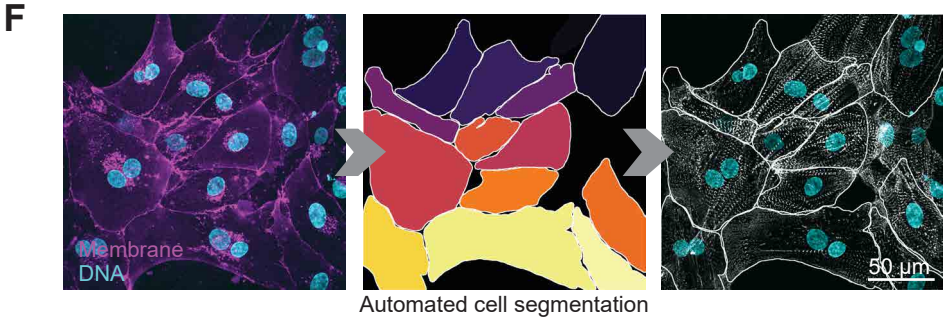


C

Exp.	Live timepoints	Fixed timepoints
1	D18, D25,D32	D18, D25, D32
2	D18, D25, D32	None
3	D18, D25, D32	D18, D32
4	D18, D25, D32	None
5	D18, D25	None
6	None	D25
7	None	D25







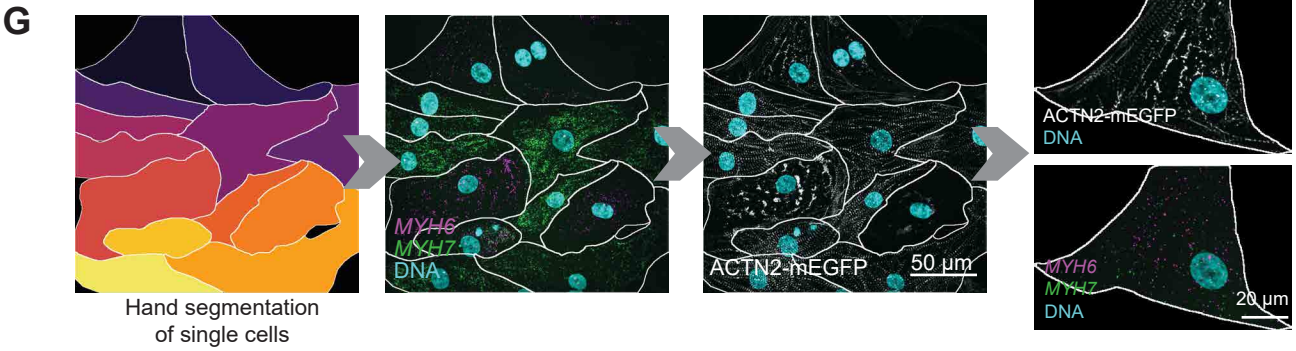


Figure S1, related to Figure 1: hiPSC-derived cardiomyocytes as a model system for studying the relationship between transcript abundance, protein abundance, and cellular organization

A. Example images of immunofluorescence staining in hiPSC-derived cardiomyocytes, endogenously tagged with alpha-actinin-2-mEGFP (white) and stained for titin (N-terminal, encoded by *TTN* gene, left images) and myosin light chain 2a (MLC2a, encoded by *MYL2* gene, right images). In all images green = titin, magenta = MLC2a, white = alpha-actinin-2-mEGFP, cyan = DNA. White boxes indicate areas enlarged for detail in lower images. Scale bar = 50 μm, insets are 30 μm x 30 μm.
B. Schematic outlining directed differentiation protocol used. Further details of this protocol can be found in the Materials and Methods section.
C. Table outlining which experimental replicates of hiPSC-CMs were used for each live- and fixed-imaging timepoint.
D. Flow cytometry chart for each replate outlined in panel (C). The top row shows histograms of the intensity of staining for cardiac troponin T (cTnT), with isotype controls shown as blue curves and stained samples shown as red curves. Numbers indicate % cTnT+ cells as gated against isotype controls. The bottom row shows intensity of cTnT versus intensity of mEGFP for stained samples.
E. Box and whisker plots of % cTnT+ cells (left plot) and % GFP+ cells of cTnT+ cells (right plot) from re-plates outlined in panel (C).Outliers are defined as outside 1.5 times the interquartile range (IQR).
F. Workflow outlining live imaging of hiPSC-CMs. Cells are stained with a wheat germ agglutinin (WGA) conjugate and NucViolet nuclear stain prior to imaging (see Materials and Methods for detailed protocol). Cell outlines are used for an ML-based segmentation workflow to obtain single cell outlines. These segmentations are then used to profile alpha-actinin-2-mEGFP organization in single cells across timepoints and datasets. Scale bar = 50 μm.
G. Outline of workflow for obtaining single-cell structure and transcript data. Cells were fixed, probed by RNA FISH, and imaged. Expert annotators performed manual segmentations of all fields of view (FOVs) by referencing all imaged channels including brightfield, alpha-actinin-2-mEGFP, and FISH probes; these segmentations were used to generate single cell data for both alpha-actinin-2-mEGFP organization and probe abundance. Scale bar = 20 μm.

Figure S2

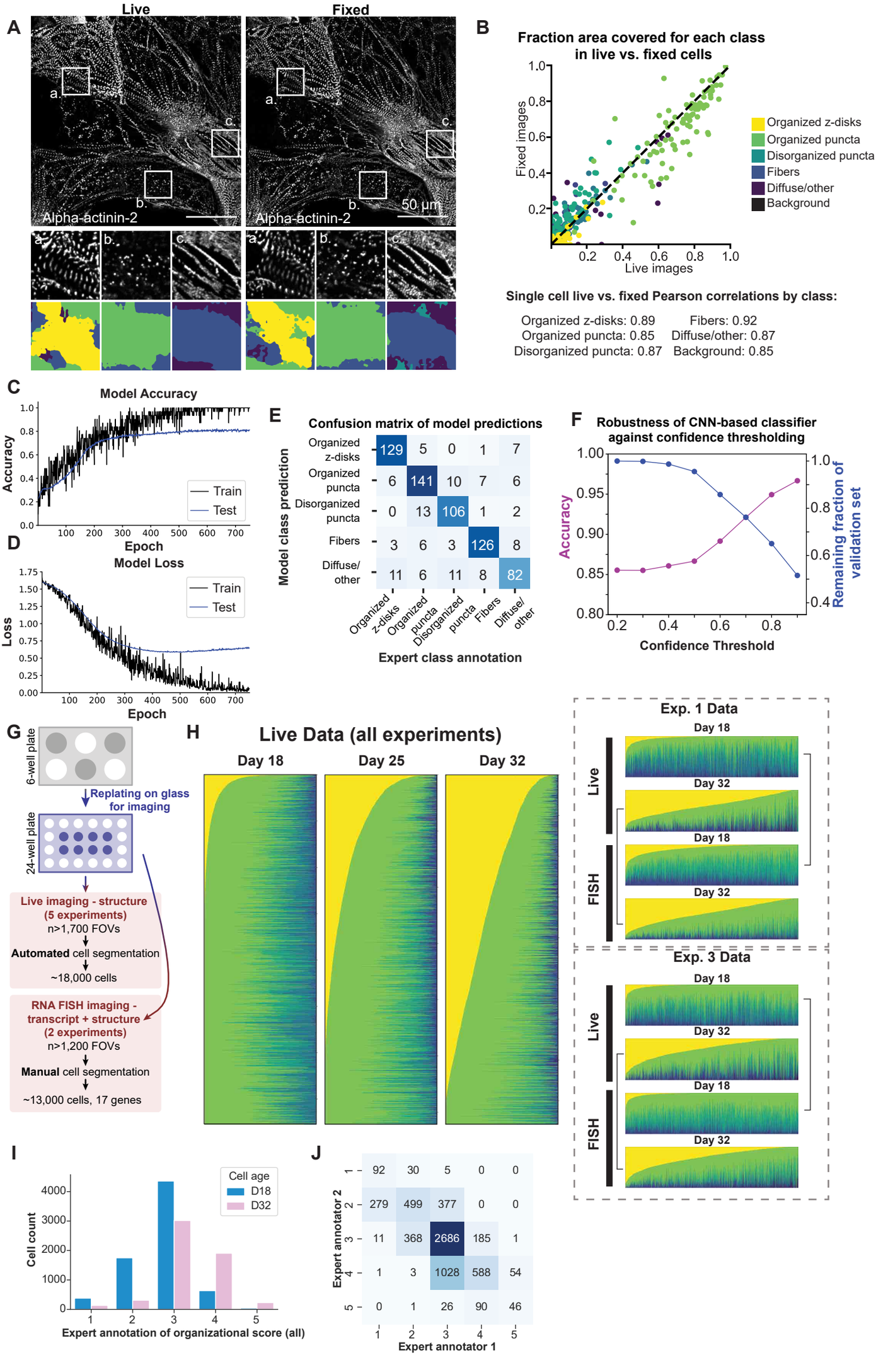
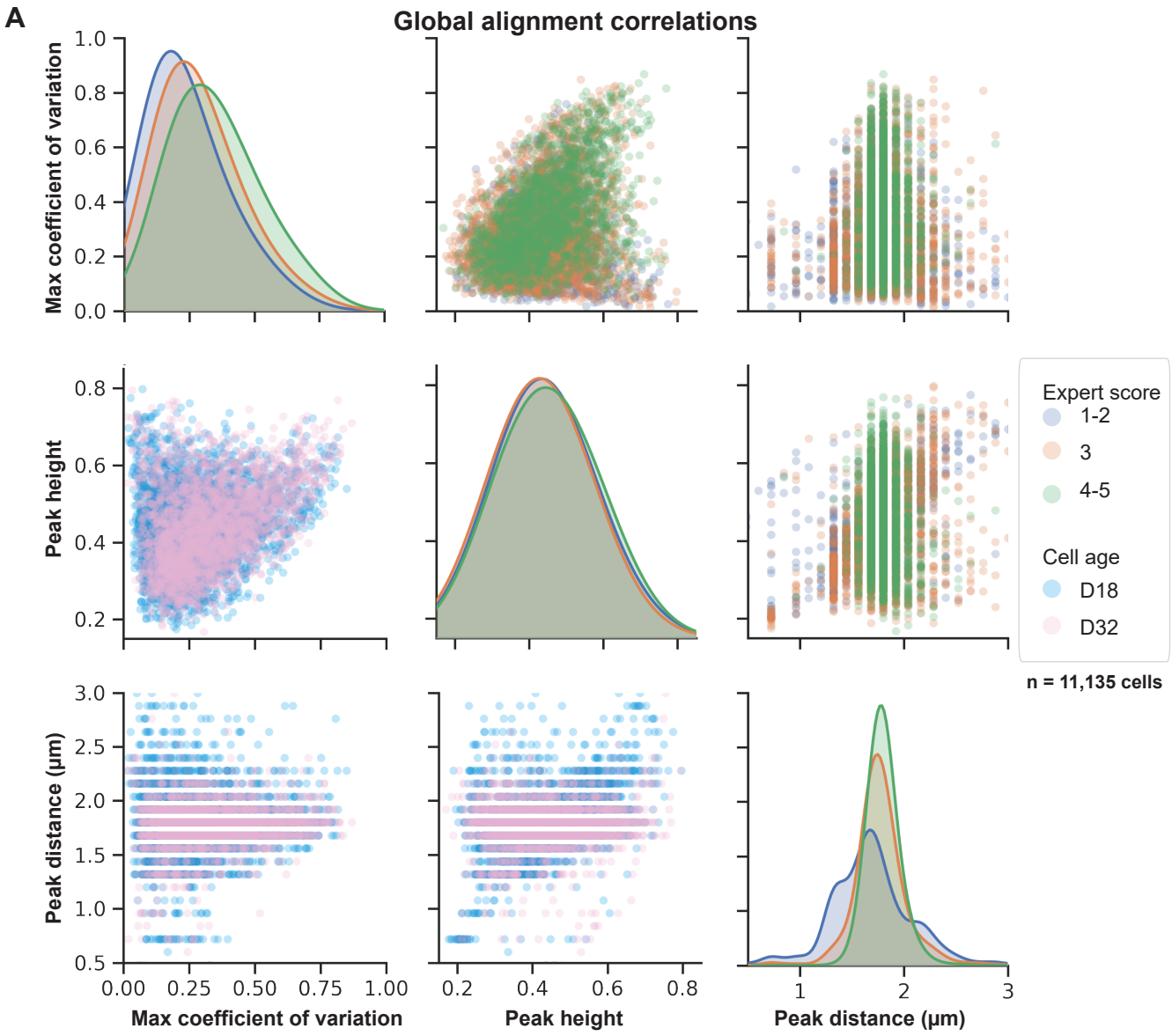

A
Live
Fixed
a.
b.
c.
Alpha-actinin-2
50 μm
B
Fraction area covered for each class in live vs. fixed cells
Fixed images
Live images
Organized z-disks
Organized puncta
Disorganized puncta
Fibers
Diffuse/other
Background
Single cell live vs. fixed Pearson correlations by class:
Organized z-disks: 0.89
Organized puncta: 0.85
Disorganized puncta: 0.87
Fibers: 0.92
Diffuse/other: 0.87
Background: 0.85
C
Model Accuracy
Accuracy
Epoch
Train
Test
D
Model Loss
Loss
E
Confusion matrix of model predictions
Model class prediction
Expert class annotation
F
Robustness of CNN-based classifier against confidence thresholding
Accuracy
Remaining fraction of validation set
Confidence Threshold
G
6-well plate
Replating on glass for imaging
24-well plate
Live imaging - structure (5 experiments)
n>1,700 FOVs
Automated cell segmentation
~18,000 cells
RNA FISH imaging - transcript + structure (2 experiments)
n>1,200 FOVs
Manual cell segmentation
~13,000 cells, 17 genes
H
Live Data (all experiments)
Day 18
Day 25
Day 32
Exp. 1 Data
Exp. 3 Data
Live
FISH
I
Cell age
D18
D32
Cell count
Expert annotation of organizational score (all)
J
Expert annotator 2
Expert annotator 1

Figure S2, related to Figure 2: Classification of local alpha-actinin-2 patterns using a machine-learning-based model

A. Comparison of a live and fixed image showing the classification model to demonstrate that the model trained on data collected with live cardiomyocytes can be applied to fixed samples. Example shown is a field-of view that was imaged live and again after fixation. Insets show features and classifications using the same classification groups and color scheme shown in Figure 2B. Scale bar = 50 µm, image insets are 20 µm x 20 µm.
B. Scatter plot showing fraction of area covered by classes for matched single cells from live (x-axis) versus fixed (y-axis) images. Cells were imaged live and then were fixed; the same positions were imaged again after fixation (as shown in example images in Figure S2A). Classes for single cells were calculated using manually segmented single-cell outlines. Data includes 14 FOVs and 92 segmented cells in live and fixed images. Points are color coded by class using the same classification groups and color scheme shown in Figure 2B.
C. Improvement in deep learning classifier model accuracy over training epochs.
D. Model loss across training epochs. Loss refers to the function that is used to determine how accurate the model currently is and to calculate gradients for optimization. A lower loss score is indicative of a more accurate model.
E. Confusion matrix showing concordance between human class annotations and model predictions for alpha-actinin-2 patterns of local organization.
F. Evaluation of model confidence in predictions. The model outputs a five-element vector; each element has a value between 0 and 1, representing the probability (or amount of "confidence" the model has) that the image patch is of the corresponding class, with all elements summing to 1. The purple line shows the model's accuracy when restricted to a given confidence threshold; the blue line shows the percent of classifications in the validation set which had a confidence above a given value.
G. Schematic illustrating division of samples for live and fixed imaging; from an individual replating experiment, cells were imaged live at D18, D25 and D32, and imaged fixed at corresponding timepoints to capture both structure and transcript.
H. Heatmaps in which individual cells (y-axis) have been ranked according to the fraction of the cell area consisting of organized z-disks (yellow). Colors within the heatmap represent the same organizational classes shown in **Figure 2**. Heatmaps are shown for all live datasets (left) at each timepoint (n = 7,544 for D18, n = 6,214 for D25 and n = 4,287 for D32) and for fixed versus live data from re-plates that were used for measuring both transcript abundance and structural organization (right).
I. Histograms showing expert annotation of organizational score by two independent experts for D18 (blue) and D32 (pink) scored from least organized (1) to most organized (5) as shown in **Figure 2A**. Each cell was scored by both experts; thus, each cell is plotted twice in the histogram. n = 6,370 cells
J. Confusion matrix showing concordance between expert annotators. n = 6,370 cells

Figure S3


A
Global alignment correlations
Max coefficient of variation
Peak height
Peak distance (µm)
Max coefficient of variation
Peak height
Peak distance (µm)
Expert score
1-2
3
4-5
Cell age
D18
D32
n = 11,135 cells

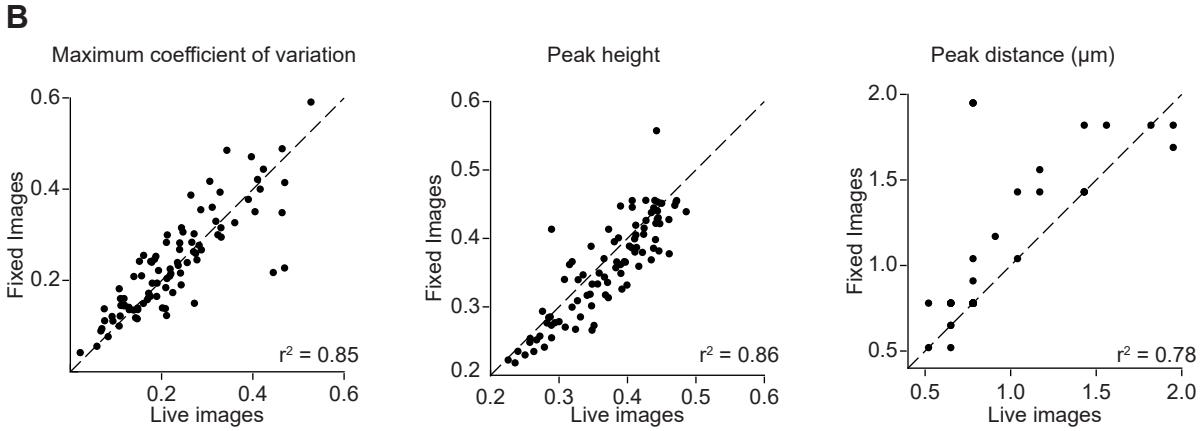
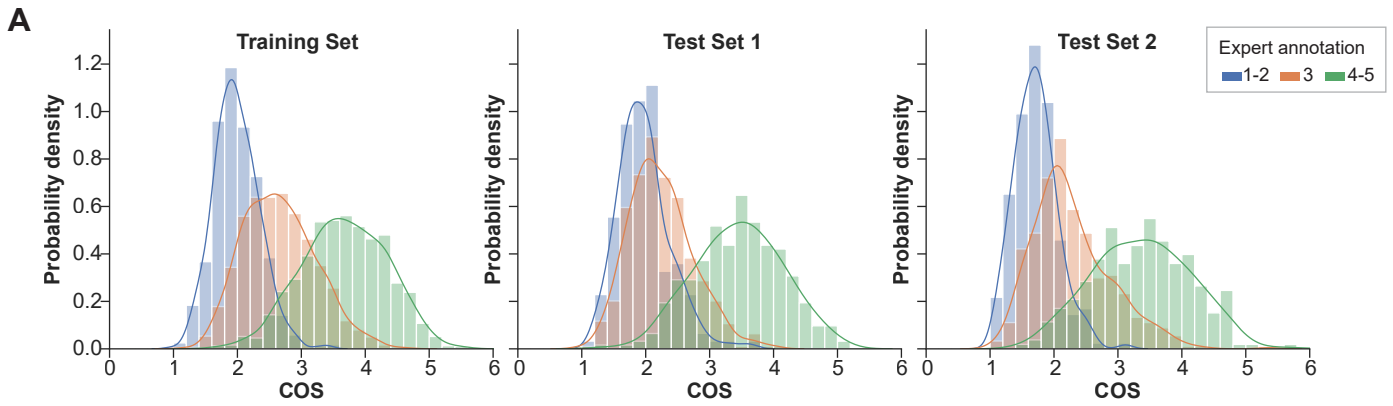
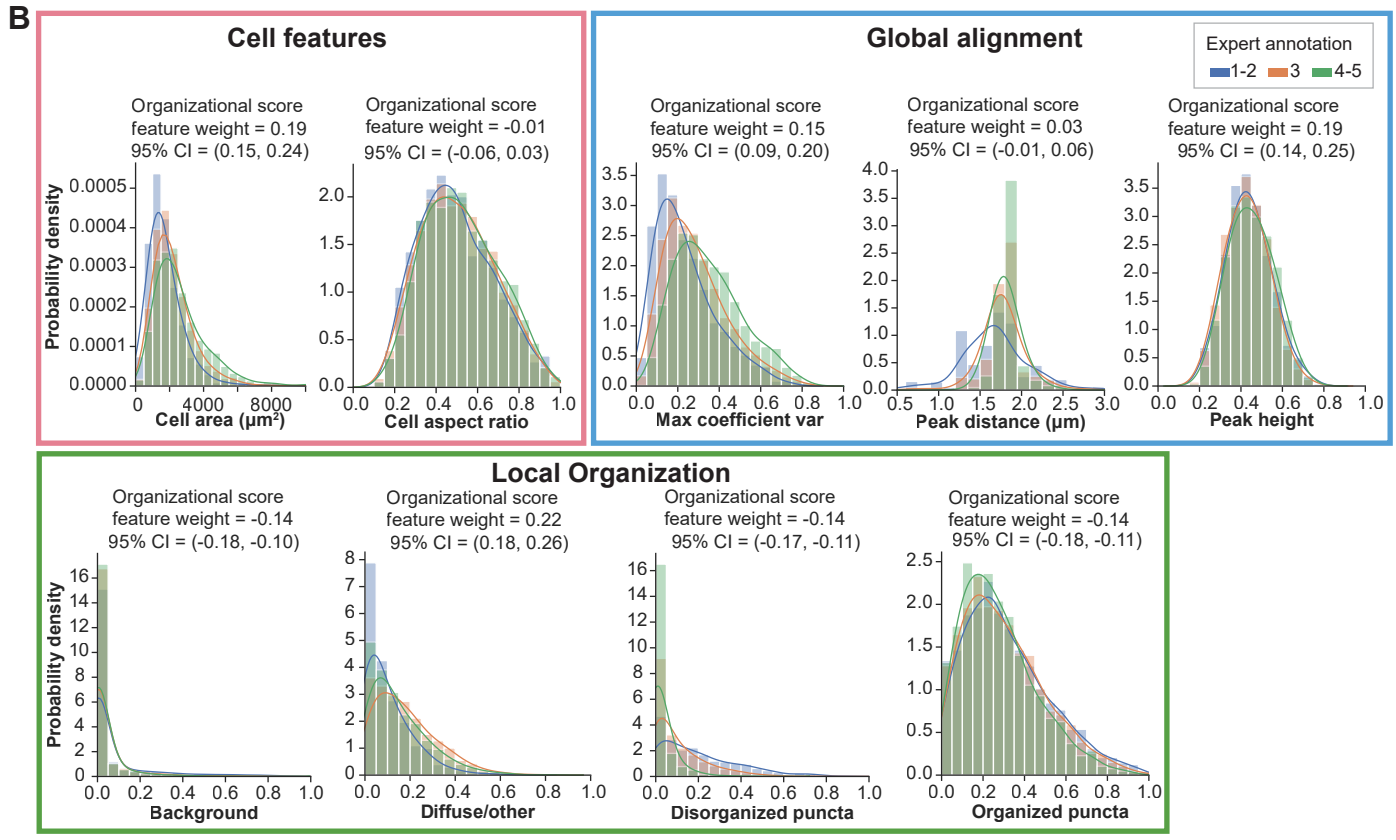

B
Maximum coefficient of variation
Peak height
Peak distance (µm)
Fixed Images
Live images
r² = 0.85
r² = 0.86
r² = 0.78

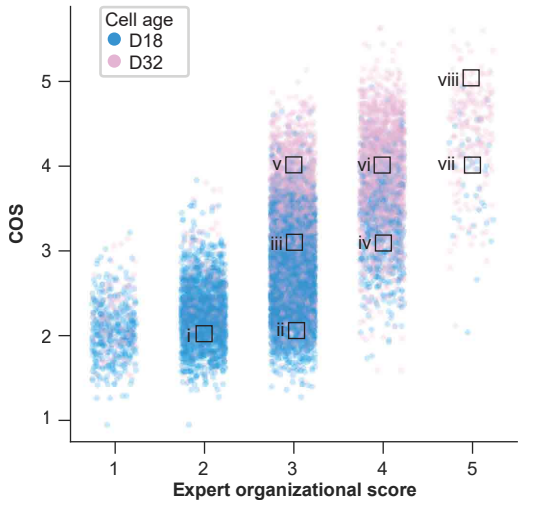
Figure S3, related to Figure 3: Global alignment of alpha-actinin-2 organization quantified using Haralick correlation

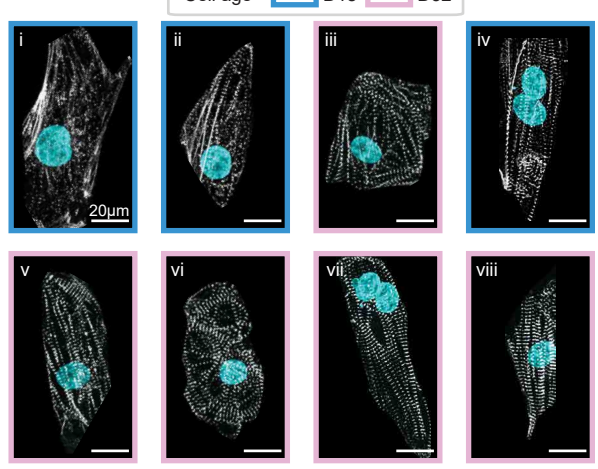
A. Plots showing correlations between features calculated to measure global alignment (Maximum coefficient of variation, Peak height, and Peak distance) and expert annotation score (blue/orange/green) or cell age (pink and blue). n = 11,135 cells.
B. Scatter plots showing alignment metrics for matched single cells from live (x-axis) versus fixed (y-axis) images. Cells were imaged live and fixed; the same positions were imaged again after fixation (as shown in example images in Figure S2A). Metrics for single cells were calculated using manually segmented single-cell outlines. Data includes 14 FOVs and 92 segmented cells in live and fixed images. Plots show correlations for maximum coefficient of variation, peak height, and peak distance.

Figure S4


A
Training Set
Test Set 1
Test Set 2
Probability density
COS
Expert annotation
1-2
3
4-5


B
Cell features
Organizational score
feature weight = 0.19
95% CI = (0.15, 0.24)
Cell area (μm²)
Organizational score
feature weight = -0.01
95% CI = (-0.06, 0.03)
Cell aspect ratio
Global alignment
Expert annotation
1-2
3
4-5
Organizational score
feature weight = 0.15
95% CI = (0.09, 0.20)
Max coefficient var
Organizational score
feature weight = 0.03
95% CI = (-0.01, 0.06)
Peak distance (μm)
Organizational score
feature weight = 0.19
95% CI = (0.14, 0.25)
Peak height
Local Organization
Organizational score
feature weight = -0.14
95% CI = (-0.18, -0.10)
Background
Organizational score
feature weight = 0.22
95% CI = (0.18, 0.26)
Diffuse/other
Organizational score
feature weight = -0.14
95% CI = (-0.17, -0.11)
Disorganized puncta
Organizational score
feature weight = -0.14
95% CI = (-0.18, -0.11)
Organized puncta
Probability density


C
Cell age
D18
D32
i
ii
iii
iv
v
vi
vii
viii
COS
Expert organizational score


D
Cell age
D18
D32
20μm
i
ii
iii
iv
v
vi
vii
viii

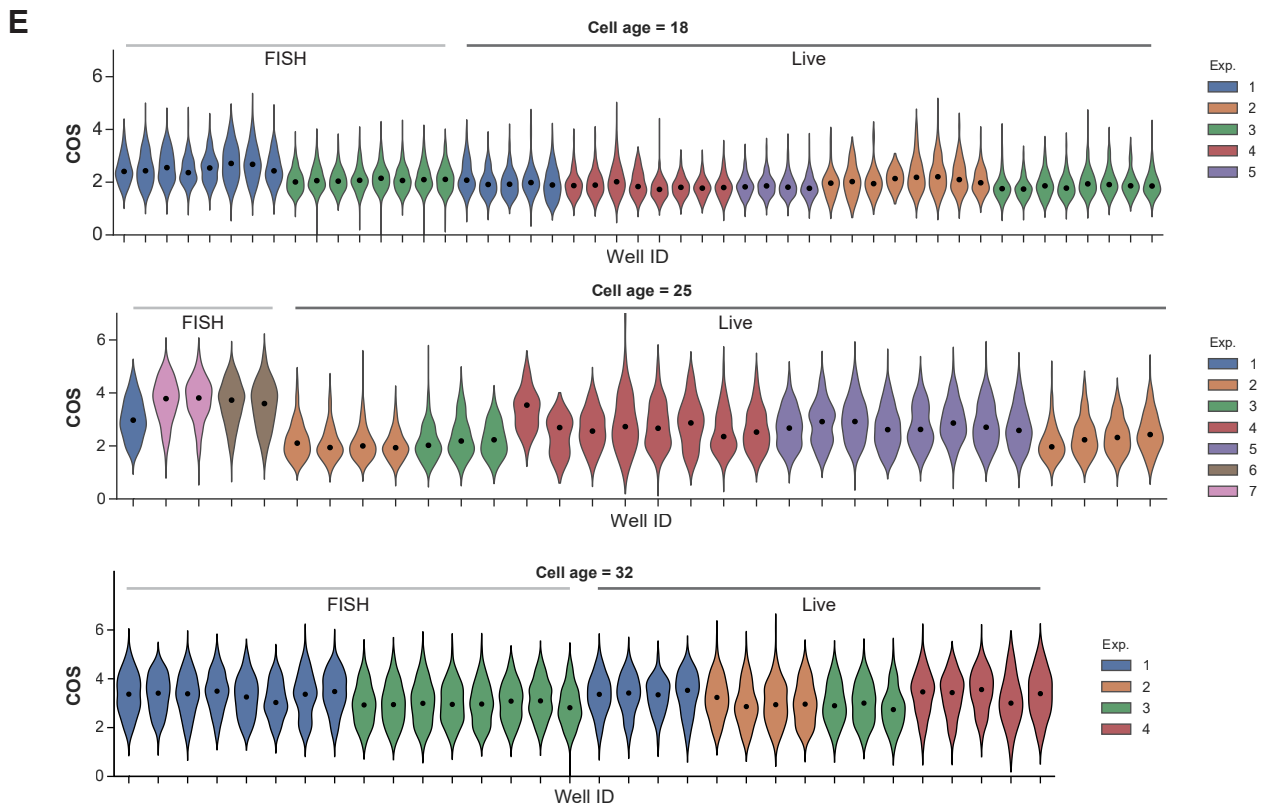

E
Cell age = 18
FISH
Live
Cell age = 25
FISH
Live
Cell age = 32
FISH
Live
COS
Well ID
Exp.
1
2
3
4
5
6
7

Figure S4, related to Figure 5: A single metric combining multiple cell features can quantify cardiomyocyte organization at scale across datasets and timepoints

A. Histograms for the original training set and each of the test sets comparing the expert annotations of organization to the combined organizational score, colored by expert annotation score; 1-2 (blue); 3 (orange); or 4-5 (green). Training set: n = 4,823 cells, test set 1: n = 932 cells, test set 2: n = 922 cells.
B. Histograms of 9 of the features used as input for the linear model used to derive the combined organizational score are shown in relation to expert annotation of organization. Features are organized by type: cell features (pink box), local organization (green box), or global alignment metric (blue box); histograms colored by expert annotation score given as 1-2 (blue), 3 (orange), or 4-5 (green). Organizational score feature weight and confidence interval are displayed for each feature (also shown in **Figure 4 Step F**).
C. Graph showing the expert organizational score per cell (x-axis) versus combined organizational score as computed using the linear regression model (y-axis). Both expert annotations are included for each cell and therefore each cell is represented twice. Each dot represents a cell, colored by time point (D18: blue, D32: pink). n = 6,370 cells. Example cells are shown in D and noted with boxes labeled i-viii.
D. Example cells from different regions of the plot representing a range of scores annotated in panel C. Alpha-actinin-2 protein (mEGFP-tagged line) localization is shown in white and nuclei in cyan. Box boundaries indicate age of cell, D18: blue, D32: pink. Scale bars = 20 μm
E. Violin plots showing well-to-well and experiment-to-experiment variation in COS for all three time points, D18 (top panel), D25 (middle panel), and D32 (bottom panel). One violin is shown per well, which is colored by experiment number. Dot indicates median.

Figure S5

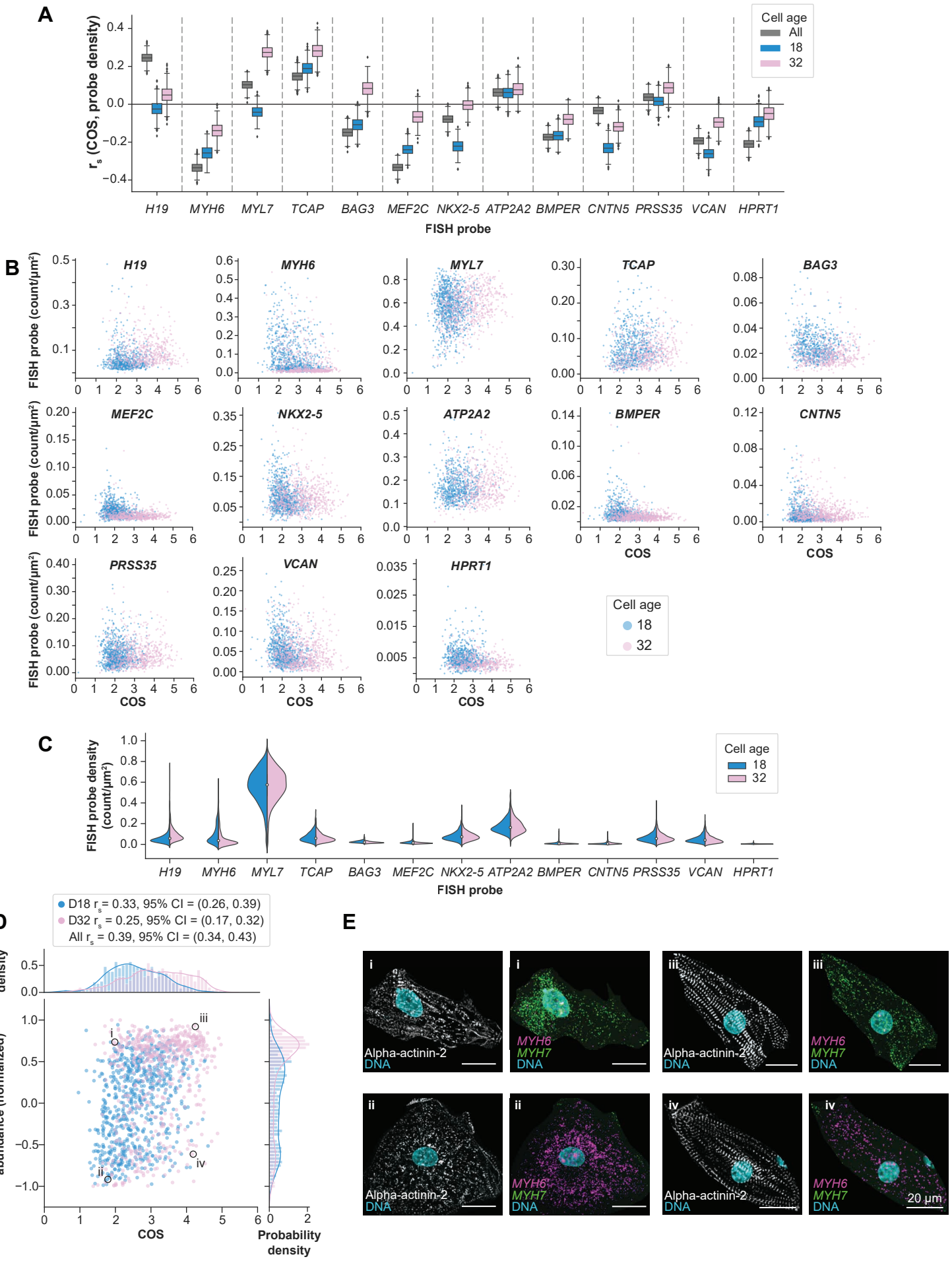

A
Cell age
All
18
32
r_s (COS, probe density)
H19
MYH6
MYL7
TCAP
BAG3
MEF2C
NKX2-5
ATP2A2
BMPER
CNTN5
PRSS35
VCAN
HPRT1
FISH probe
B
FISH probe (count/µm²)
COS
C
FISH probe density (count/µm²)
D
D18 r_s = 0.33, 95% CI = (0.26, 0.39)
D32 r_s = 0.25, 95% CI = (0.17, 0.32)
All r_s = 0.39, 95% CI = (0.34, 0.43)
Probability density
MYH7-MYH6 relative transcript abundance (normalized)
E
Alpha-actinin-2
DNA
MYH6
MYH7
20 µm

Figure S5, related to Figure 6: Measurement of correlation between cell organization and transcript abundance allows for identification of genes that correlate with cell organization, timepoint, or both

A. Box and whisker plots showing the bootstrapped Spearman rank correlation between combined organizational score and transcript abundance of the genes listed. Correlation metrics are shown for each of the genes that were probed by RNA FISH in alpha-actinin-2-mEGFP cardiomyocytes for each time point and as combined populations. Grey boxplots indicate the correlation across all cells; blue and pink boxplots show the correlation for cells stratified by age (D18: blue, D32: pink). Outliers are defined as outside 1.5 times the interquartile range (IQR).
B. Scatter plots showing transcript abundance from RNA FISH, reported as density (count/μm^2) (y axis) versus combined organizational score (COS) (x-axis) for each evaluated gene. Transcript abundance is measured by FISH probe spot density (counts per μm^2) in single cells. Each plot represents one gene and each dot represents a single cell (n = 719 D18, n = 578 D32, n = 1,297 total cells for *MYH6* and *MYH7;* n = 717 D18, n = 487 D32, n = 1,204 total cells for *COL2A1* and *HPRT1;* n = 639 D18, n = 522 D32, n = 1,161 total cells for *H19* and *ATP2A2;* n = 639 D18, n = 522 D32, n = 1,161 total cells for *BAG3* and *TCAP;* n = 868 D18, n = 766 D32, n = 1,634 total cells for *BMPER* and *VCAN*; n = 902 D18, n = 672 D32, n = 1,574 total cells for *PLN* and *PRSS35*; n = 813 D18, n = 750 D32, n = 1,563 total cells for *NKX2-5* and *CNTN5*; and n = 896 D18, n = 645 D32, n = 1,541 total cells for *MEF2C* and *MYL7*). Colors indicate cell age (D18: blue, D32: pink).
C. Violin plot showing distributions of transcript abundance for each gene by timepoint. Transcript abundance is measured by FISH probe spot density (counts per μm^2) in single cells. Dot indicates median. Colors indicate cell age (D18: blue, D32: pink). (Cell counts are the same as in (B); n = 1,297 cells for *MYH6* and *MYH7,* n = 1,204 cells for *COL2A1* and *HPRT1,* n = 1,161 cells for *H19* and *ATP2A2,* n = 1,161 cells for *BAG3* and *TCAP,* n = 1,634 for *BMPER* and *VCAN*, n = 1,574 for *PLN* and *PRSS35*, n = 1,563 for *NKX2-5* and *CNTN5* and n = 1,541 for *MEF2C* and *MYL7*).
D. Scatter plot with normalized relative abundance of *MYH7* compared to *MYH6* transcripts (y-axis) calculated from RNA FISH transcript abundance plotted against the combined organizational score (x-axis) across single cells (n = 1,297) (see Materials and Methods for details of relative expression calculation; levels of *MYH7* were 4-5x higher than those of *MYH6*, thus the difference between these two genes is scaled from -1 (100% *MYH6*) to +1 (100% *MYH7*), with the median of 0 representing cells expressing both genes). Colors indicate cell age (D18: blue, D32: pink). Top histogram shows the marginal distribution of the organizational scores stratified by cell age as a probability density normalized such that the area under each curve is equal to one. Right histogram shows the marginal distribution of the relative expression of *MYH7* and *MYH6*, also stratified by cell age. Spearman correlations (r_s) and confidence intervals (CIs) are as noted in the figure with “All” referring to correlations of combined D18 and D32 cells. Circles marked by Roman numerals (i-iv) refer to example cells shown in panel E.
E. Example images of four cells representing a range of transcript abundance and combined organizational scores as indicated in panel D, showing alpha-actinin-2-mEGFP protein localization (white) on the left for each cell and *MYH6* transcript (magenta), and *MYH7* transcript (green) by RNA FISH on the right. DNA is shown in cyan. Example cells were cropped from full field-of-view images based on manual annotation of cell boundaries. Scale bar = 20 μm.

Figure S6

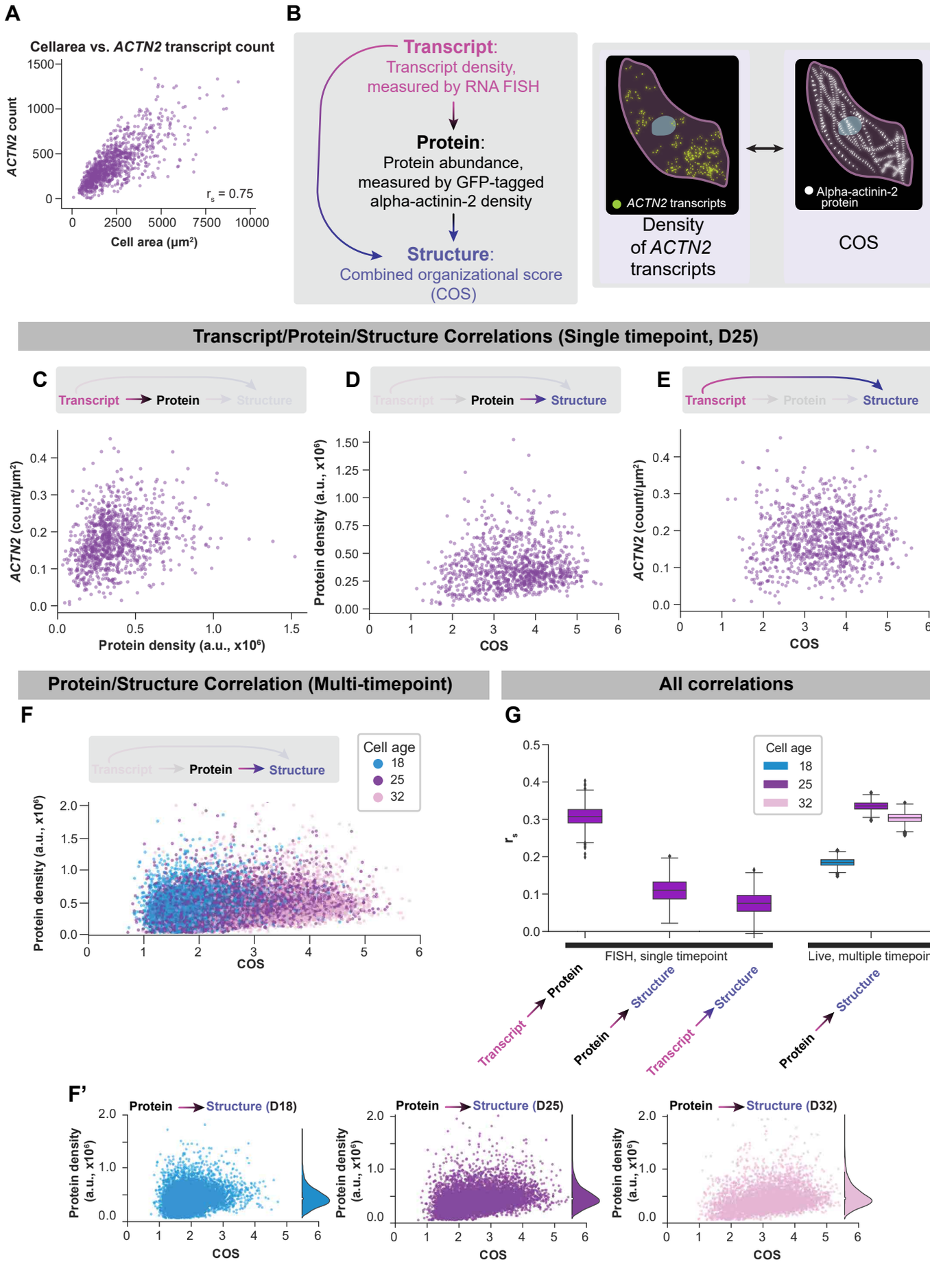
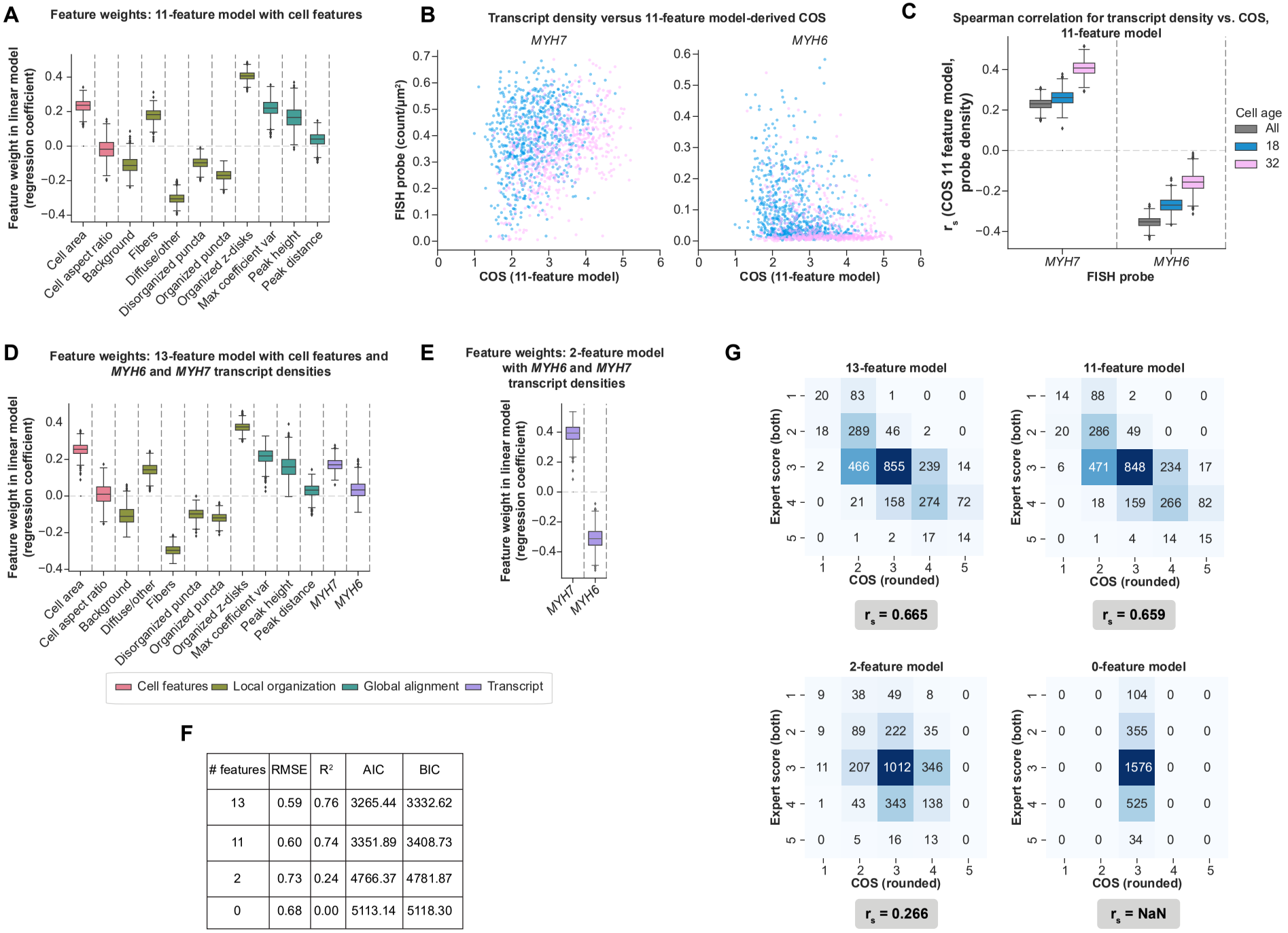

A
Cellarea vs. ACTN2 transcript count
ACTN2 count
Cell area (μm²)
r_s = 0.75
B
Transcript:
Transcript density, measured by RNA FISH
Protein:
Protein abundance, measured by GFP-tagged alpha-actinin-2 density
Structure:
Combined organizational score (COS)
ACTN2 transcripts
Alpha-actinin-2 protein
Density of ACTN2 transcripts
COS
Transcript/Protein/Structure Correlations (Single timepoint, D25)
C
Transcript → Protein
ACTN2 (count/μm²)
Protein density (a.u., x10⁶)
D
Protein → Structure
Protein density (a.u., x10⁶)
COS
E
Transcript → Structure
ACTN2 (count/μm²)
COS
Protein/Structure Correlation (Multi-timepoint)
F
Transcript → Protein → Structure
Cell age
18
25
32
Protein density (a.u., x10⁶)
COS
All correlations
G
r_s
Cell age
18
25
32
FISH, single timepoint
Live, multiple timepoints
Transcript → Protein
Protein → Structure
Transcript → Structure
Protein → Structure
F'
Protein → Structure (D18)
Protein → Structure (D25)
Protein → Structure (D32)
Protein density (a.u., x10⁶)
COS

Figure S6, related to Figure 6: Simultaneous measurement of transcript abundance, protein abundance, and organization of alpha-actinin-2 provides experimental context for correlations

A. Scatter plot showing the relationship between cell area (x-axis) and *ACTN2* transcript count; given the strong correlation (r_S = 0.75) between cell area and probe transcript count, we use probe density as our readout for transcript abundance in all subsequent data analysis.
B. Schematic illustrating the experimental goal of establishing correlations between gene expression (measured by RNA FISH), protein abundance of alpha-actinin-2 (measured by mEGFP intensity), and structural organization (measured by COS).
C. Scatter plot for D25 fixed cells showing alpha-actinin-2-mEGFP protein amount, reported as density (intensity (a.u.)/μm^2) (x-axis) versus *ACNT2* transcript abundance (y-axis). Transcript abundance is measured by FISH probe spot density (counts per μm^2) in single cells. Spearman correlation (r_S) equals 0.309 with a 95% CI of 0.252 to 0.362. n = 1,031 cells.
D. Scatter plot for D25 fixed cells showing combined organizational score (x-axis) versus alpha-actinin-2-mEGFP protein amount, reported as density (intensity (a.u.)/μm^2) (y-axis). Spearman correlation (r_S) equals 0.107 with a 95% CI of 0.049 to 0.169. n = 1,031 cells.
E. Scatter plot for D25 fixed cells showing combined organizational score (x-axis) versus *ACNT2* transcript abundance (y-axis). Transcript abundance is measured by FISH probe spot density (counts per μm^2) in single cells. Spearman correlation (r_S) equals 0.075 with a 95% CI of 0.008 to 0.135. n = 1,031 cells.
F. Scatter plot for D18, D25 and D32 live cells showing combined organizational score (x-axis) versus alpha-actinin-2-mEGFP protein amount, reported as density (intensity (a.u.)/μm^2) (y-axis). n = 18,045 cells. Subplots in F' show correlation for individual timepoints; density plots show the alpha-actinin-2-mEGFP protein level distributions for each timepoint. n = 7,544 for D18, n = 6,214 for D25, n = 4,287 for D32. D18: blue, D25: purple, D32: pink.
G. Box and whisker plot showing the bootstrapped Spearman rank correlation between transcript abundance, protein abundance or combined organizational score for live or fixed samples, as indicated. D18: blue, D25: purple, D32: pink. Outliers are defined as outside 1.5 times the interquartile range (IQR).

Figure S7



F

# features	RMSE	R^2	AIC	BIC
13	0.59	0.76	3265.44	3332.62
11	0.60	0.74	3351.89	3408.73
2	0.73	0.24	4766.37	4781.87
0	0.68	0.00	5113.14	5118.30

Figure S7, related to Figure 4: Incorporating *MYH6* and *MYH7* transcript abundance as features for predicting expert scores has a minimal impact on model performance

A. Relative contribution of features to 11 feature linear regression model used to calculate a cell's combined organizational score. The weight (regression coefficient) is plotted for each feature, colored by feature type. The training set contained both D18 and D32 cells (n = 1,297 cells in total probed for *MYH6* and *MYH7* transcript with RNA FISH, n = 11 total features). Box and whisker plots show regression coefficients from 1000 bootstrap rounds (see Materials and Methods). Outliers are defined as outside 1.5 times the interquartile range (IQR).Units for features: cell area in μm^2; all local organization metrics (in green) as fraction of cell area; peak distance in µm.
B. Scatter plot showing *MYH7* and *MYH6* transcript abundance from RNA FISH, reported as density (count/μm^2) (y-axis) versus COS (x-axis) from 11 feature regression model (model without transcript features). Transcript abundance is measured by FISH probe spot density (counts per μm^2) in single cells; each dot represents a single cell (n = 1,297 cells probed for *MYH6* and *MYH7* transcript with RNA FISH).
C. Box and whisker plots showing the bootstrapped Spearman rank correlation between transcript abundance of MYH*7* and *MYH6* and COS from 11 feature regression model (model without transcript features). Correlation metrics are shown for each gene assayed by RNA FISH in alpha-actinin-2-mEGFP cardiomyocytes, for each time point and as combined populations. Grey box plots indicate the correlation across all cells; blue and pink boxplots show the correlation for cells stratified by age (D18: blue, D32: pink; n = 1,297 cells in total probed for *MYH6* and *MYH7* transcript with RNA FISH). Outliers are defined as outside 1.5 times the interquartile range (IQR).
D. Relative contribution of features to 13 feature linear regression model (11 features from Panel A in addition to *MYH6* and *MYH7* transcript density) used to calculate a cell's combined organizational score. The weight (regression coefficient) is plotted for each feature, colored by feature type. The training set contained both D18 and D32 cells (n = 1,297 cells in total probed for *MYH6* and *MYH7* transcript with RNA FISH, n = 13 total features). Box and whisker plots show regression coefficients from 1000 bootstrap rounds (see Materials and Methods). Outliers are defined as outside 1.5 times the interquartile range (IQR). Units for features: cell area in μm^2; all local organization metrics (in green) as fraction of cell area; peak distance in µm; transcript (in purple) as RNA FISH spot density in count/μm^2.
E. Relative contribution of transcript features to 2 feature linear regression model (*MYH6* and *MYH7* transcript densities only) used to calculate a cell's combined organizational score. The weight (regression coefficient) is plotted for each gene. The training set contained both D18 and D32 cells (n = 1,297 cells in total probed for *MYH6* and *MYH7* transcript with RNA FISH, n = 2 total features). Box and whisker plots show regression coefficients from 1000 bootstrap rounds (see Materials and Methods). Outliers are defined as outside 1.5 times the interquartile range (IQR). Transcript feature units are RNA FISH spot density in count/μm^2.
F. Model performance metrics for 13, 11, 2, and 0 (intercept only) feature regression models. RMSE=root mean squared error; R^2=adjusted coefficient of determination; AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion.
G. Duplexed RNA FISH with *MYH6* and *MYH7* probes was performed on a training set of 1,297 cells, and each cell was scored for level of sarcomere organization by two expert scorers. Expert scores were the dependent variable in the linear regression models with 0, 2, 11, and 13 features. Confusion matrices show concordance between the rounded COS and both expert annotators. Overall correlation values between the COS and human scores are given below each confusion matrix.

Table S1: Differentiation and re-plate metadata, related to STAR Methods
Sample metadata for all differentiation and re-plate experiments used in this study (AICS-0075 cl.85 ACTN2-mEGFP). For each of the 7 experiments (Experiment number), the time point and sample types (live imaging or fixed for RNA FISH) are listed. Each experiment resulted from harvesting wells of differentiated cardiomyocytes (number of wells from a 6-well-plate are listed in column F), and flow cytometry was performed for quality control. Percent cTnT is the percent of the harvested population that expressed cardiac troponin T by flow cytometry analysis, and Percent GFP is the double positive population (GFP+/cTnT+) indicating percent of cardiomyocytes (cTnT+) that are also expressing an endogenously tagged mEGFP-alpha-actinin-2 protein (GFP+).

Table S2: Feature weights for the Combined Organizational Score, related to Figures 4 and 5 and STAR Methods
Feature weights from the linear regression model used to calculate a cell's combined organizational score (COS) as shown in Figure 4. For each feature, the feature weights (mean, Column B) and 95 percent confidence intervals are given (Columns C and D).

Table S3: Genes evaluated using RNA FISH, related to Figure 6 and Figure S5
Genes evaluated using RNA FISH were chosen for either known biological relevance (biological relevance, Column F), from an independent single cell RNA-seq study (scRNA-seq, Column F), or both as explained in the Results. RNA FISH probe sets for each listed gene can be ordered from Molecular Instruments/Molecular Technologies using the unique probe ID listed here (Column D). Protein name (Column B) and NCBI accession number (Column C) for the sequence used to design probe sets is also given.

Table S4: Correlation coefficients and confidence intervals, related to Figure 6 and Figures S5, S6
Mean Spearman correlations and 95 percent confidence intervals for COS vs. FISH probe density (boxplots from 1000 bootstraps shown in Figure 6B, Figure S5A, D), *ACTN2* probe count vs. cell area, *ACTN2* FISH probe density vs. alpha-actinin-2 protein density, and COS vs. alpha-actinin-2 protein density (Figure S6). For COS vs FISH probe density, values are given for D18, D25 (select genes only), D32, and All (All = D18 and D32 combined) for each gene.