

Single-Cell Sequencing of the Healthy and Diseased Heart Reveals Ckap4 as a New Modulator of Fibroblasts Activation

Running Title: *Gladka et al.; Single-Cell Sequencing of the Adult Heart*

Monika M. Gladka, PhD^{1#}; Bas Molenaar, MSc^{1#}; Hesther de Ruiter, BS¹;

Stefan van der Elst, BS¹; Hoyee Tsui, PhD¹; Danielle Versteeg, BS^{1,2};

Grègory P.A. Lacraz, PhD¹; Manon M.H. Huibers, PhD³; Alexander van Oudenaarden, PhD¹;

Eva van Rooij, PhD^{1,2*}

[#]These authors contributed equally to this work



¹Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Centre, Utrecht, The Netherlands; ²Department of Cardiology, University Medical Centre, Utrecht, The Netherlands; ³Department of Pathology, University Medical Centre, Utrecht, The Netherlands

***Address for Correspondence:**

Eva van Rooij, PhD
Hubrecht Institute, KNAW
Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.
Tel: +31 30 2121956
Email: e.vanrooij@hubrecht.eu

Abstract

Background—Genome-wide transcriptome analysis has greatly advanced our understanding of the regulatory networks underlying basic cardiac biology and mechanisms driving disease. However, so far, the resolution of studying gene expression patterns in the adult heart has been limited to the level of extracts from whole tissues. The use of tissue homogenates inherently causes the loss of any information on cellular origin or cell type-specific changes in gene expression. Recent developments in RNA amplification strategies provide a unique opportunity to use small amounts of input RNA for genome-wide sequencing of single cells.

Methods—Here, we present a method to obtain high quality RNA from digested cardiac tissue from adult mice for automated single-cell sequencing of both the healthy and diseased heart.

Results—After optimization, we were able to perform single-cell sequencing on adult cardiac tissue under both homeostatic conditions and after ischemic injury. Clustering analysis based on differential gene expression unveiled known and novel markers of all main cardiac cell types. Based on differential gene expression we were also able to identify multiple subpopulations within a certain cell type. Furthermore, applying single-cell sequencing on both the healthy and the injured heart indicated the presence of disease-specific cell subpopulations. As such, we identified cytoskeleton associated protein 4 (*Ckap4*) as a novel marker for activated fibroblasts that positively correlates with known myofibroblast markers in both mouse and human cardiac tissue. *Ckap4* inhibition in activated fibroblasts treated with TGF β triggered a greater increase in the expression of genes related to activated fibroblasts compared to control, suggesting a role of *Ckap4* in modulating fibroblast activation in the injured heart.

Conclusions—Single-cell sequencing on both the healthy and diseased adult heart allows us to study transcriptomic differences between cardiac cells, as well as cell type-specific changes in gene expression during cardiac disease. This new approach provides a wealth of novel insights into molecular changes that underlie the cellular processes relevant for cardiac biology and pathophysiology. Applying this technology could lead to the discovery of new therapeutic targets relevant for heart disease.

Key Words: Adult heart; single-cell sequencing; ischemic injury; myofibroblast; *Ckap4*

Clinical Perspective

What is new?

- Single-cell sequencing can be used to identify subpopulation-specific and novel cell type-specific markers in the healthy and diseased heart.
- Identification of new disease-driven cell populations provides insights into gene expression changes that are triggered by ischemic injury.
- Myozenin2-enriched cardiomyocytes form a distinct subpopulation of cardiomyocytes in the healthy heart.
- CKAP4 is a novel marker for activated fibroblasts that positively correlates with known myofibroblast markers in both murine and human diseased hearts.
- *In vitro* experiments suggest a modulating function for CKAP4 in myofibroblast activation.



What are the clinical implications?

- Single-cell sequencing of the adult heart allows us to examine molecular mechanisms that drive the cellular processes underlying heart disease.
- New biology discovered by single-cell sequencing can ultimately lead to the development of novel therapeutic strategies.
- We identified CKAP4 as a new marker for activated cardiac fibroblasts during ischemic injury that appears to attenuate myofibroblast activation.

Introduction

The heart consists of a collection of different cell types that coordinately regulate cardiac function¹. Changes in cellular composition and function are mechanistically responsible for cardiac remodeling and repair during disease. Identifying the underlying differences in gene expression between cell types or transcriptome heterogeneity across cells of the same type will greatly help to improve our understanding of cellular changes under both healthy and diseased conditions². RNA amplification strategies have provided the opportunity to use small amounts of input RNA for genome-wide gene expression analysis at single cell resolution. The analysis of individual cells randomly drawn from a sample allows for an unbiased view of all mRNAs present in different cell types of an organ, which will provide a more accurate classification of cellular diversity through differences in gene expression²⁻⁵. Recently, transcriptional profiling of cardiac cell populations during heart development by single-cell sequencing has led to the identification of lineage-specific gene programs that underlie early cardiac development^{4,6}. However, so far, no studies have focused on single-cell transcriptomics of the adult heart.

Here we present a method for obtaining high quality RNA from digested adult cardiac tissue for automated single-cell sequencing of both the healthy and diseased heart. Using this approach, we were able to collect reliable gene expression data for all main cardiac cell types. Clustering analysis uncovered both known and novel markers of certain cell populations, and led to the identification of multiple subpopulations within a certain cell type. Single-cell sequencing analysis of both healthy hearts and hearts suffering from ischemic injury indicated that cardiac damage gives rise to new subpopulations of known cell types. Using our single-cell sequencing data we were able to identify CKAP4 as a novel marker for activated fibroblasts, which we were able to validate in cardiac samples from patients suffering from ischemic heart disease. In a set

of *in vitro* experiments, we were able to show that CKAP4 is involved in myofibroblast activation.

Altogether, our data for the first time show the feasibility of using single-cell sequencing on the adult heart to study transcriptomic differences between cardiac cell types and the heterogeneity in gene expression within one cell population. This method will greatly advance the molecular insights into cellular mechanisms that are relevant for cardiac remodeling and function.

Methods

An Expanded Methods section, a step-by-step protocol for the single-cell sequencing approach and any associated references are available in the Online data supplements. The data, analytic methods, and study materials will be made available by the authors to other researchers for purposes of reproducing the results or replicating the procedure.

Experimental animals

All animal studies were performed in accordance with institutional guidelines and regulations of the Animal Welfare Committee of the Royal Netherlands Academy of Arts and Sciences.

C57BL/6J mice were subjected to sham (control) or ischemia reperfusion (IR) surgery as previously reported ⁷. Hearts were collected and analyzed 3 days after surgery.

Digestion of the heart

After collecting the infarcted areas (infarct and border zone region) or the corresponding region of control hearts, the tissue was digested for 15 min and used for subsequent RNA isolation or single-cell sorting and sequencing.

Flow cytometry

The freshly collected cardiac cell lysates were resuspended in DMEM and living single cells were sorted into 384-well plates based on multiple scatter properties and DAPI exclusion. After cell sorting, the plates were immediately centrifuged and stored at -80°C.

Library preparation and sequencing of single cells

The SORT-seq procedure was applied as described previously⁵. Illumina sequencing libraries were prepared using the TruSeq small RNA primers (Illumina) and sequenced paired-end at 75 bp read length with Illumina NextSeq.

Data analysis of single-cell RNA sequencing

Paired-end reads from Illumina sequencing were mapped with BWA-ALN⁸ to the reference genome GRCm38/mm10. For quantification of transcript abundance, the number of transcripts containing unique molecular identifiers (UMIs) per cell-specific barcode were counted for each gene. Next, the RaceID2 algorithm was used to cluster cells based on K-medoids clustering, to visualize cell clusters using t-distributed stochastic neighbour embedding (t-SNE) and to compute genes up- or down-regulated in all cells within the cluster compared to cells not in the cluster^{2,3}.

Bulk sequencing and data analysis

For bulk sequencing, RNA from cardiac tissue was isolated using Trizol. Subsequently, libraries were prepared and sequenced using a similar protocol as described for single-cell RNA sequencing.

Pathway analysis and gene ontology

To investigate whether differentially expressed genes in subgroups of cells share a similar biological function, enrichment analyses on these genes were performed using the gene ontology

(GO) enrichment tool and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis using DAVID⁹. Significant enrichment of genes in GO terms and KEGG pathway analyses are shown as p values corrected for multiple testing using the Benjamini-Hochberg method.

Human heart samples

Approval for studies on human tissue samples was obtained from the Medical Ethics Committee of the University Medical Center Utrecht, The Netherlands (12#387). In this study we included tissue from the left ventricular free wall of patients with ischemic heart disease (infarct, border zone, and remote) and left ventricular free wall of non-failing donor hearts.

Gene expression values obtained by qPCR were plotted for correlation analysis.



Statistical analysis (qPCR). The number of samples (n) used in each experiment is indicated in the legend or shown in the figures. The results are presented as mean \pm standard error of the mean. For qPCR analysis, statistical analyses were performed using PRISM (GraphPad Software Inc.). If 2 groups were compared a Student's *t*-test was used.

Results

Isolation of high quality RNA from digested hearts from adult mice

To create a reliable gene expression atlas of all cardiac cell types we first aimed to determine the optimal method for tissue digestion and RNA extraction to obtain high quality RNA from single-cell suspensions of the adult heart. To do so, mice were euthanized after which the hearts were perfused with cold perfusion buffer. After collecting the anterior wall of the left ventricle, the tissue was washed in ice-cold perfusion buffer, kept on ice, minced into small pieces and transferred into a glass vial with cold digestion buffer. Following digestion of the tissue the cell

suspension was gently pipetted up and down after which the lysate was passed through a 100 μ m cell strainer. The strainer was then rinsed with DMEM and the cells were collected for subsequent RNA extraction or flow cytometry (**Figure 1A-B**).

To optimize our digestion protocol, we started out by testing 4 different digestion solutions that are commonly used to digest muscle tissue, containing either Liberase, Collagenase II, Pancreatin or Trypsin¹⁰⁻¹³. Based on cellular imaging and assessment of RNA quality by RNA Integrity Number (RIN) the digestion solution containing Liberase appeared most optimal for dispersing adult cardiac tissue while maintaining intact RNA (**Figure S1A-B**).

To compare both the influence of the solution containing the single-cell suspension and time the samples were kept on ice before further processing, we tested both DPBS with 5% FBS¹⁴ and DMEM and collected the cells for RNA analysis either immediately or after having been on ice for 30 or 60 minutes. While the time on ice did not seem to influence the RIN, using DMEM appeared to provide higher quality RNA when lysing cardiac tissue (**Figure S1C**).

Next, we examined whether digestion in a 37°C incubator would be better than a 37°C shaking water bath using two different concentrations of the digestion enzyme Liberase. Based on RIN, 0.5 mg Liberase yielded good quality RNA, and visually it was evident that using the shaking water bath was better at digesting the pieces of cardiac tissue into suspension (**Figure S1D**). RNA isolated from mechanically homogenized cardiac tissue was taken along as a positive control (control heart).

To test whether the method of RNA isolation would influence the RNA quality we used both the mirVana RNA isolation kit (Thermo Fisher Scientific) and Trizol (Invitrogen) to isolate RNA from the digested cardiac cells. RNA integrity was comparable for samples isolated with Trizol or mirVana (**Figure S1E**).

Together these data indicated that based on cell morphology and RNA quality using Liberase to digest adult cardiac tissue for 15 min in a 37°C shaking water bath prior to Trizol RNA isolation, was optimal for obtaining a single-cell suspension of the heart (**Figure S1E**, indicated by an arrow).

Single-cell sorting strategy

After enzymatically dispersing cardiac tissue, flow cytometry was used to separate the cells. Empirically, we found that using a large nozzle size (130 μm) allowed for sorting all range of cardiac cells without damaging larger cardiomyocytes. We based our gating strategy on multiple scatter properties including DAPI to sort for living cells, and green autofluorescence to sort for more complex cells that contain cytoskeletal filaments (**Figure 1C**)^{15, 16}. Our results indicated that on average 89.4% of the cells from control hearts were viable after sorting, of which 92% showed an autofluorescent signal (**Figure 1C**). Additionally, to enrich for bigger cells, we selected for cells with a higher forward scatter width (FSC-Width). Imaging the cells after sorting indicated that the cells remained intact and suitable for further sequencing applications (**Figure 1D-E**). Since cardiomyocytes are notoriously difficult to isolate by sorting strategies we wanted to confirm the quality of the isolated cardiomyocytes, post sorting. To do so we crossed $\alpha\text{MHC-Cre}$ transgenic mice with R26 lox-Stop-lox tdTomato mice to mark the cardiomyocyte population (**Figure S2A**). After sorting we obtained 88.8% living cells (**Figure S2B**), in line with data obtained in **Figure 1C**. To confirm these were living cells, we resorted the single-cell lysate, which indicated 99.5% of these sorted cells were alive. (**Figure S2C**). Imaging of the Tomato signals showed the cardiomyocytes to be intact after the sorting procedure (**Figure S2D**). To ensure that the quality of RNA remained intact even after sorting we collected different quantities of cardiac cells and isolated RNA. Bioanalyzer results indicate that after cell

separation the RNA isolated from the dispersed and sorted cells remained of good quality as indicated by RIN (**Figure 1E**).

Single-cell sequencing to identify gene expression signatures in all main cardiac cell types

Using the SORT-Seq protocol,⁵ on average we detected 16874 of raw unique reads per cell. The distribution of the readcounts across cells indicated the reads come from single cells, since we did not observe the bi-modal or multi-modal distribution of reads one would expect when detecting transcript from doublets or multiplets (**Figure S2E**). After applying filtering procedures for quality and input (see online-only Data Supplement), a total of 426 cells from three different control hearts were used for downstream *in-silico* analysis.

For the identification and analysis of all main cardiac cell types we used the RaceID2¹¹ algorithm^{3,17}. K-medoids clustering of 1-Pearson correlation coefficients revealed 14 distinct cell clusters in the adult heart (**Figure 2A**). The separation between the different cell clusters was further validated by a t-Distributed Stochastic Neighbor Embedding (t-SNE) map showing lower intra-cluster cell-to-cell distance compared to the inter-cluster distances (**Figure 2B**). Next, we assessed which genes were differentially expressed within each cluster compared to all other clusters (**Database S1**). We used the abundance of known marker genes to determine the cell identity of the different cardiac cell clusters (**Figure 2C and Database S1**). We were able to identify clusters belonging to all major cell populations in the heart (**Figure S3A**). T-SNE maps showing the expression of some of these marker genes indicated the presence of cardiomyocytes (**Figure 2D**), fibroblasts (**Figure 2E**), endothelial cells (**Figure 2F**) and macrophages (**Figure 2G**).

Taken together, these findings demonstrate that by using our method we are able to generate a single cell gene expression profiling of the adult heart, which can identify all major cell types.

Contribution of mitochondrial transcripts differ for individual cell types

While investigating gene expression signatures in all main cardiac cell types (**Figure S3A and Database S1**), we observed that the contribution of mitochondrial and genomic transcripts varied among different cells. All cardiomyocytes have higher percentages of mitochondrial transcripts (58-86% of total transcripts) when compared to other cardiac cell types (**Figure S3B-E**). Since mitochondrial transcripts are so abundant we excluded them from the clustering analysis and focused only on the differential expression of genomic genes.



Single-cell sequencing identifies cell-type specific subpopulations in the healthy heart

Single-cell sequencing of the adult heart revealed multiple clusters within the same cell type (**Figure 2C and Figure S3**). These cells are bioinformatically clustered based on the differential expression of genes, while they also express marker genes that identify them as a specific cell type. According to the gene expression of marker genes we were able to identify 4 different clusters of cardiomyocytes, 2 clusters of endothelial cells, 2 clusters of fibroblasts, 2 clusters of macrophages, 1 cluster of smooth muscle cells and a cluster of erythrocytes (**Figure 2C and Figure S3A**).

To explore these clusters in more detail we focused on the cardiomyocyte clusters. These clusters are defined as cardiomyocytes based on the enriched expression of cardiomyocyte marker genes compared to other cells (**Figure 3A, indicated in red and Figure 3B**). The enriched expression of a divergent set of genes classifies them to as separate cell clusters based on the RaceID2 parameters (**Figure 3A, indicated in black and Figure 3B**). Interestingly, the relative

expression of the well-known cardiomyocyte markers genes, showed a high level of variation between the different cardiomyocyte clusters and as expected we observed them to be enriched compared to the fibroblasts. (**Figure 3C**).

Based on differential expression of cardiomyocyte marker genes, cluster 4 appeared to be the most divergent from the other cardiomyocyte clusters (**Figure 3C**). A t-SNE map confirmed that this cluster does express cardiomyocyte markers, such as cardiac troponin T, *Tnnt2* (**Figure 4A**), however, it is the only subpopulation of cardiomyocytes that is enriched for *Myozenin2* (*Myoz2*) expression (**Figure 3** and **Figure 4B**). To determine whether the clustering of *Myoz2* expression is a product of stochastic gene expression within the cardiomyocytes of the heart, we aimed to validate this cluster in an independent mouse model. To do so we specifically sorted and sequenced cardiomyocytes from cardiac tissue from mice in which we labelled the cardiomyocytes with tdTomato (α MHC-Cre transgenic mice crossed with R26 lox-Stop-lox tdTomato mice) (**Figure S2A**). Similar to the control hearts (**Figure S4A**), we also detected a *Myoz2*-enriched cardiomyocyte cluster (**Figure S4B**). In comparing the highly expressed genes in the *Myoz2*-enriched cardiomyocyte clusters from either C57BL/6J mice or tdTomato mice, we identified a large overlap in enriched genes from both clusters (**Figure S4C**). The overlap in gene enrichment strongly suggest the clustering data to be reliable and that the *Myoz2*-enriched cardiomyocyte cluster is indeed different from the other cardiomyocyte clusters.

Immunohistochemistry indicated MYOZ2 to be enriched in a layer of cardiomyocytes (co-stained with TNNT2) located at the epicardial surface of the heart (**Figure 4C**). Gene ontology analysis of the most differentially regulated genes in the *Myoz2*-enriched cluster versus other cardiomyocytes, indicated the highly detected genes to be involved in degenerative disorders of the central nervous system (**Figure 4D**), while the lowly expressed genes appeared

to be involved in cardiac diseases (**Figure 4E**). This is interesting as *Myoz2*, also known as Calsarcin-1, is an inhibitor of the pathological, pro-hypertrophic phosphatase Calcineurin.^{18, 19}

In conclusion, our data show that single-cell sequencing analysis on cardiac cells can serve to identify subpopulations of a certain cell type, thus, indicating a large heterogeneity between cells from the same cell type. Additionally, this approach allows for the detection of novel cell type-enriched gene expression, providing a basis for discovering new gene functions in certain cell types.

Single-cell sequencing of the injured heart

Ischemic heart disease is the most common form of cardiovascular disease that induces a remodelling response across the damaged area that involves fibroblast activation, immune cell infiltration, neoangiogenesis and a change in cardiomyocyte function^{20, 21}. The identification of new regulators, transcription factors and molecular pathways that are relevant for these cellular processes, could eventually lead to the development of new therapeutic strategies for patients suffering from this disease.

To determine whether our method would allow for studying the influence of disease on inter- and intracellular changes in gene expression, we exposed mice to ischemic injury (ischemia reperfusion, IR) and collect samples 3 days post IR (3dpIR),⁷ Using the same isolation procedure as described above we isolated 509 individual cells from the infarcted region of heart. (**Figure S5 and S6**). We pooled these cells with the 426 cells obtained previously from the corresponding region of control hearts for *in-silico* analysis with RaceID2. Using our optimized protocol, we were able to obtain good quality RNA from single cell suspensions from the infarct region, as assessed by RIN (**Figure S6B**). K-medoids clustering of 1-Pearson correlation revealed a total of 17 different cell cluster in all pooled cells from control and 3 dpIR hearts

(**Figure S6C**). Using the abundance of known marker genes to determine the cell identity of the different cardiac cell clusters we were able to identify clusters belonging to all major cell populations in the heart (**Figure S7A**), which was validated by t-SNE maps showing enriched marker gene expression in these individual clusters (**Figure S7B-E**). Based on marker gene expression (**Figure 3B**), we could identify all main cardiac cell types among the different clusters (**Figure S7**). To see how ischemic injury would affect expression of mitochondrial and genomic genes within all cell types, we compared the average relative expression of genomic and mitochondrial genes within each cluster obtained from either healthy and diseased hearts (**Figure S8A-D**). On average, the expression of mitochondrial genes decreased after ischemic injury within all cell clusters.



Inter- and intracellular gene expression changes induced by ischemic injury

By generating an t-SNE map to indicate transcriptome similarities between all individual cells, we were able to see that the majority of clusters contained cells from both control and injured hearts (**Figure 5A-B**). However, injury triggered the appearance of a disease-enriched cell clusters (**Figure 5A-D and S8E**).

Focussing on the fibroblasts, we were able to identify 3 different fibroblast clusters (7, 13 and 15) (**Figure 5 B-D**). Examination of the library origin, we observed that the cells in clusters 13 and 15 predominantly stemmed from the injured hearts, whereas cluster 7 contained cells from both conditions (**Figure 5C-D**). Cells from clusters 13 and 15 were characterized by the relatively high expression of *Postn*, *Wisp1* and *Tnc*, previously associated with fibroblast activation²²⁻²⁴, and t-SNE maps again confirmed the enriched expression of these genes in the disease-enriched fibroblast clusters (**Figure 5E and Figure S9**). The disease-specific induction

of these genes was validated by qPCR on cardiac tissue from either control or injured hearts (**Figure 5F**).

Ckap4 expression is specifically increased in activated fibroblasts

Having identified a population of fibroblasts stemming from both healthy and diseased hearts (cluster 7) and a population stemming predominantly from diseased hearts (cluster 13 and 15), we next determined the differentially expressed genes between these populations. Using a Log2 fold change of 1.5 or -1.5 or more with a p value of 0.05 we found a total of 11 genes up-regulated and 20 genes down-regulated in the disease-enriched fibroblasts population (**Figure 5G**). In using the same parameters for defining differentially expressed genes in larger cell clusters, the number of differentially expressed genes between cells from the healthy and ischemic hearts became much larger. For example, we found 205 genes differentially expressed between healthy and ischemic cardiomyocytes coming from clusters 1, 3, 4 and 8, while 191 genes were differentially expressed in macrophages coming from either the healthy or ischemic hearts forming cluster 5, 9, 14 and 16 (data not shown). The upregulated genes in these disease-enriched clusters were related to various processes associated with extracellular matrix deposition and collagen deposition (**Figure 5H**), a hallmark of the fibrotic response after ischemic injury of the heart²⁵. In line with this observation, among the enriched genes, we detected various collagen genes and markers known to be expressed in activated fibroblasts, such as *Postn*.²² (**Figure 5I**). The identification of clusters that are enriched for disease-related genes from ischemic hearts further underscores the validity of our clustering approach.

In addition to known markers for activated fibroblasts, we identified cytoskeleton associated protein 4 (*Ckap4*) to be upregulated in this population of activated fibroblasts. *Ckap4* is a transmembrane protein and can act as a receptor for various ligands in different cells^{26, 27},

however its function in cardiac fibroblasts remains unknown. We could confirm the upregulation of *Ckap4* upon ischemic injury in the whole heart by bulk RNA-sequencing (**Figure 5J**).

However, our single-cell sequencing data allowed us to detect the upregulation of *Ckap4* specifically in activated fibroblasts and not in any other cell types, which was confirmed by t-SNE maps (**Figure 5K-L**). By immunohistochemistry we validated the induction of CKAP4 in hearts post ischemic injury, which overlapped with cells expressing the fibroblasts marker vimentin (VIM) (**Figure 6A**). To investigate the functional relevance of *Ckap4* expression in activated fibroblasts we inhibited *Ckap4* expression in fibroblasts, after which we exposed them to TGF β , a well-known inducer of myofibroblasts formation (**Figure 6B**)²⁸. TGF β was able to induce *Ckap4* expression when compared to control. SiRNA-mediated inhibition of *Ckap4* resulted in a dose dependent reduction of *Ckap4* (**Figure 6C**), on the mRNA and the protein level (**Figure 6C-D**). Quantitative PCR analysis showed that TGF β treatment led to an increase in the expression of markers for activated fibroblasts and that inhibition of *Ckap4* further potentiated this effect under such conditions (**Figure 6E**). While this warrants further investigation, this seems to imply that CKAP4 in activated fibroblast functions to dampen the expression of genes related to activated fibroblast.

Also in cardiac samples from patients suffering from ischemic heart disease we were able to confirm a positive correlation in expression between CKAP4 and genes known to be induced in activated cardiac fibroblasts (**Figure 7A-C**). In addition, immunohistochemistry showed a strong overlap between *CKAP4* expression and various well-established fibroblast markers in human ischemic hearts (**Figure 8D**), suggesting *CKAP4* also has a role in activated fibroblasts in humans.

Taken together, our single cell RNA sequencing data on healthy and disease hearts demonstrate that we are able to identify disease-specific subpopulations of various cell types. Comparing gene expression patterns between healthy and disease subpopulations within cell types allowed us to detect cell type-specific upregulation of various genes. Using this approach, we identified *Ckap4*, as a marker specifically upregulated in activated fibroblasts.

Discussion

Here we show that our optimized technique to isolate and sort adult cardiac cells in combination with a high throughput method to sequence single cells with SORT-seq^{2, 5, 29}, provides a unique opportunity to reliably obtain single-cell gene expression data of the adult mammalian heart (Figure S10 and step-by-step protocol in Online Data Supplements). Among the sequenced cells, we were able to identify all major cardiac cell types, including cardiomyocytes, fibroblasts, endothelial cells and macrophages^{30,1}.

A major advantage of single-cell sequencing is the ability to detect heterogeneity within a certain cell type in the heart. Clustering analysis of our single-cell sequencing data shows differentially expressed genes in subsets of cells which are likely to contribute to the functional diversity within different cell types. For example, we found 4 clusters of cardiomyocytes that compared to fibroblast, as expected, show a significant enrichment for cardiomyocyte marker gene expression. However, our data also show that there is a substantial heterogeneity in expression of well-established cardiomyocyte markers between the different cardiomyocyte subpopultaion (Figure 3C). While we currently do not know the biological relevance of this observation, it could imply that there are functionally different cardiomyocyte populations already in a healthy heart.

In addition, we show that only a subset of cardiomyocytes express Myoz2, a protein belonging to a family of sarcomeric proteins¹⁸. Myoz2 has been shown to tether α -actinin to Calcineurin, a well-known inducer of cardiomyocyte hypertrophy³¹, thereby inhibiting the pathological hypertrophic response of cardiomyocytes¹⁸. Myoz2 expression limited to only a subset of cardiomyocytes, predominantly located towards the epicardial region of the heart, raises the possibility that subpopulations of cardiomyocytes respond differently to Calcineurin-mediated hypertrophy, with some being more resistant than others. A challenge for single cell sequencing studies is to discriminate between difference based on stochastic changes in gene expression or biology³². To show that the differences in expression in the Myoz2-enriched cardiomyocyte cluster are biologically relevant, we additionally clustered sequenced cardiomyocytes from an additional mouse model (Figure S4). Also in this study we were able to identify a distinct cluster of cardiomyocytes to be enriched for Myoz2 expression. When comparing the enriched genes in both Myoz2 clusters, we detected a large overlap suggesting that these clusters of Myoz2-enriched cardiomyocytes indeed resemble each other and might be functionally different compared to the other cardiomyocytes. Determining the biological function of this subset of cardiomyocytes, could yield valuable insights into cellular mechanisms responsible for cardiac function and pathologies.

Comparing single-cell sequencing data sets from both the healthy and injured heart revealed that disease gives rise to new subpopulations of known cell types that appear specific for or enriched with cells coming from the diseased heart. While it is known that the cellular composition of the heart changes during pathological stress³³, we are limited in our ability to detect genome-wide changes in expression specifically occurring within each cell type during cardiac stress. Our single-cell sequencing data now allows us to identify transcriptome-wide

differences in all major cardiac cell types between different conditions. By defining disease-related cell subpopulations and the associated changes in gene expression, we expect to identify novel molecular mechanisms that are relevant for the cellular changes underlying heart disease.

The potential of the described approach is nicely exemplified by the identification *Ckap4* as a novel marker for activated fibroblasts. *Ckap4* was previously reported to have a function in cell proliferation during tumour progression³⁴, but its function in the heart so far has been unstudied. In our data set we found *Ckap4* to be expressed in the same cell population as *Postn*, *Ctrc1* and *Fnl1*, well-known markers for cardiac myofibroblasts²²⁻²⁴. Immunohistochemistry on both control and ischemic hearts confirmed the expression of CKAP4 to be specific for the stressed heart and to overlap with Vimentin, a marker for fibroblasts, in the ischemic heart. *In vitro* loss-of-function of *Ckap4* showed an over activation of myofibroblast related genes before and after TGF β stimulation suggesting a role for *Ckap4* during fibroblast activation. Moreover, we observed a positive correlation between *CKAP4* and fibroblasts markers in cardiac biopsies from patients suffering from ischemic heart disease.

In an attempt to collect living cells for our single-cell analysis we biased our flow cytometry gating towards larger cells. As a result, we preferentially isolated cardiomyocytes (71% for sequenced cells in control heart and 59% in diseased hearts) while it is estimated that 30% of cardiac cells are cardiomyocytes³⁰. Nonetheless, we were still able to detect also other cell populations. Further optimization of our gating strategy is required to obtain a representative spectrum of cardiac cell types after cell sorting by flow cytometry.

Inherent to the technique, the low sequencing efficiency prevents us from detecting lower expressed genes in a cell. Since currently the majority of reads are being spend on mitochondrial genes (up to 23-84% of reads for cardiomyocytes clusters from both sham and IR), efforts to

separate out transcripts derived from the mitochondria could greatly enhance the sequencing efficiency.

While future developments will continue to optimize the single-cell sequencing technology on organs, our study for the first time indicates the feasibility of using this technique on adult cardiac tissue. Our data indicate the potential of this method to identify transcriptional differences between cardiac cell types and to study the heterogeneity in gene expression between the different subpopulations. These discoveries together create major opportunities to unveil new gene functions for cellular biology and organ function.

Acknowledgments



We thank Judith Vivié, Susanne van den Brink, Jean-Charles Boisset, Anna Alemany Arias and Mauro Muraro (Hubrecht Institute) for technical assistance and helpful discussions on single cell sequencing. We gratefully acknowledge Thomas Cahill for providing his expertise related to macrophages biology.

Sources of Funding

This work was supported by the Leducq Foundation (14CVD04) and the European Research Council under the European Union's Seventh Framework Programme (ERC Grant Agreement AdG 294325 GeneNoiseControl and CoG 615708 MICARUS). M.M.G. was supported by a Dr. Dekker postdoctoral fellowship from the Dutch Heart Foundation (NHS2016T009).

Author Contributions

M.M.G., B.M., G.P.A.L. and E.v.R. designed experiments. M.M.G., B.M., H.d.R., H.T., D.V. and S.v.d.E performed all experiments. M.M.H.H. provided human biopsies. M.M.G. and B.M. analyzed data. M.M.G., B.M. and E.v.R. wrote manuscript. M.M.G. and B.M contributed equally as joint first authors.

Disclosures

The authors declare no conflict of interest.

References



1. Pinto AR, Ilinykh A, Ivey MJ, Kuwabara JT, D'Antoni ML, Debuque R, Chandran A, Wang L, Arora K, Rosenthal NA and Tallquist MD. Revisiting Cardiac Cellular Composition. *Circ Res*. 2016;118:400-409.
2. Grun D and van Oudenaarden A. Design and Analysis of Single-Cell Sequencing Experiments. *Cell*. 2015;163:799-810.
3. Grun D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, van den Born M, van Es J, Jansen E, Clevers H, de Koning EJ and van Oudenaarden A. De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell*. 2016;19:266-277.
4. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gorham JM, Kathiriya IS, Hinson JT, Homsy J, Gray J, Pu W, Bruneau BG, Seidman JG and Seidman CE. Single-Cell Resolution of Temporal Gene Expression during Heart Development. *Dev Cell*. 2016;39:480-490.
5. Muraro MJ, Dharmadhikari G, Grun D, Groen N, Dielen T, Jansen E, van Gurp L, Engelse MA, Carlotti F, de Koning EJ and van Oudenaarden A. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst*. 2016;3:385-394 e3.
6. Li G, Xu A, Sim S, Priest JR, Tian X, Khan T, Quertermous T, Zhou B, Tsao PS, Quake SR and Wu SM. Transcriptomic Profiling Maps Anatomically Patterned Subpopulations among Single Embryonic Cardiac Cells. *Dev Cell*. 2016;39:491-507.
7. Hullinger TG, Montgomery RL, Seto AG, Dickinson BA, Semus HM, Lynch JM, Dalby CM, Robinson K, Stack C, Latimer PA, Hare JM, Olson EN and van Rooij E. Inhibition of miR-15 protects against cardiac ischemic injury. *Circ Res*. 2012;110:71-81.
8. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
9. Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.

10. Masson-Pevet M, Jongsma HJ and De Bruijne J. Collagenase- and trypsin-dissociated heart cells: a comparative ultrastructural study. *J Mol Cell Cardiol.* 1976;8:747-757.
11. Spada F, Fuoco C, Pirro S, Paoluzi S, Castagnoli L, Gargioli C and Cesareni G. Characterization by mass cytometry of different methods for the preparation of muscle mononuclear cells. *N Biotechnol.* 2016;33:514-523.
12. Liu L, Cheung TH, Charville GW and Rando TA. Isolation of skeletal muscle stem cells by fluorescence-activated cell sorting. *Nat Protoc.* 2015;10:1612-1624.
13. Etchevers H. Primary culture of chick, mouse or human neural crest cells. *Nat Protoc.* 2011;6:1568-1577.
14. Yuelling LW, Du F, Li P, Muradimova RE and Yang ZJ. Isolation of distinct cell populations from the developing cerebellum by microdissection. *J Vis Exp.* 2014;(91):52034.
15. Chorvat D, Jr., Kirchnerova J, Cagalinec M, Smolka J, Mateasik A and Chorvatova A. Spectral unmixing of flavin autofluorescence components in cardiac myocytes. *Biophys J.* 2005;89:L55-57.
16. Gao G, Johansson U, Rundquist I and Ollinger K. Lipofuscin-induced autofluorescence of living neonatal rat cardiomyocytes in culture. *Mech Ageing Dev.* 1994;73:79-86.
17. Grun D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H and van Oudenaarden A. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature.* 2015;525:251-255.
18. Frey N, Richardson JA and Olson EN. Calsarcins, a novel family of sarcomeric calcineurin-binding proteins. *Proc Natl Acad Sci U S A.* 2000;97:14632-14637.
19. Frey N, Barrientos T, Shelton JM, Frank D, Rutten H, Gehring D, Kuhn C, Lutz M, Rothermel B, Bassel-Duby R, Richardson JA, Katus HA, Hill JA and Olson EN. Mice lacking calsarcin-1 are sensitized to calcineurin signaling and show accelerated cardiomyopathy in response to pathological biomechanical stress. *Nat Med.* 2004;10:1336-1343.
20. Tarone G, Balligand JL, Bauersachs J, Clerk A, De Windt L, Heymans S, Hilfiker-Kleiner D, Hirsch E, Iaccarino G, Knoll R, Leite-Moreira AF, Lourenco AP, Mayr M, Thum T and Tocchetti CG. Targeting myocardial remodelling to develop novel therapies for heart failure: a position paper from the Working Group on Myocardial Function of the European Society of Cardiology. *Eur J Heart Fail.* 2014;16:494-508.
21. van Berlo JH, Maillet M and Molkenin JD. Signaling effectors underlying pathologic growth and remodeling of the heart. *J Clin Invest.* 2013;123:37-45.
22. Li L, Fan D, Wang C, Wang JY, Cui XB, Wu D, Zhou Y and Wu LL. Angiotensin II increases periostin expression via Ras/p38 MAPK/CREB and ERK1/2/TGF-beta1 pathways in cardiac fibroblasts. *Cardiovasc Res.* 2011;91:80-89.
23. Colston JT, de la Rosa SD, Koehler M, Gonzales K, Mestrlil R, Freeman GL, Bailey SR and Chandrasekar B. Wnt-induced secreted protein-1 is a prohypertrophic and profibrotic growth factor. *Am J Physiol Heart Circ Physiol.* 2007;293:H1839-1846.
24. Maqbool A, Spary EJ, Manfield IW, Ruhmann M, Zuliani-Alvarez L, Gamboa-Esteves FO, Porter KE, Drinkhill MJ, Midwood KS and Turner NA. Tenascin C upregulates interleukin-6 expression in human cardiac myofibroblasts via toll-like receptor 4. *World J Cardiol.* 2016;8:340-350.
25. Kong P, Christia P and Frangogiannis NG. The pathogenesis of cardiac fibrosis. *Cell Mol Life Sci.* 2014;71:549-574.

26. Conrads TP, Tocci GM, Hood BL, Zhang CO, Guo L, Koch KR, Michejda CJ, Veenstra TD and Keay SK. CKAP4/p63 is a receptor for the frizzled-8 protein-related antiproliferative factor from interstitial cystitis patients. *J Biol Chem*. 2006;281:37836-37843.
27. Chavda B, Ling J, Majernick T and Planey SL. Antiproliferative factor (APF) binds specifically to sites within the cytoskeleton-associated protein 4 (CKAP4) extracellular domain. *BMC Biochem*. 2017;18:13.
28. Chen H, Yang WW, Wen QT, Xu L and Chen M. TGF-beta induces fibroblast activation protein expression; fibroblast activation protein expression increases the proliferation, adhesion, and migration of HO-8910PM [corrected]. *Exp Mol Pathol*. 2009;87:189-194.
29. Hashimshony T and Yanai I. Revealing developmental networks by comparative transcriptomics. *Transcription*. 2010;1:154-158.
30. Bergmann O, Zdunek S, Felker A, Salehpour M, Alkass K, Bernard S, Sjostrom SL, Szewczykowska M, Jackowska T, Dos Remedios C, Malm T, Andra M, Jashari R, Nyengaard JR, Possnert G, Jovinge S, Druid H and Frisen J. Dynamics of Cell Generation and Turnover in the Human Heart. *Cell*. 2015;161:1566-1575.
31. Panther F, Williams T and Ritter O. Inhibition of the calcineurin-NFAT signalling cascade in the treatment of heart failure. *Recent Pat Cardiovasc Drug Discov*. 2009;4:180-186.
32. Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, Myers RM and Wold BJ. From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res*. 2014;24:496-510.
33. Kehat I and Molkentin JD. Molecular pathways underlying cardiac remodeling during pathophysiological stimulation. *Circulation*. 2010;122:2727-2735.
34. Kimura H, Fumoto K, Shojima K, Nojima S, Osugi Y, Tomihara H, Eguchi H, Shintani Y, Endo H, Inoue M, Doki Y, Okumura M, Morii E and Kikuchi A. CKAP4 is a Dickkopf1 receptor and is involved in tumor progression. *J Clin Invest*. 2016;126:2689-2705.

Circulation

Figure Legends

Figure 1. Sorting single cells from the adult heart. **A**, Schematic representation of the heart and all main cardiac cell types. **B**, Work flow of the protocol. **C**, Gating strategy to sort single cells based on different scatter properties. **D**, Schematic image of the heart highlighting the area selected for enzymatic digestion; images of cells before **D'**, and after sorting **D''**. **E**, Representative bioanalyzer results for RNA quality for the indicated number of sorted cells from control heart. This quality step was performed on each heart used for digestion and downstream single-cell analysis.



Figure 2. Clustering of cardiac cells based on gene expression differences. **A**, Heatmap showing distances in cell-to-cell transcriptomes of 426 cells obtained from 3 hearts. Distances are measured by $1 - \text{Pearson's correlation coefficient}$. K-medoids clustering identified a total of 14 different cell clusters depicted on the x-axis and y-axis of the heatmap. **B**, t-SNE map indicating transcriptome similarities between all individual cells. Different numbers and colors highlight different clusters identified by K-medoids clustering shown in **A**. **C**, Tables showing list of known marker genes of main cardiac cell types used to identify the subpopulations of cells identified in **A**. **D-G**, t-SNE maps indicating the expression of selected, well-established cellular markers in cell populations identified as **D**, cardiomyocytes, **E**, fibroblasts, **F**, endothelial cells and **G**, macrophages. Data are shown as normalized transcript counts on a color-coded logarithmic (\log_2) scale.

Figure 3. Gene expression differences across cardiomyocyte subpopulations. **A**, Table showing the top 25 highest expressed and enriched genes per cluster. Genes highlighted in red are known cardiomyocyte markers. **B**, t-SNE maps showing the distribution in expression of selected cardiomyocyte markers that are enriched in indicated cardiomyocyte clusters. Expression is shown as normalized transcript counts on a color-coded scale. **C**, Heatmap of average normalized number of reads across the 4 cardiomyocyte clusters and a fibroblast cluster as a control.

Figure 4. Myozenin 2 expression is restricted to a subset of cardiomyocytes. **A-B**, t-SNE maps showing expression of **A**, *Tnnt2* and **B**, *Myoz2* across cells from control hearts. Expression is shown as normalized transcript counts on a color-coded scale. **C**, Representative confocal images of control hearts stained for TNNT2 (green), MYOZ2 (red) and DAPI (blue). Immunohistochemistry was performed on 3 control hearts. **D-E**, KEGG pathway analysis of top 200 **D**, up-regulated or **E**, down-regulated genes in the *Myoz2* expressing cardiomyocyte cluster (Cluster 4) compared to all other cardiomyocyte clusters (Cluster 1, 3 and 9).

Figure 5. Single-cell sequencing of the ischemic heart. **A-B**, t-SNE map indicating transcriptome similarities between 935 individual cells. Colors in **A**, highlight the conditions of the hearts from which the cells were derived (control in green and 3 dpIR in pink). Numbers in **B**, highlight the clusters identified in Figure S6C. **C-D**, Enlargement of the t-SNE map from **A**, and **B**, focussing in on the fibroblast clusters. The dotted circle highlights clusters containing mainly/exclusively cells from the diseased hearts. **E**, Enlargement of the t-SNE map from fibroblast clusters showing higher expression of specific genes in the clusters from the diseased

hearts compared to clusters from both control and 3 dpIR hearts. Expression is depicted as normalized transcript count on a color-coded logarithmic (\log_2) scale. **F**, Validation of the increased expression of genes found up-regulated in fibroblasts from diseased compared to control hearts by qPCR on whole hearts. $n=5$ Two-sample t-test; * $p<0.05$ control vs 3 dpIR. **G**, Pie graph showing the number of significantly ($p<0.05$) up- and down-regulated genes in diseased fibroblast clusters compared to the control fibroblast cluster. Genes were selected with a \log_2 fold change of at least 1.5 or -1.5 respectively. **H**, KEGG pathway analysis and GO term enrichment on genes that were significantly up-regulated in **G**. **I**, Expression of the up-regulated 11 genes in the diseased fibroblast clusters compared to the control fibroblast cluster as identified in **G**. **J**, Heatmap showing the differential expression of the 11 up-regulated genes in the diseased fibroblast clusters in bulk sequencing of the infarct region of control and 3 dpIR hearts. **K**, t-SNE map showing the expression of *Ckap4* across all cells sequenced from control and diseased hearts. Expression is depicted as normalized transcript count on a color-coded scale. **L**, Enlargement of the t-SNE map from fibroblast clusters showing a higher expression of *Ckap4* in the diseased fibroblast clusters (Cluster 13 and 15) compared to the control fibroblast cluster (Cluster 7).

Figure 6. *Ckap4* is specifically expressed in fibroblasts post I/R, and in activated fibroblasts in vitro. **A**, Representative confocal images of control (left) and 3 dpIR (right) mouse heart stained for CKAP4 (red) and known markers for different cardiac cell types (green): endothelial cells (PECAM1 – upper panel), cardiomyocytes (ACTN2 – middle panel) and fibroblasts (VIM – lower panel). Immunohistochemistry was performed on a total of 3 hearts per condition. **B**, Schematic overview of *Ckap4* knockdown experiments in activated NIH/3T3 fibroblasts. **C**,

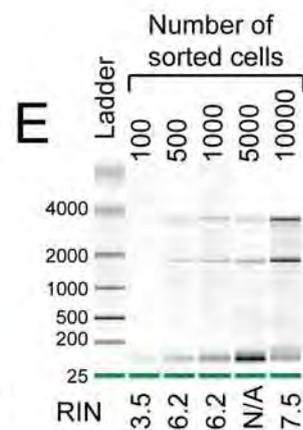
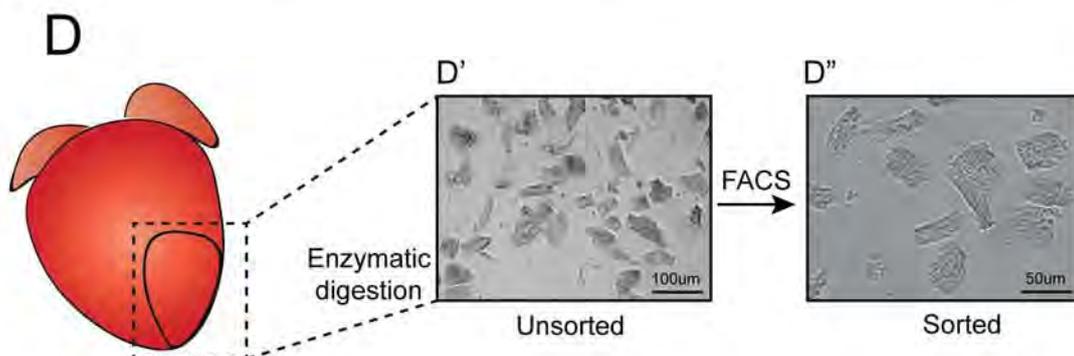
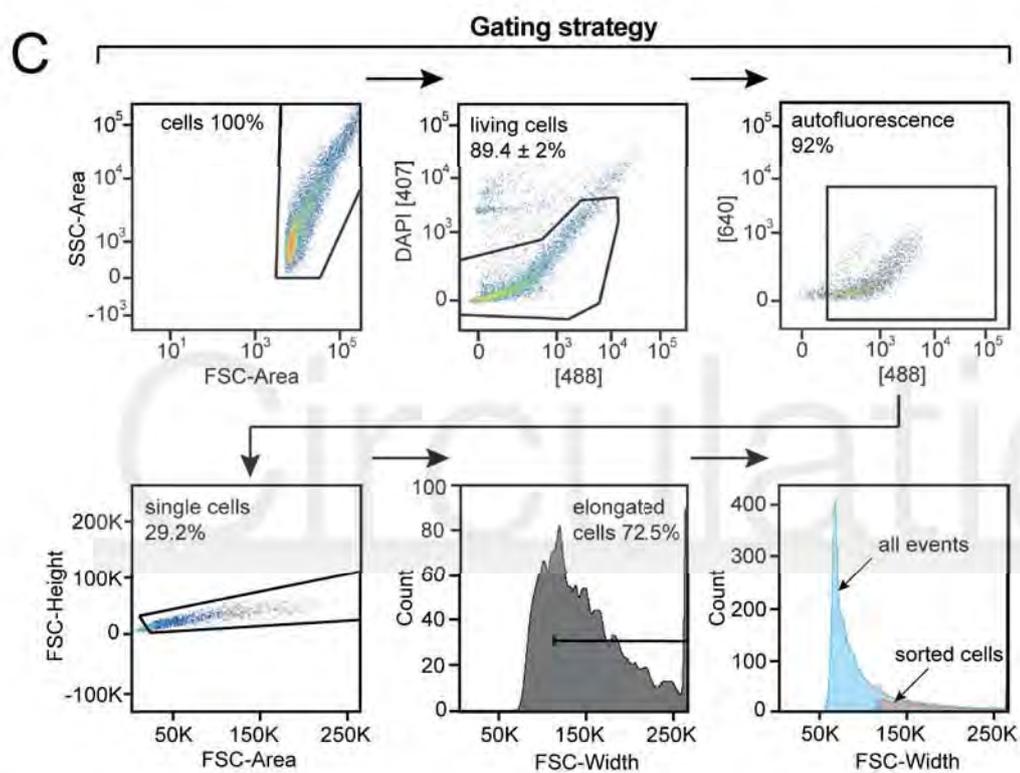
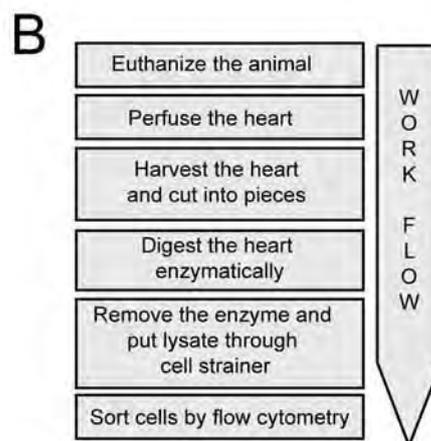
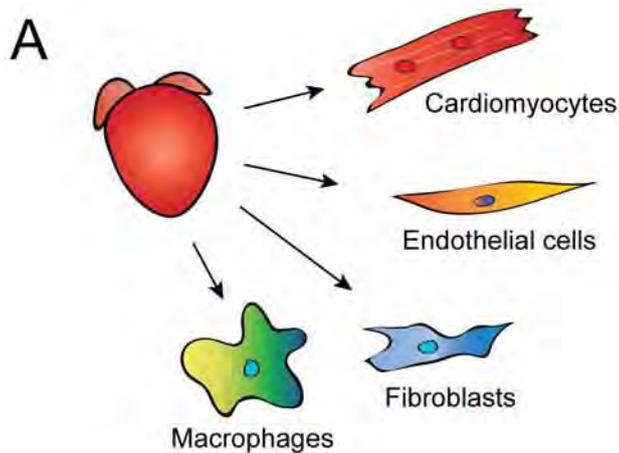
qPCR for *Ckap4* after TGF β stimulation in NIH/3T3 fibroblasts post *Ckap4* siRNA treatment or scrambled siRNA treatment. Expression levels are relative to vehicle treated NIH/3T3 cells transfected with scrambled siRNA. n=3-10; Two-sample t-test; # p<0.005 vs scrambled siRNA vehicle treated; * p<0.005 vs scrambled siRNA vehicle treated; \$ p<0.005 vs scrambled siRNA TGF β treated. **D**, Representative confocal images of NIH/3T3 cells upon vehicle or TGF β treatment without or with *Ckap4* knockdown. **E**, qPCR for marker genes of activated fibroblasts in NIH/3T3 cells upon vehicle or TGF β treatment without or with *Ckap4* knockdown. Expression levels are relative to vehicle treated NIH/3T3 cells transfected with scrambled siRNA. n=5-9; Two-sample t-test; # p<0.05 vs scrambled siRNA vehicle treated; * p<0.05 vs scrambled siRNA vehicle treated; \$ p<0.005 vs scrambled siRNA vehicle treated.



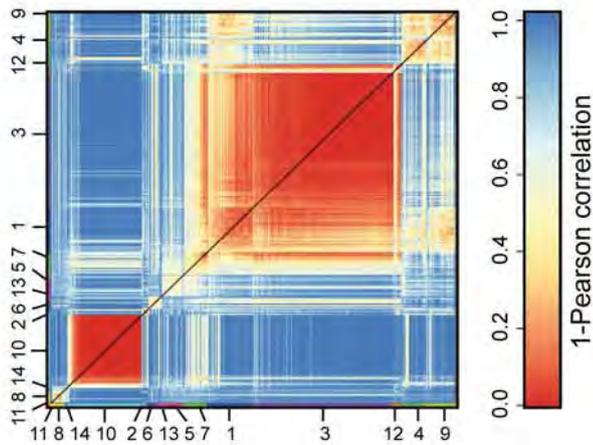
Figure 7. CKAP4 is co-expressed with markers of activated fibroblasts in human ischemic

hearts. A-C, Pearson correlation between expression of *CKAP4* and markers for activated fibroblasts **A**, *POSTN*, **B**, *CTHRC1*, and **C**, *FNI*, determined by qPCR analysis on human cardiac tissue (n=30 including control hearts and three different parts of the ischemic heart:

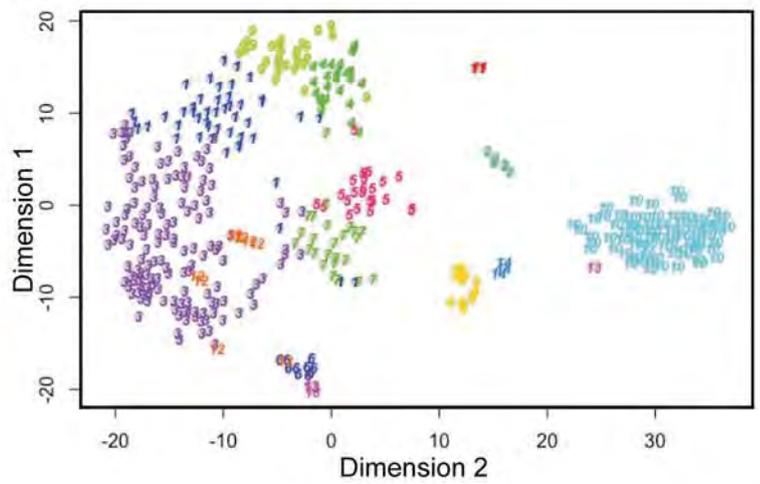
remote, border zone and infarct n=35). **D**, Representative confocal images from human ischemic hearts stained for CKAP4 (red) and markers for different cardiac cell types (green): endothelial cells (PECAM1 – upper, left panel), cardiomyocytes (ACTN2 – upper, right panel) and fibroblasts (ACTA2 – middle, left panel, DDR2 – middle, right panel, VIM – lower, left panel, PDGFR – lower, right panel). Immunohistochemistry was performed on a total of 3 hearts per condition.



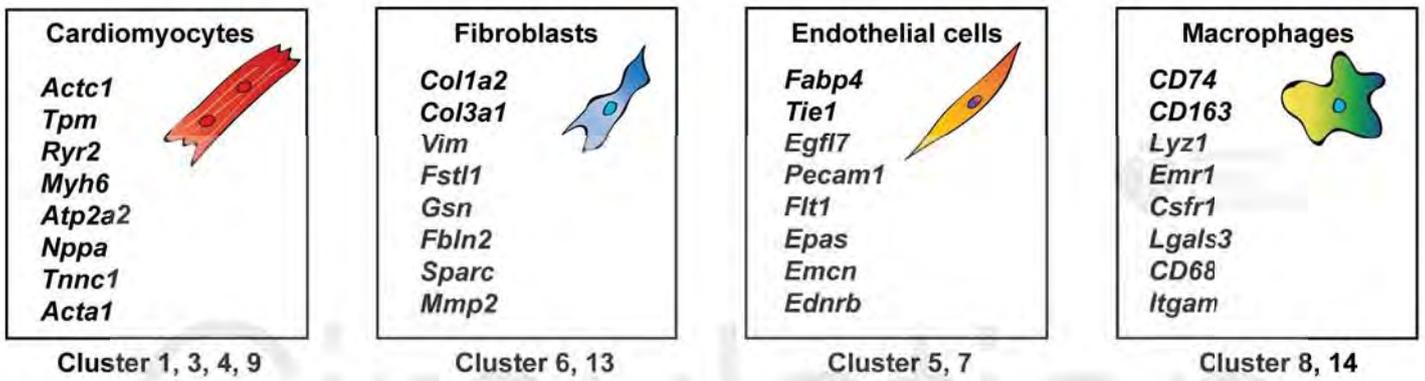
A



B

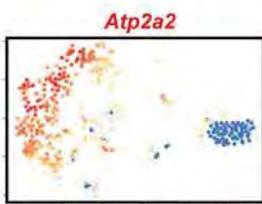
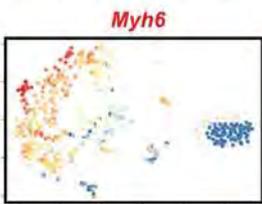
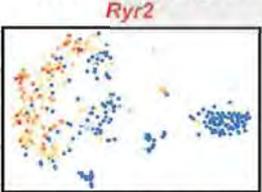


C



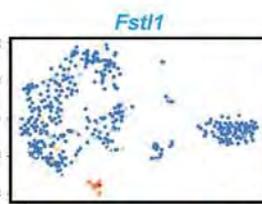
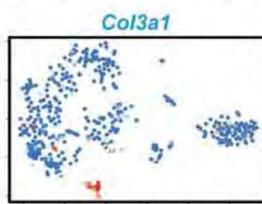
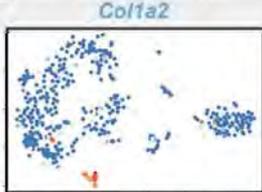
D

Cardiomyocytes



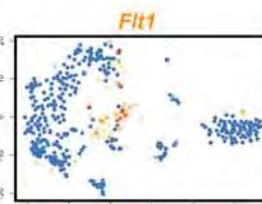
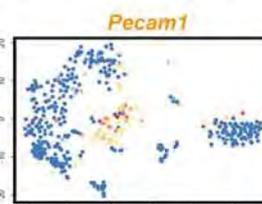
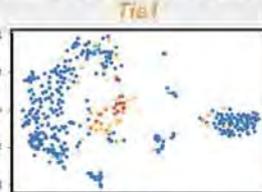
E

Fibroblasts



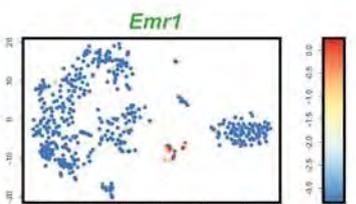
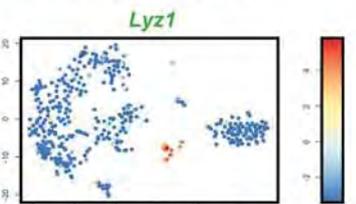
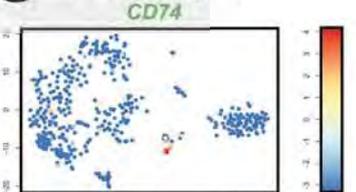
F

Endothelial cells



G

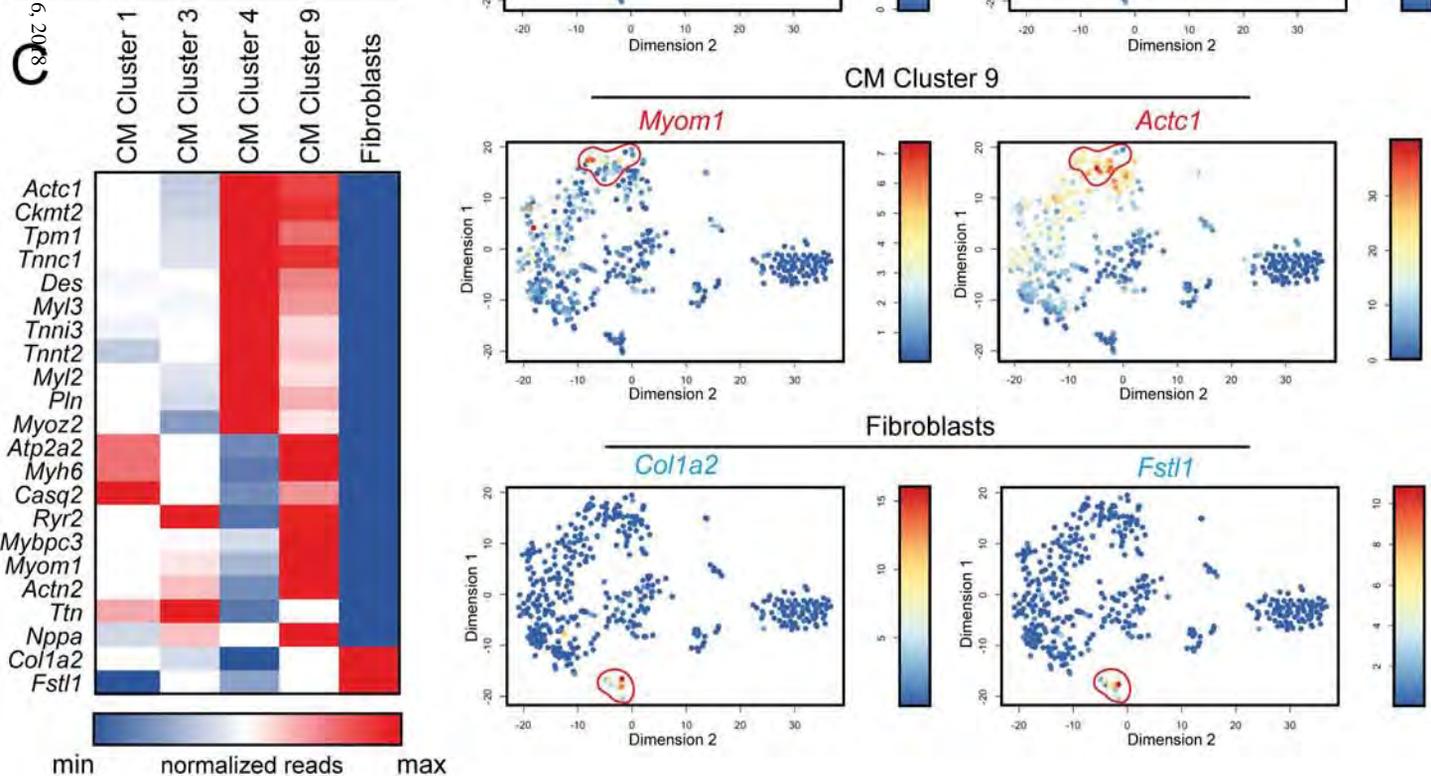
Macrophages

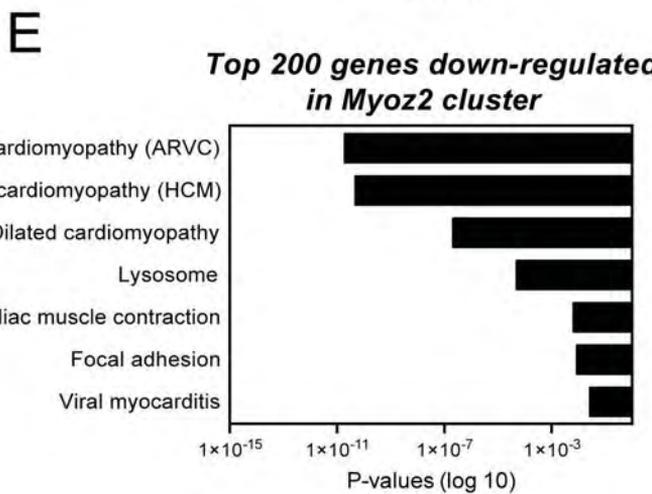
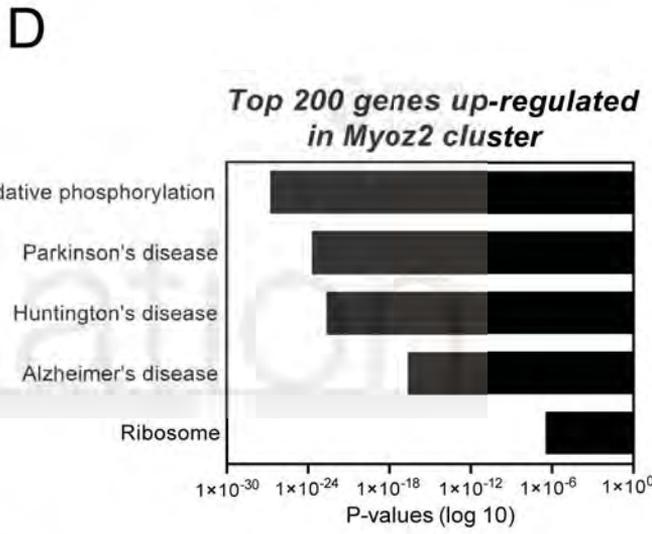
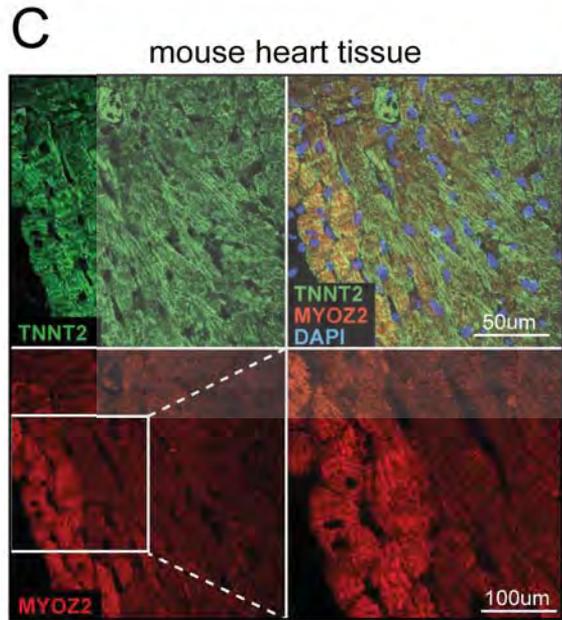
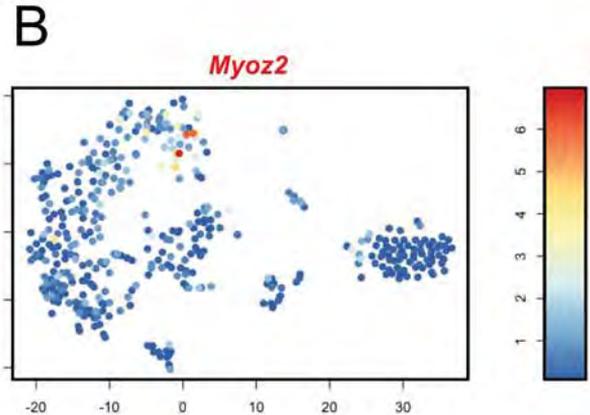
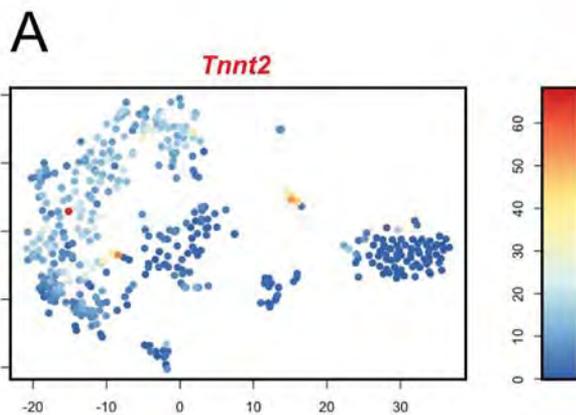


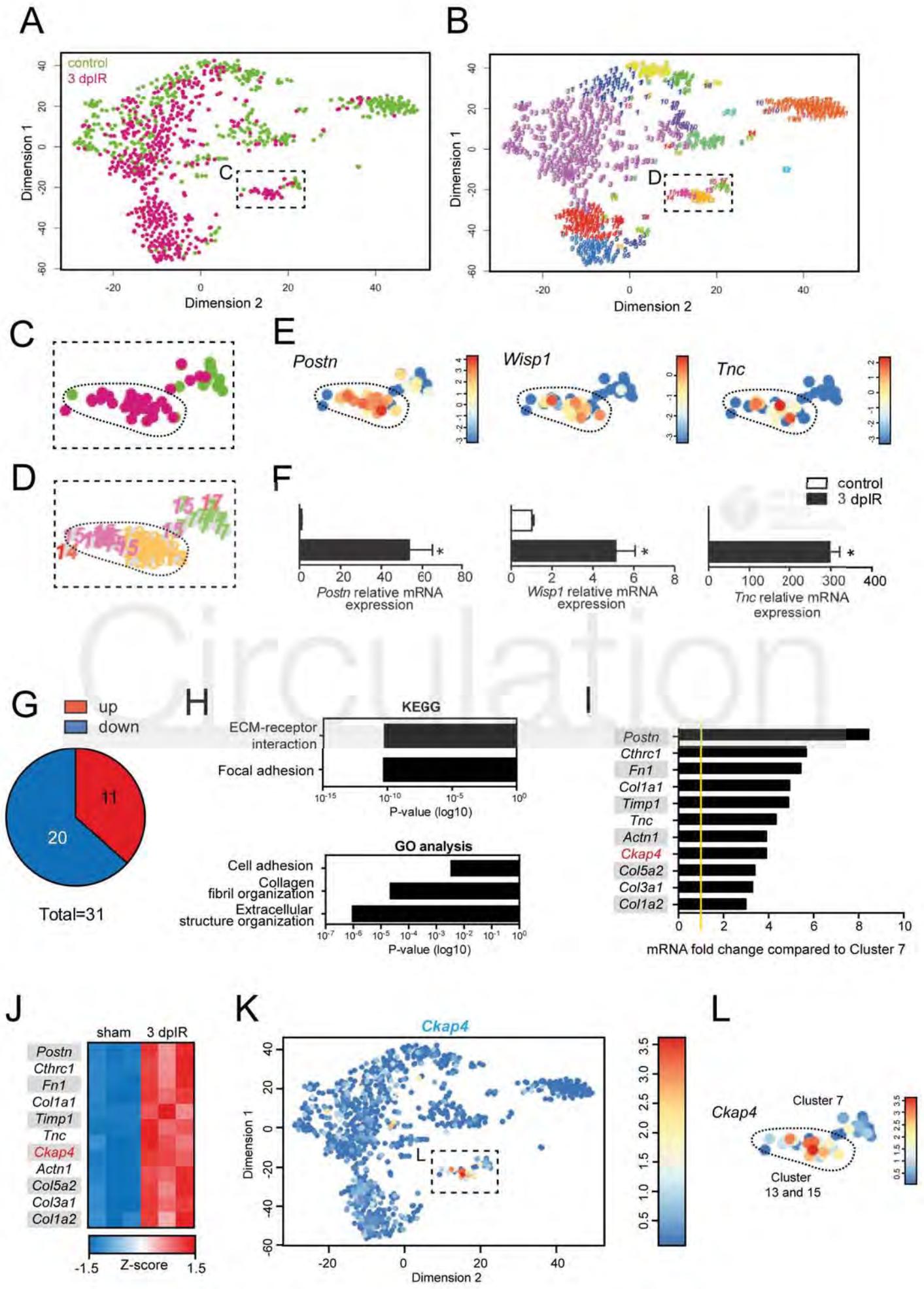
A

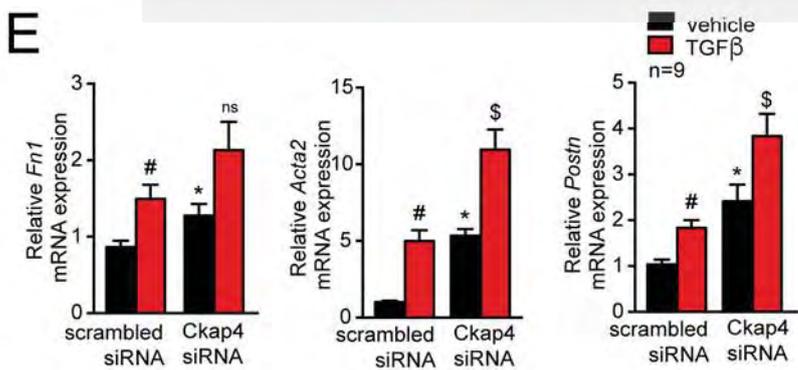
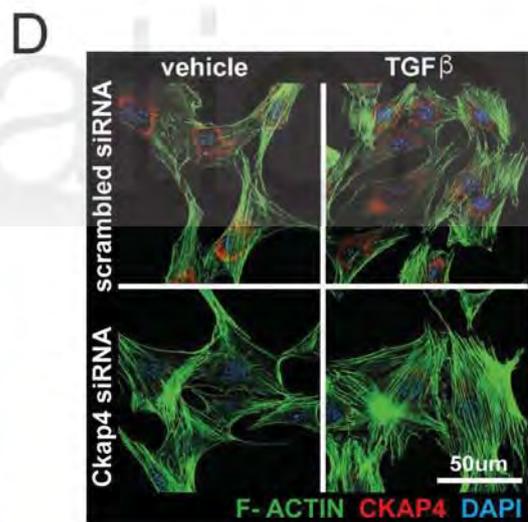
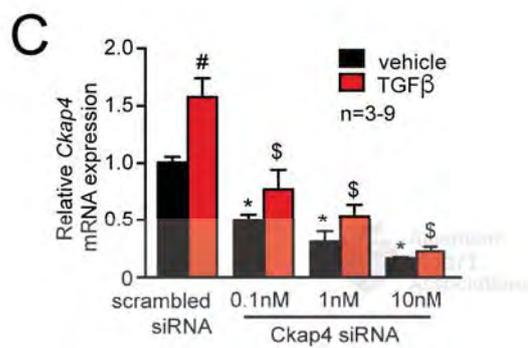
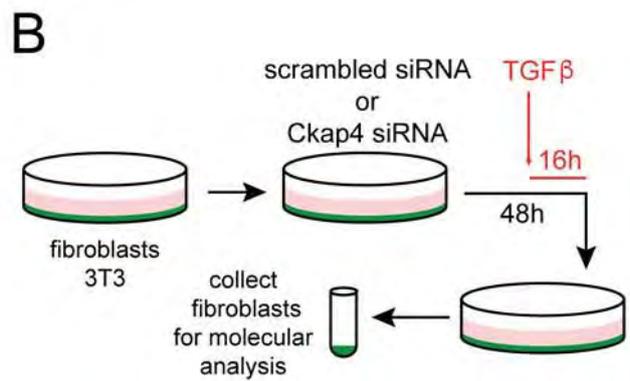
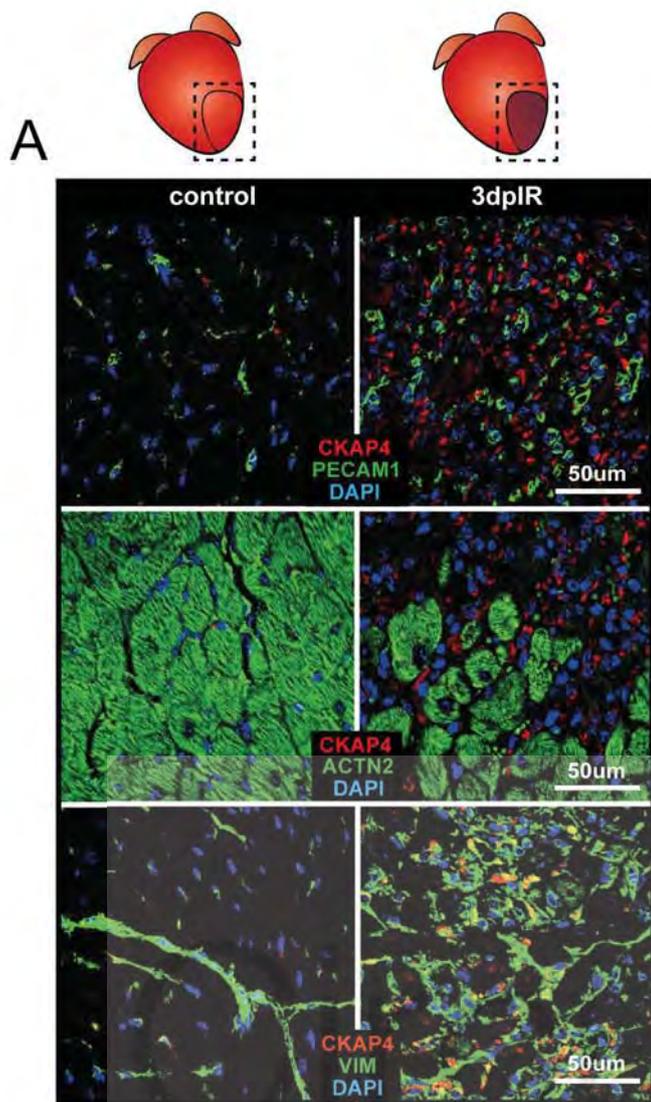
CM Cluster 1	<i>Atp2a2</i>	<i>Cryab</i>	<i>Ckmt2</i>
	<i>Myh6</i>	<i>Casq2</i>	<i>Clu</i>
	<i>Actc1</i>	<i>Cox4i1</i>	<i>Oxct1</i>
	<i>Lpl</i>	<i>Aldoa</i>	<i>Eno3</i>
	<i>Slc25a4</i>	<i>Ckm</i>	<i>Pygm</i>
	<i>Cox6a2</i>	<i>Hrc</i>	<i>Cox7b</i>
	<i>Myl3</i>	<i>Ldhb</i>	<i>Kcnk3</i>
	<i>Atp5a1</i>	<i>Tecr</i>	<i>Srl</i>
	<i>Cst3</i>		
CM Cluster 3	<i>Rn45s</i>	<i>Tnni3</i>	<i>Hspa5</i>
	<i>Atp2a2</i>	<i>Lars2</i>	<i>Ryr2</i>
	<i>Ttn</i>	<i>Bsg</i>	<i>Clu</i>
	<i>Mir5109</i>	<i>Cryab</i>	<i>Ctsd</i>
	<i>Myh6</i>	<i>Hrc</i>	<i>Tcap</i>
	<i>Tnnt2</i>	<i>Casq2</i>	<i>Oxct1</i>
	<i>Lpl</i>	<i>Mybpc3</i>	<i>Ptgds</i>
	<i>Cox6a2</i>	<i>Atp1a2</i>	<i>Uqcrc1</i>
	<i>Myl3</i>		
CM Cluster 4	<i>Actc1</i>	<i>Atp5h</i>	<i>Ybx1</i>
	<i>Myl2</i>	<i>Atp5c1</i>	<i>Cyca</i>
	<i>Tpm1</i>	<i>Des</i>	<i>Gapdh</i>
	<i>Pln</i>	<i>Eif1</i>	<i>Cox5a</i>
	<i>Tp1</i>	<i>Cox6c</i>	<i>Cox7b</i>
	<i>Fth1</i>	<i>Aes</i>	<i>Myoz2</i>
	<i>Atp5g3</i>	<i>Hspb6</i>	<i>Uqcrcq</i>
	<i>Fabp3</i>	<i>Cox5b</i>	<i>Uqcrcfs1</i>
	<i>Atp5g1</i>		
CM Cluster 9	<i>Atp2a2</i>	<i>Des</i>	<i>Ckmt2</i>
	<i>Myh6</i>	<i>Nppa</i>	<i>Cox5a</i>
	<i>Actc1</i>	<i>Ndr2</i>	<i>Uqcrc1</i>
	<i>Mb</i>	<i>Mdh1</i>	<i>Ogdh</i>
	<i>Slc25a4</i>	<i>Tcap</i>	<i>Uqcrcfs1</i>
	<i>Tpm1</i>	<i>Tnnc1</i>	<i>Myom1</i>
	<i>Atp5b</i>	<i>Slc25a3</i>	<i>Gapdh</i>
	<i>Pln</i>	<i>Ldhb</i>	<i>Acadl</i>
	<i>Fabp3</i>		

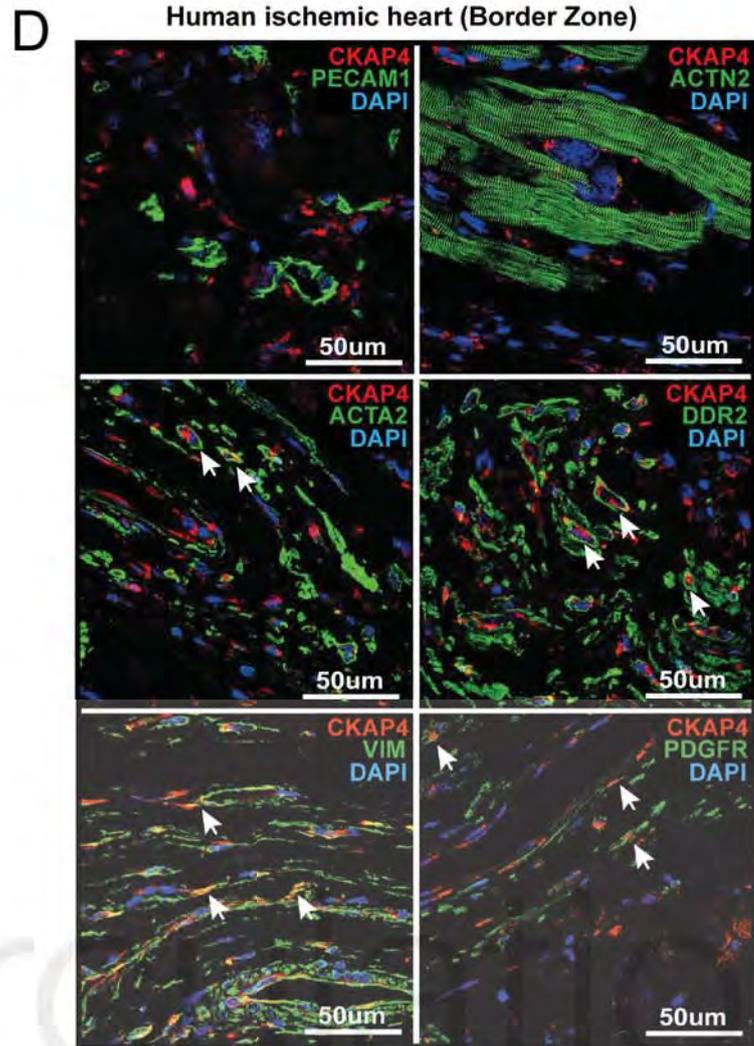
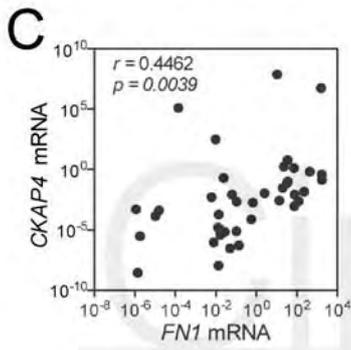
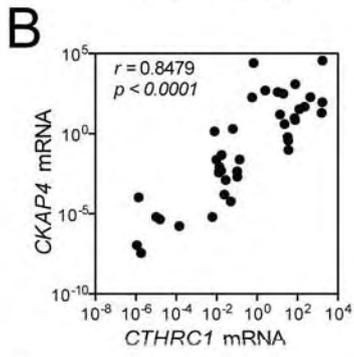
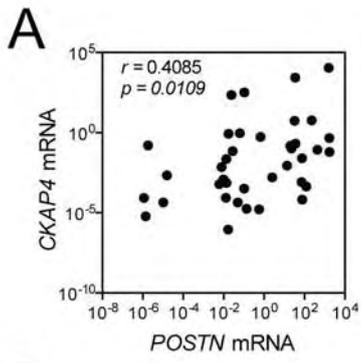
B











American
Heart
Association.

Single-Cell Sequencing of the Healthy and Diseased Heart Reveals Ckap4 as a New Modulator of Fibroblasts Activation

Monika M. Gladka, Bas Molenaar, Hesther de Ruiter, Stefan van der Elst, Hoyee Tsui, Danielle Versteeg, Grègory P. A. Lacraz, Manon M. H. Huibers, Alexander van Oudenaarden and Eva van Rooij

Circulation. published online January 31, 2018;

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Copyright © 2018 American Heart Association, Inc. All rights reserved.

Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circ.ahajournals.org/content/early/2018/01/30/CIRCULATIONAHA.117.030742>

Data Supplement (unedited) at:

<http://circ.ahajournals.org/content/suppl/2018/01/30/CIRCULATIONAHA.117.030742.DC1>

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

Reprints: Information about reprints can be found online at:
<http://www.lww.com/reprints>

Subscriptions: Information about subscribing to *Circulation* is online at:
<http://circ.ahajournals.org/subscriptions/>

SUPPLEMENTAL MATERIAL

Single-cell sequencing of the healthy and diseased heart reveals Ckap4 as a new modulator of fibroblasts activation.

Short title: Gladka Single-cell sequencing of the heart

Monika M. Gladka, PhD^{1#}, Bas Molenaar, MSc^{1#}, Hesther de Rooter, BS¹, Stefan van der Elst, BS¹, Hoyee Tsui, PhD¹, Danielle Versteeg, BS^{1,2}, Grègory P. A. Lacraz, PhD¹, Manon M.H. Huibers, PhD³, Alexander van Oudenaarden, PhD¹ and Eva van Rooij, PhD^{1,2*}

These authors contributed equally to this work

¹ Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Centre (UMC) Utrecht, The Netherlands

² Department of Cardiology, University Medical Centre, Utrecht, The Netherlands

³ Department of Pathology, University Medical Centre, Utrecht, The Netherlands

***Correspondence to:**

Eva van Rooij, PhD

Hubrecht Institute, KNAW

Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

Phone: +31 30 2121956

e.vanrooij@hubrecht.eu

Expanded Methods and Results

Animals. All animal studies were performed in accordance with institutional guidelines and regulations of the Animal Welfare Committee of the Royal Netherlands Academy of Arts and Sciences. All animal experiments were performed on adult C57BL/6J male mice from Charles River Laboratories (8-9 weeks of age) or α MHC-Cre transgenic mice¹ crossed with R26 lox-Stop-lox tdTomato (The Jackson Laboratory, #007914) to allow for cardiomyocyte-specific sorting.

Ischemia-reperfusion or sham surgery was performed in adult mice by temporary ligation (1h) of the left anterior descending artery (LAD)². In brief, mice were anaesthetized with a mixture of ketamine and xylazine by IP injection and hairs were shaved from the thorax. A tracheal tube was placed and the mouse was connected to a ventilator (UNO Microventilator UMV-03, Uno BV.) The surgical side was cleaned with iodine and 70% ethanol. Using aseptic technique with sterile instruments the skin was incised at the midline to allow access to the left third intercostal space. Pectoral muscles were retracted and the intercostal muscles cut caudal to the third rib. Wound hooks were placed to allow access to the heart. The pericardium was incised longitudinally and LAD identified. A 7.0 silk suture was placed around the LAD and a piece of 2-3 mm PE 10 tubing. Following 1h of ischemic conditions, the PE tubing was removed and the ligature cut to allow for reperfusion via the LAD. After the surgery, the mice were injected with 0.05-0.1mg/kg of buprenorphine (temgesic), after which the rib cage was closed with a 5.0 silk suture and skin closed with a wound clip. The animal was disconnected from the ventilator by removing the tracheal tube and placed on a nose cone with 100% oxygen. During the whole procedure and recovery period the animals are placed on a 38°C heating mat.

Tissue collection. Mice were euthanized and the chest was opened to expose the heart. After removing the right atrium, the heart was immediately perfused by injecting 10ml of cold

perfusion buffer into the left ventricle. After perfusion, the heart was removed and washed in cold perfusion buffer.

Perfusion buffer: 135mM NaCl, 4mM KCl, 1mM MgCl₂, 10mM HEPES, 0.33mM NaH₂PO₄, 10mM glucose, 10mM 2,3-butanedione monoxime (Sigma, St Louis, MO), 5mM taurine (Sigma), adjust pH to 7.2 at 37°C. Make a 10x stock without BDM and taurine, filter and store at 4°C. Add BDM and taurine fresh.

Table S1

Perfusion buffer		
	1x pH7,2	10x (500ml) pH7,2 (sterile H ₂ O)
NaCl	135mM	39.1g
KCl	4mM	1.5g
MgCl ₂	1mM	0.476g
HEPES	10mM	11.9g
NaH ₂ PO ₄	0.33mM	0.227g
Glucose	10mM	9g
<i>BDM</i>	<i>10mM</i>	<i>add 504mg in 500ml 1x</i>
<i>Taurine</i>	<i>5mM</i>	<i>add 312mg in 500ml 1x</i>

Digestion of the heart. After collecting the infarcted area (infarct and borderzone region) or the corresponding region of a control heart, the tissue was minced into fine pieces using a scalpel and transferred into a glass vial with 1.5ml of cold digestion buffer. Digestion buffers were prepared as described in detail below.

1x digestion solutions (Suitable for half or quarter of the heart):

The tissues were digested by either using a shaking (100rpm) 37°C waterbath or by placing the pieces of tissue in a 37°C incubator for 15min. The obtained cell suspension was gently pipetted up and down (10x) and transferred onto a 100µm cell strainer placed on top of a 50ml Falcon tube. The tissue was gently rubbed through the strainer using the plunger of a 1ml syringe, after which the strainer was rinsed with 8.5ml of DMEM to obtain a total volume of 10ml, which was centrifuged for 6 min at 4°C at 300 *xg*.

Table S2

Liberase (Roche, #5401020001)		Pancreatin (Sigma, #P3292)	
150µl	Liberase TL (final conc. 0.5mg/ml)	2mg	Pancreatin
30µl	DNase I (final conc. 20µg/ml)	4mg	Collagenase II
15µl	HEPES 1M	30µl	DNase I (final conc. 20µg/ml)
1305µl	DMEM	1470µl	DPBS
<u>Comments:</u> Dilute 1 vial in 1ml RNase free water to final concentration of 5mg/ml. Rotate 30min on roller bank at 4°C.		<u>Comments:</u> DNase I (Worthington, LK003172): Dilute 1 vial in 500µl RNase free water to a final concentration of 1mg/ml (2000U/ml)	
Collagenase II (Gibco, #17101-015)		Trypsin (Gibco, #15400-054)	
3mg	Collagenase II (final conc. 2mg/ml)	80µl	Trypsin 10x Stock
1500µl	DPBS	1420µl	DPBS
<ul style="list-style-type: none"> Filter the total volume of 1.5ml through a 0.2µM filter and keep on ice. 			

The protocol we decided to use was to digest the tissue in a waterbath instead of using an incubator. This was because visually the digested tissue appeared in better condition when using the waterbath. We also opted to use DMEM instead of DPBS because DMEM was suggested as preferred medium by the supplier of Liberase (Roche).

Imaging of the cells. After digestion of the heart, 5000 cells were sorted into a 96-well plate and imaged using EVOS Cell Imaging Systems to visualize the morphology of the cells.

RNA isolation and quality control. To isolate RNA from the collected cells we used either the *mirVana*[™] miRNA Isolation kit (Invitrogen, #AM1560) according to the manufacturer's instructions or we resuspended the pelleted cells in 1ml of Trizol reagent (Invitrogen, #15596026) for direct RNA isolation. The RNA concentration was assessed by a NanoDrop Spectrophotometer. To assess the RNA quality an aliquot of RNA was diluted to 200-5000pg/μl and put on a bioanalyzer using the Agilent RNA 6000 Pico Kit (Agilent technologies, 5067-1513) according to the manufacturer's instructions.

Flow cytometry. The freshly collected cardiac cell lysates were resuspended in 2ml DMEM-Dulbecco's Modified Eagle Medium, high glucose, GlutaMAX[™], Supplement, pyruvate (Gibco, #31966021) with DAPI 1:1000 (Invitrogen, #D3571) and stored on ice until sorting using a FACS Aria II (BD bioscience). Live single cells were sorted into 384-well plates based on multiple scatter strategies. Using forward scatter and side scatter area (FSC-Area and SSC-Area) we selected for cells that were DAPI negative and autofluorescent in 488 and 460 channels; considering that living, more complex cells containing filaments like cardiomyocytes are more autofluorescent at those wavelengths. Next, based on FSC-Height and FSC-Area we selected for single cells and excluded cell debris. Finally, by using FSC-Width we were able to pick more elongated cells. After cell sorting, 384-well plates were immediately centrifuged and frozen down at -80°C. For RNA quality control we additionally sorted 100, 500, 1000, 5000 and 10000 cells into a 96-well plate containing Trizol. The RNA was isolated and RIN was assessed using the Bioanalyzer.

For sorting tdTomato positive cardiomyocytes we employed one additional gating that allowed us to select only for red fluorescent cells.

Library preparation and sequencing of single cells. The SORT-seq procedure was applied and described previously³. Cells were sorted into 384 well plates containing 5 μ l of VaporLock oil and an aqueous solution of 100nl containing reverse transcriptase (RT) primers, spike-in RNA molecules, dinucleotide triphosphates (dNTPs) and CEL-seq primers. CEL-seq primers consisted of a 24 bp polyT sequence followed by a 6bp unique molecular identifier (UMI), a cell-specific barcode, the 5' Illumina TruSeq2 adapter and a T7 promotor sequence. Cell lysis was formed by incubation of cells for 5min at 65°C, after which cDNA libraries were obtained by dispersion of the RT enzyme and second strand mixes with the Nanodrop II liquid handling platform (GC biotech). cDNA libraries in all wells were pooled, followed by separation of the aqueous phase from the oil phase prior to *in vitro* transcription for linear amplification as performed by overnight incubation at 37°C⁴. Next, Illumina sequencing libraries were prepared using the TruSeq small RNA primers (Illumina), followed by PCR amplification for 12-15 rounds depending on the amount of RNA after *in vitro* transcription. Afterwards, libraries were sequenced paired-end at 75 bp read length with Illumina NextSeq.

Data analysis of single-cell RNA-sequencing. Paired-end reads from Illumina sequencing were mapped with BWA-ALN⁵ to the reference genome GRCm38/mm10 downloaded from the UCSC genome browser. The left read mate was used to allocate reads to the libraries and cells, whereas the right read mate was mapped to the gene models and used to quantify transcript abundance. Reads that mapped equally well to multiple loci in the reference genome were excluded. For quantification of transcript abundance, per cell-specific barcode the number of transcripts containing unique UMI's were counted for each gene in the genome. Next, these observed transcript counts were converted into expected transcript counts using Poissonian statistics as previously described⁶, taking into account the count number for each gene and a total of 4096 different UMI's. Afterwards, all read counts for mitochondrial genes were discarded due to the high abundance of transcripts coming from these genes in cardiomyocytes, which interfered with downstream clustering.

Next, the RaceID2 algorithm was used to cluster cells based on K-medoids clustering, to visualize cell clusters using t-distributed stochastic neighbour embedding (t-SNE) and to compute genes up- or down-regulated in all cells in the cluster compared to all cells not in the cluster^{7, 8}. All default parameters of the RaceID2 algorithm were used with the exceptions described below. Per cell, the number of transcripts were down sampled to a total of 1000 unique transcripts, and cells that had a number lower than 1000 unique transcripts were discarded (min.total = 1000). This resulted in a total of 423 and 509 cells sham and 3 days post I/R conditions respectively. For clustering, only genes with a down sampled read count of 3 or higher in at least 1 cell were used (min.expr = 3, min.number = 1, max.expr = inf). For identification of outlier cells in each cluster, cells were selected that expressed at least 4 genes different from their mean expression across all cells in the cluster by a probability threshold lower than 1×10^{-6} (outlg = 4, probthr = 1e-6). Afterwards, quality of the clustering was assessed by visual inspection of the heat map of 1-pearson correlation (if the cluster of cells in the heat map show a good separation from all the other cells not in the cluster). Additionally, an additional dimension reduction algorithm, t-Distributed Stochastic Neighbor Embedding (t-SNE), was used to inspect if all cells that cluster together with K-medoid clustering also have a smaller distance between each other in a t-SNE plot compared to all other cells. Differential expression of genes between two subgroups of cells was calculated using the DESeq2 package in the R platform.⁹

siRNA transfection and TGF β treatment. For *Ckap4* knock-down we used Trilencer 27 (Origene, #SR417628). siRNA transfection was performed in NIH/3T3 cells at increasing concentrations of 0,1nM, 1nM and 10nM using Lipofectamine 2000 Transfection Reagent (Invitrogen, #11668027). TGF β (Recombinant Human TGF β 1, Hek293 derived; Preprotech #100-21) stimulation was performed at 5ng/ml for the final 16h. Cells were harvested after 48h and used for RNA isolation and immunofluorescence.

Histological analysis and immunofluorescence. Adult hearts were excised from euthanized mice, washed in cold PBS and fixed with 4% formalin at room temperature for 48h, prior to embedding in paraffin and subsequently sectioned at 4 μ m. Paraffin sections from mouse hearts as well as human ischemic hearts were used for immunofluorescence. After deparaffinization, rehydration, heat induced antigen retrieval and blocking with 0.05% BSA, sections were incubated with specific primary antibodies overnight at 4°C. After washing with PBS, the sections were incubated with secondary antibodies for 1h at RT followed by incubation with DAPI 1:5000 (Invitrogen, #D3571) for 10min at RT. The sections were washed and sealed with a mounting medium ProLongTM Gold Antifade Mountant (Invitrogen, #P36934). Images were taken using the Leica TCS SPE confocal microscope.

Antibodies used in this study include mouse anti-Troponin T (TNNT2, Abcam, #ab8295), mouse anti- α actinin (ACTN2, Sigma, #A7732) and rabbit anti-Myozenin 2 (MYOZ2, Novisbio, #NBP1-88259) to visualise cardiomyocytes. Rabbit anti-Cytoskeleton-associated protein 4 (CKAP4, Proteintech, #16686-1-AP), mouse anti-Vimentin (VIM, Santa Cruz, #sc-373717), mouse anti-Smooth muscle actin (ACTA2, Sigma-Aldrich, #A5228), mouse anti-Discoidin domain receptor tyrosine kinase 2 (DDR2, Santa Cruz, #sc-81707), mouse anti-Platelet-derived growth factor receptor (PDGFR, Santa Cruz, #sc-398206) to visualise activated fibroblasts. Goat anti-CD31 (PECAM1, Abcam, #ab56299) antibody to visualize endothelial cells. We used corresponding secondary fluorescent antibody Alexa FluorTM 488 donkey anti-mouse IgG (H+L) (Invitrogen, #A21202), Alexa FluorTM 568 donkey anti-rabbit IgG (H+L) (Invitrogen, #A10042), Alexa FluorTM 488 donkey anti-goat IgG (H+L) (Invitrogen, #A11055) Immunofluorescent staining was also performed on NIH/3T3 cells treated with control siRNA or Ckap4 siRNA. Cells were fixed with 4%PFA, quenched with NH₄Cl, permeabilized, blocked with 1% fish gelatin (Gelatin from cold water fish skin, Sigma-Aldrich, #G7765) and incubated with rabbit anti-CKAP4 antibody (Proteintech #16686-1-AP) for 25min at RT prior to incubation with the corresponding secondary Alexa FluorTM 568 donkey anti-rabbit IgG (H+L) (Invitrogen, #A10042) together with Alexa FluorTM 488 Phalloidin (Invitrogen, #A12379), to stain for F-actin, for 20min at RT. The cells were washed and sealed with a mounting medium (ProLongTM

Gold Antifade Mountant with DAPI, Invitrogen, #P36935). Images were taken using the Leica TCS SPE confocal microscope.

Quantitative PCR. Total RNA was isolated from heart ventricles or NIH/3T3 cells with Trizol reagent according to the manufacturer's instructions. Total RNA (1µg) was used for mRNA based reverse transcription using an iScript™ cDNA Synthesis Kit (Bio-Rad #1708891). qPCR was performed according to the SYBRgreen based methodology using iQ™ SYBR Green Supermix (Bio-Rad #170-8885). Transcript quantities were normalized for endogenous loading.

Table S3

qPCR primers used:

<i>Gene</i>	Sequences
<i>Postn - mouse</i>	Fw 5'- tctgggcaccaaaaagaaatac, Rv 5'- tccttccatttcatatagcc
<i>Wisp1 - mouse</i>	Fw 5'- aggtacgcaataggagtgtg, Rv 5'- agttgtacctgcagttgggtg
<i>Tnc - mouse</i>	Fw 5'- tatctggtgctgaacggactg, Rv 5'- ccggttcagcttctgtgtag
<i>Ckap4 - mouse</i>	Fw 5'- caggacttctccgctcagag, Rv 5'- tcctcacggctttctctgt
<i>Fn1 - mouse</i>	Fw 5'- ccactgtggagtacgtggtag, Rv 5'- aagcaattttgatggaatcgac
<i>Acta2 - mouse</i>	Fw 5'- ttctataacgagcttctgtg, Rv 5'- gagtccagcacaataaccagttg
<i>CKAP4 - human</i>	Fw 5'- cagccaccaggacttctcc, Rv 5'- ttgggagcttctcaagatgg
<i>POSTN - human</i>	Fw 5'- tgccctcaacagattttgg, Rv 5'- gcagcctttcattcctcc
<i>CTHRC1 - human</i>	Fw 5'- ccaaggggaagcaaaaagg, Rv 5'- cccttgtaagcacattccatta
<i>FN1 - human</i>	Fw 5'- ttcacgtctgtcactccaca, Rv 5'- tctcattcaacaagaaaccactg

Pathway analysis and gene ontology

To investigate whether genes share a similar biological function, genes that were differentially expressed between subgroups of cells with at least a log2 fold change of 1.5 up-regulated or

-1.5 down-regulated and a p value <0.05 were selected; after adjustment for multiple hypothesis testing using the Benjamini-Hochberg method. Overrepresentation of these genes was tested in the gene ontology (GO) biological process database and in the Kyoto Encyclopedia of Genes and Genomes (KEGG) using DAVID¹⁰. Significant enrichment of genes in GO and KEGG terms are shown; p values corrected for multiple testing using the Benjamini-Hochberg method.

Bulk sequencing and data analysis

For bulk sequencing, RNA of the infarct zone from left ventricle was isolated using Trizol. Afterwards, libraries were prepared and sequenced using a similar protocol as described for single-cell RNA sequencing. For quantification of transcript abundance, the same strategy was used as for single-cell sequencing, with the exception that UMIs were ignored and gene read counts were determined by the number of reads uniquely mapped to each gene. Next, these counts were divided by the total amount of mapped readcounts and multiplied by one million to obtain the read-per-million count (RPM). RPM counts were used rather than reads-per-kilobase-per-million (RPKM) due to polyA-based method for library prep, thereby allowing for sequencing of only 1 fragment containing the poly-A region per transcript. These counts were subsequently used to calculate the Z-score for genes per sample which are shown in a color-coded scale in a heatmap.

Human heart samples

Approval for studies on human tissue samples was obtained from the Medical Ethics Committee of the University Medical Center Utrecht, The Netherlands (12#387). Written informed consent was obtained or in certain cases waived by the ethics committee when obtaining informed consent was not possible due to death of the patient. In this study, we included tissue from the left ventricular free wall of patients with end-stage heart failure secondary to ischemic heart disease. The end-stage heart failure tissue was obtained during heart transplantation or at autopsy. For immunofluorescence, samples from three patients

were included from which the border zone of the infarcted hearts was used to verify localization of CKAP4 in activated fibroblasts. qPCR analysis for *Ckap4* and myofibroblast markers were done on 30 tissue samples from the left ventricular free wall of patients with ischemic heart disease (ischemic region, border zone and remote region) and 5 samples coming from left ventricular free wall of non-failing donor hearts. Gene expression values obtained by qPCR were plotted for correlation analysis.

Statistical analysis (qPCR). The number of samples (n) used in each experiment is indicated in the legend or shown in the figures. The results are presented as mean \pm standard error of the mean. For qPCR analysis, statistical analyses were performed using PRISM (GraphPad Software Inc.). Student's *t*-test and two-way ANOVA with Bonferroni's multiple comparison tests were used to test for statistical significance. For the analysis of human ischemic samples the Pearson correlation (*r*) between 2 genes was used.

References

1. Palermo J, Gulick J, Colbert M, Fewell J and Robbins J. Transgenic remodeling of the contractile apparatus in the mammalian heart. *Circ Res.* 1996;78:504-509.
2. Hullinger TG, Montgomery RL, Seto AG, Dickinson BA, Semus HM, Lynch JM, Dalby CM, Robinson K, Stack C, Latimer PA, Hare JM, Olson EN and van Rooij E. Inhibition of miR-15 protects against cardiac ischemic injury. *Circ Res.* 2012;110:71-81.
3. Muraro MJ, Dharmadhikari G, Grun D, Groen N, Dielen T, Jansen E, van Gorp L, Engelse MA, Carlotti F, de Koning EJ and van Oudenaarden A. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* 2016;3:385-394 e3.
4. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Dor Y, Regev A and Yanai I. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 2016;17:77.
5. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754-1760.
6. Grun D, Kester L and van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods.* 2014;11:637-640.
7. Grun D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, van den Born M, van Es J, Jansen E, Clevers H, de Koning EJ and van Oudenaarden A. De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell.* 2016;19:266-277.
8. Grun D and van Oudenaarden A. Design and Analysis of Single-Cell Sequencing Experiments. *Cell.* 2015;163:799-810.
9. Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
10. Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4:44-57.

Figure S1

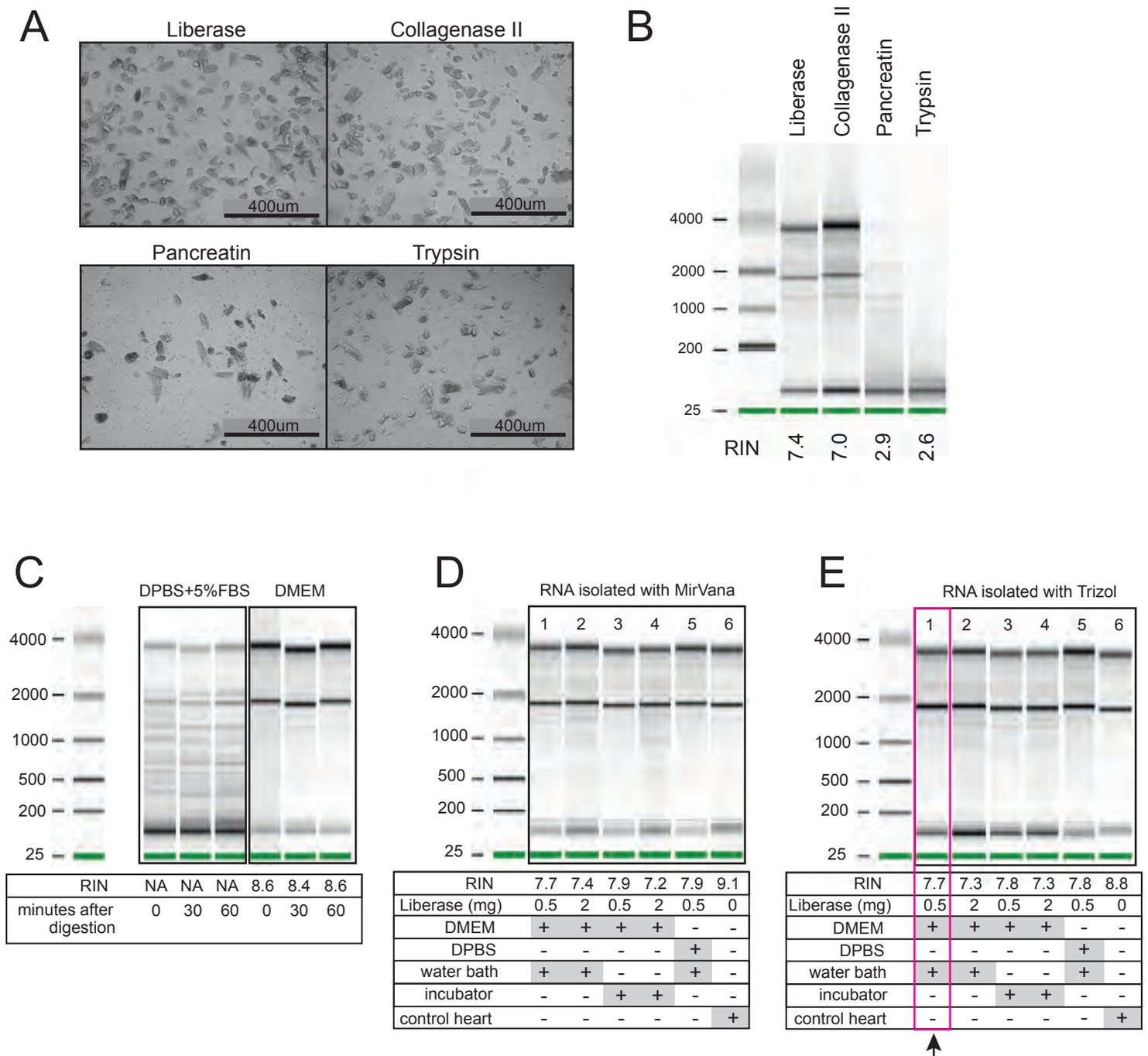


Figure S1. RNA quality control after digesting adult cardiac tissue. **A**, Representative images of cells after digestion with four different enzymatic solutions. **B**, Bioanalyzer-based quality control of RNA integrity from conditions tested in experiment shown in **A**. **C**, Bioanalyzer-based quality control of RNA integrity testing 2 different buffers and the influence of incubation time. **D**, Bioanalyzer plots from RNA isolated with MirVana kit. **E**, Bioanalyzer plots from RNA isolated with Trizol, red lined box highlights the condition chosen for downstream single-cell RNA sequencing analysis.

Figure S2

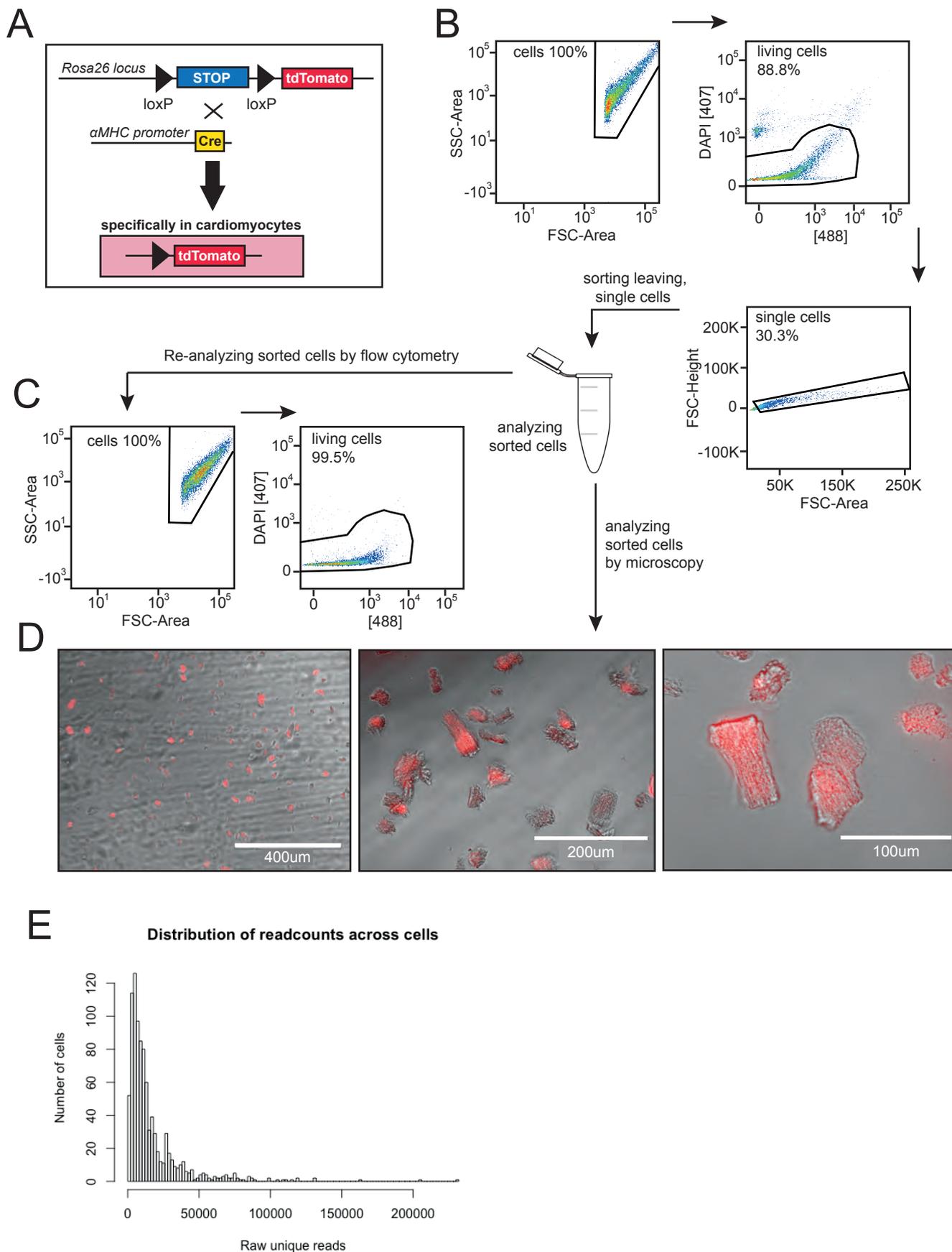


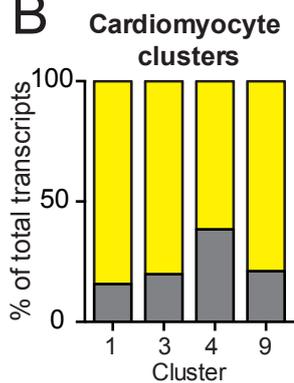
Figure S2. Viability of sorted cells. **A**, Schematic overview of breeding strategy to generate mice with tdTomato labelled cardiomyocytes. **B**, Gating strategy to sort living and single cells as shown in details in Figure 1C. **C**, Re-analyzing sorted living, single cells by flow cytometry based on DAPI exclusion. **D**, Wide-field immunofluorescence images of all sorted cells as shown in B, showing tdTomato positive cardiomyocytes. **E**, Distribution of total number of unique reads in all cells included in the analysis.

Figure S3

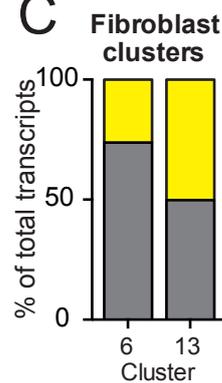
A

Cluster 1	Cardiomyocytes
Cluster 2	Unknown
Cluster 3	Cardiomyocytes
Cluster 4	Cardiomyocytes
Cluster 5	Endothelial cells
Cluster 6	Fibroblasts
Cluster 7	Endothelial cells
Cluster 8	Macrophages/Immune cells
Cluster 9	Cardiomyocytes
Cluster 10	Unknown
Cluster 11	Erythrocytes
Cluster 12	Smooth muscle cells
Cluster 13	Fibroblasts
Cluster 14	Macrophages

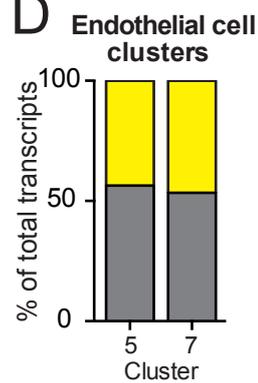
B



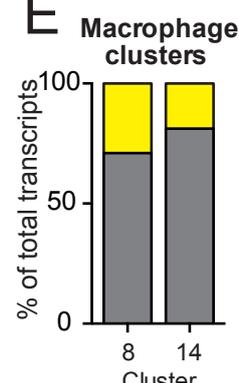
C



D



E



Mitochondrial transcripts
 Genomic transcripts

Figure S3. All cell types identified by single-cell sequencing of the control heart. A, Table highlighting cell identity of each cluster after K-medoids clustering. **B,** Percentage of expression of all genomic (grey) and all mitochondrial (yellow) genes of **B,** cardiomyocytes, **C,** fibroblasts, **D,** endothelial cells and **E,** macrophages.

Figure S4

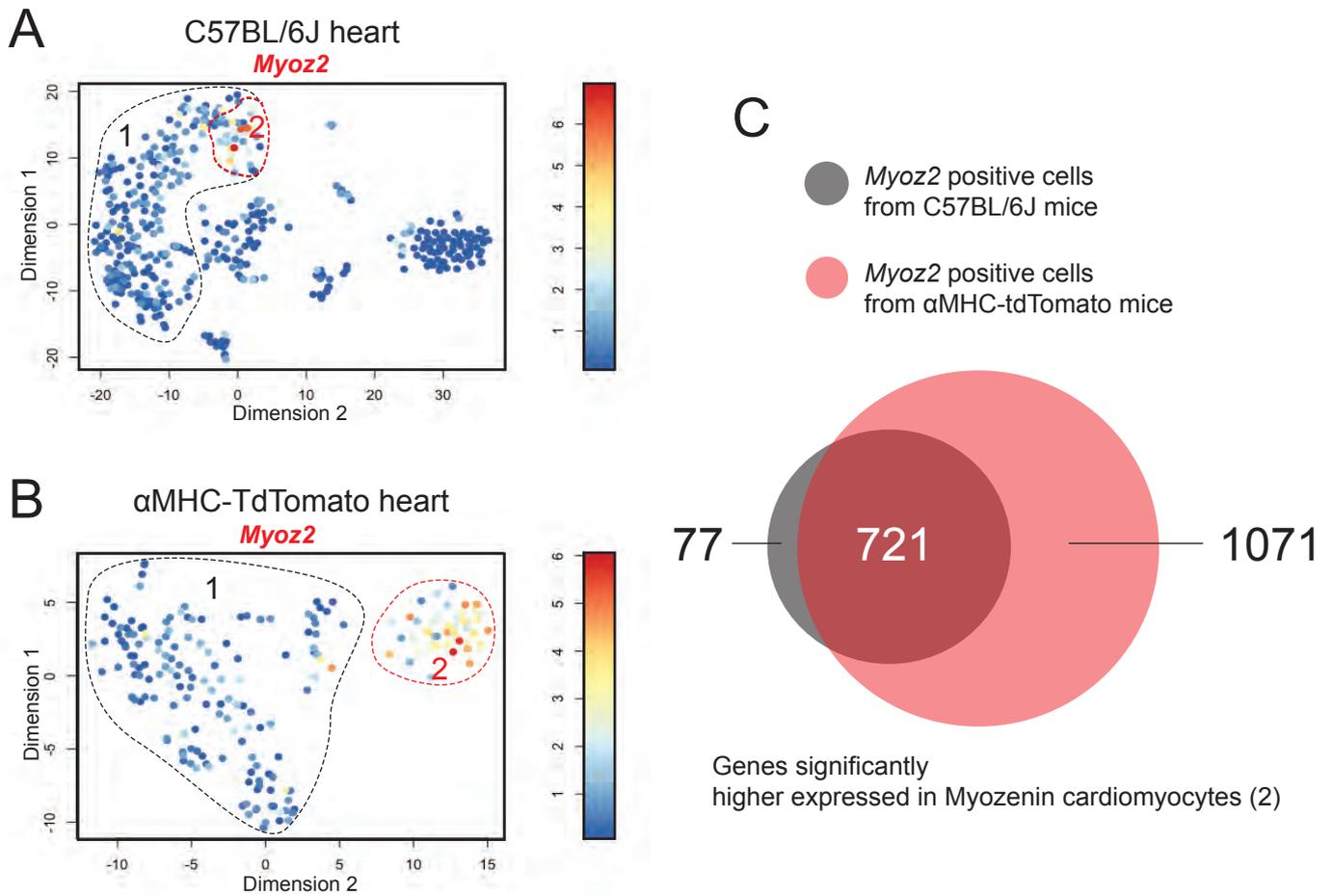


Figure S4. *Myoz2* is expressed in a subset of tdTomato positive cardiomyocytes. A-B, t-SNE maps of all cardiac cells **A**, or after sorting tdTomato positive cardiomyocytes **B**, indicating the *Myoz2*-expressing cells. Normalized readcount of *Myoz2* in all cells is depicted as a color-coded scale. The black dotted circle (1) depicts all cardiomyocyte clusters that have low *Myoz2* expression, the red dotted circle (2) depicts the cardiomyocyte cluster that has high *Myoz2* expression. **C**, Venn-diagram showing overlap between significantly higher expressed genes in (2) versus (1) if all cells are sorted or if only tdTomato positive cardiomyocytes are sorted.

Figure S5

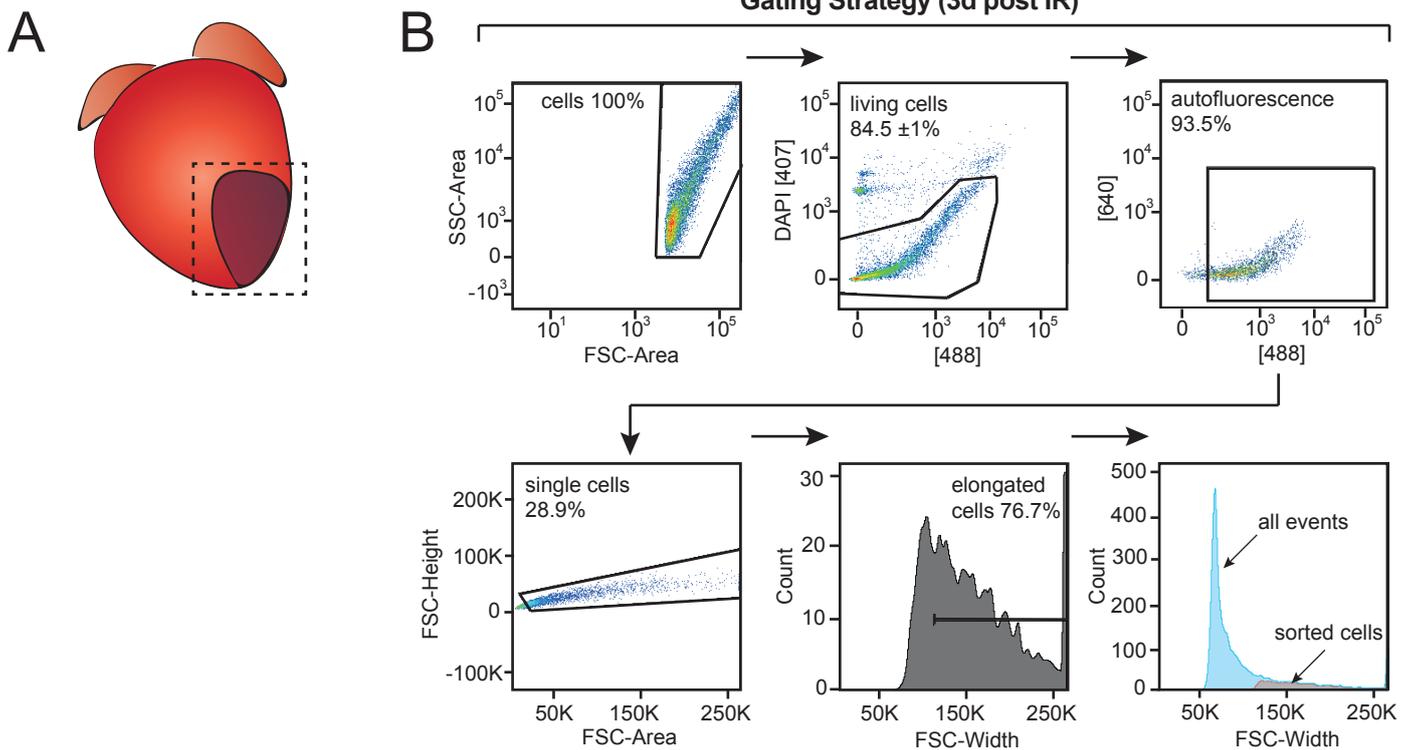


Figure S5. Sorting strategy to obtain single cells of an adult heart 3 dpIR. **A**, Schematic representation of a heart 3 days post ischemia/reperfusion injury (3 dpIR), dotted box indicates the area that was selected for enzymatic digestion. **B**, Gating strategy for sorting the cells for single-cell sequencing by flow cytometry based on different scatter properties and DAPI exclusion.

Figure S6

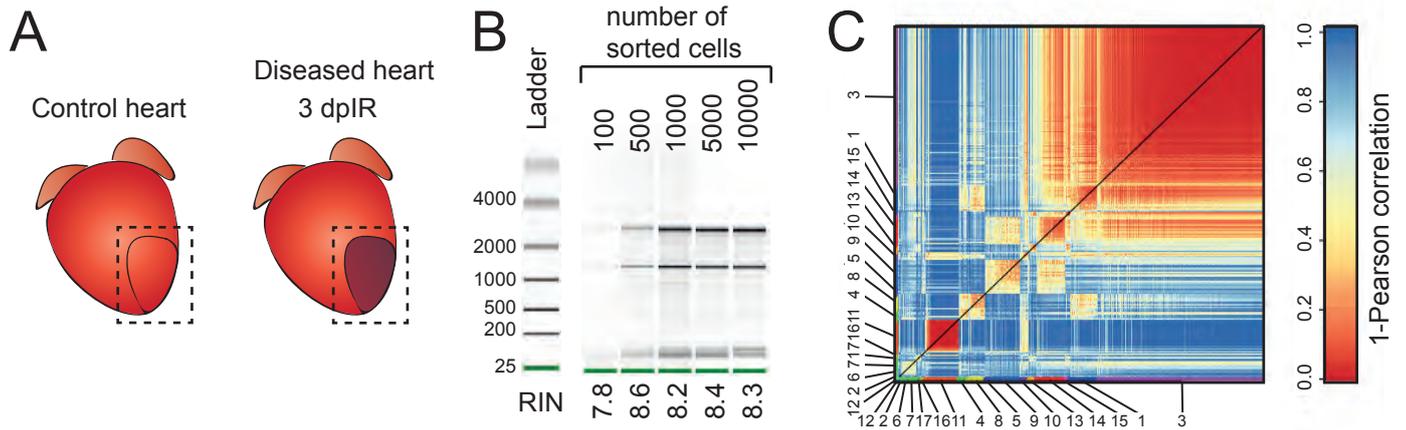


Figure S6. Clustering of cell populations in the healthy and injured heart. **A**, Schematic representation of control and diseased hearts 3 dpIR, dotted box highlights the regions taken for enzymatic digestion and single-cell analysis. **B**, Representative bioanalyzer results for RNA quality of the indicated number of sorted cells from 3 dpIR heart. This quality step was performed on each heart used for digestion and downstream single-cell analysis. **C**, Heatmap showing distances in cell-to-cell transcriptomes between cells from both conditions (control and 3 dpIR), measured by $1 - \text{Pearson's correlation coefficient}$. K-medoid clustering identified a total of 17 clusters, highlighted left and below the heatmap. Cells from control hearts were the same as used in Figure 2. For 3 dpIR, a total of 509 cells were sequenced from a total of 3 hearts.

Figure S7

A

Cluster 1	Cardiomyocytes
Cluster 2	Smooth muscle + other cells
Cluster 3	Cardiomyocytes
Cluster 4	Cardiomyocytes
Cluster 5	Macrophages/Immune cells
Cluster 6	Endothelial cells
Cluster 7	Fibroblasts
Cluster 8	Cardiomyocytes
Cluster 9	Macrophages/Immune cells
Cluster 10	Endothelial cells
Cluster 11	Unknown
Cluster 12	Erythrocytes
Cluster 13	Fibroblasts
Cluster 14	Macrophages/Immune cells
Cluster 15	Fibroblasts
Cluster 16	Macrophages/Immune cells
Cluster 17	Unknown

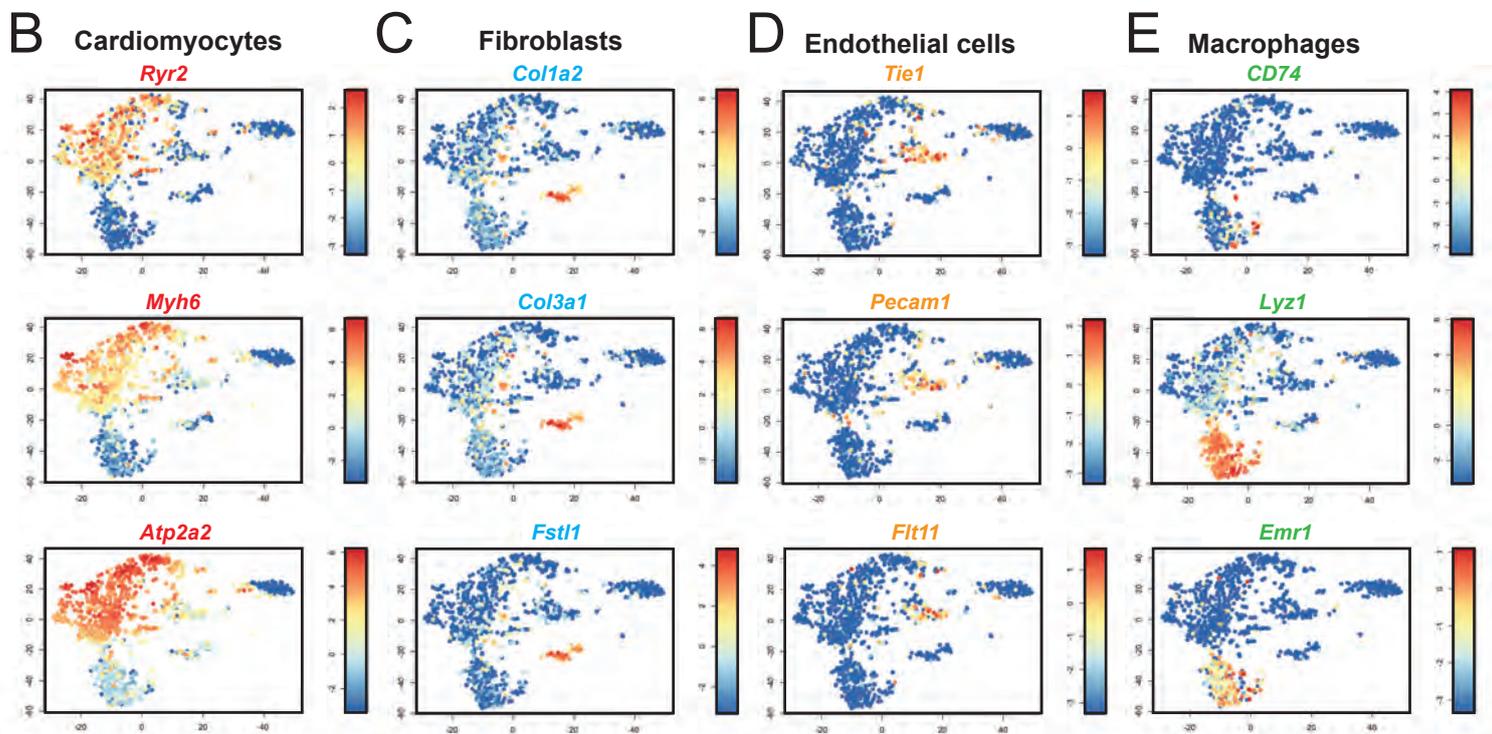


Figure S7. Expression pattern of established cellular markers in cardiac cells from control and 3 dpiR hearts. **A**, Cell clusters identified in pooled cells from control and 3 dpiR hearts. **B-E**, t-SNE maps similar as in Figure 5 A-B highlighting the expression of well-established cellular markers in clusters identified as **B**, cardiomyocytes, **C**, fibroblasts, **D**, endothelial cells and **E**, macrophages. Data shown as normalized transcript count on a color-coded logarithmic (\log_2) scale.

Figure S8

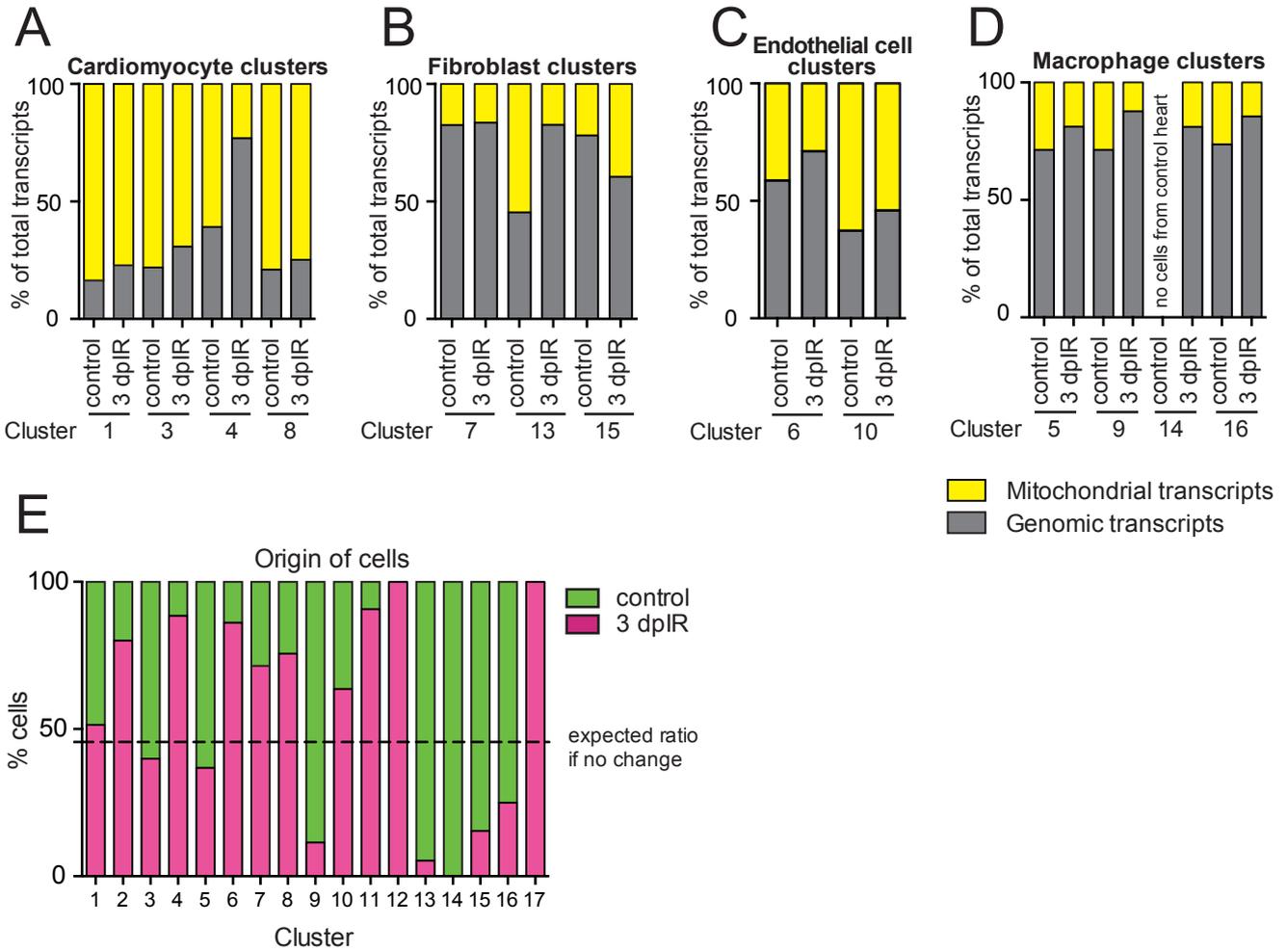


Figure S8. Mitochondrial versus genomic gene expression and origin of cells from control or 3 dpIR hearts per cluster. **A-D**, Percentage of expression of all genomic (grey) and all mitochondrial (yellow) genes of **A**, cardiomyocytes, **B**, fibroblasts, **C**, endothelial cells and **D**, macrophages. **E**, Bargraph of the relative contribution of cells from control and diseased conditions (3 dpIR) for each cluster. The dashed line indicates the expected contribution if the number of cells from both conditions would contribute equally to the clusters.

Figure S9

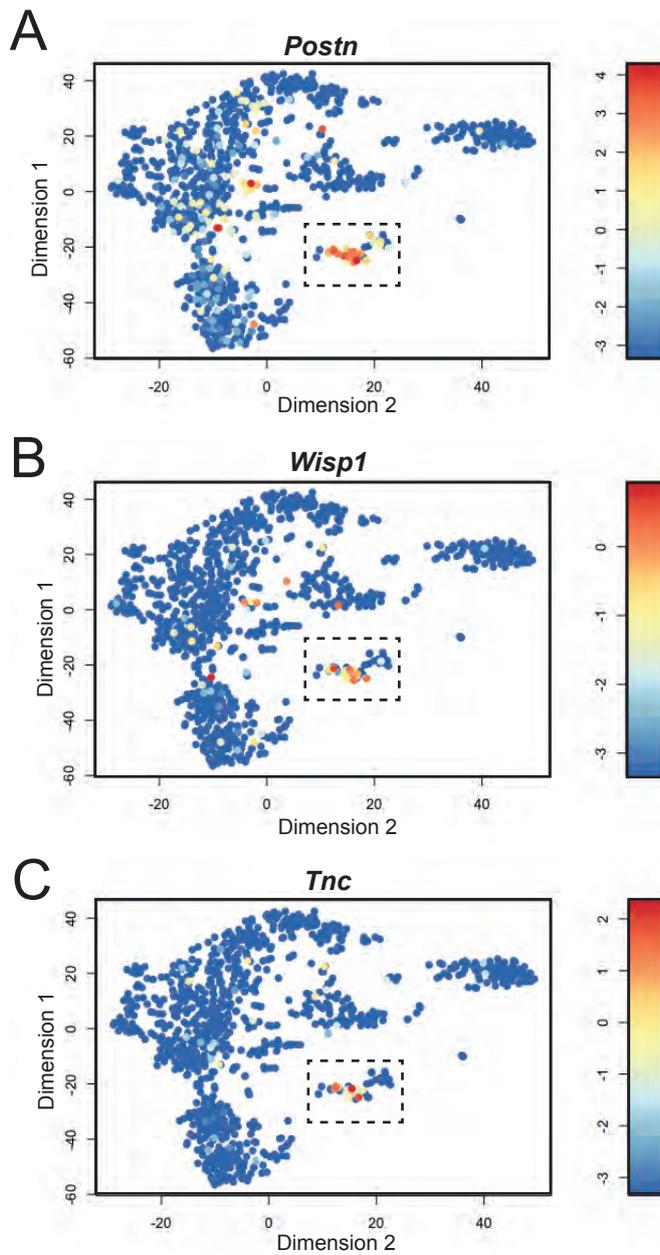


Figure S9. Expression of disease-specific fibroblast genes. A-C, t-SNE maps highlighting the expression of marker genes for activated fibroblasts in all cells from control and 3dpIR hearts. Data shown as normalized transcript count on a color-coded logarithmic (log₂) scale. Dotted box indicates cells identified as fibroblasts.

Figure S10

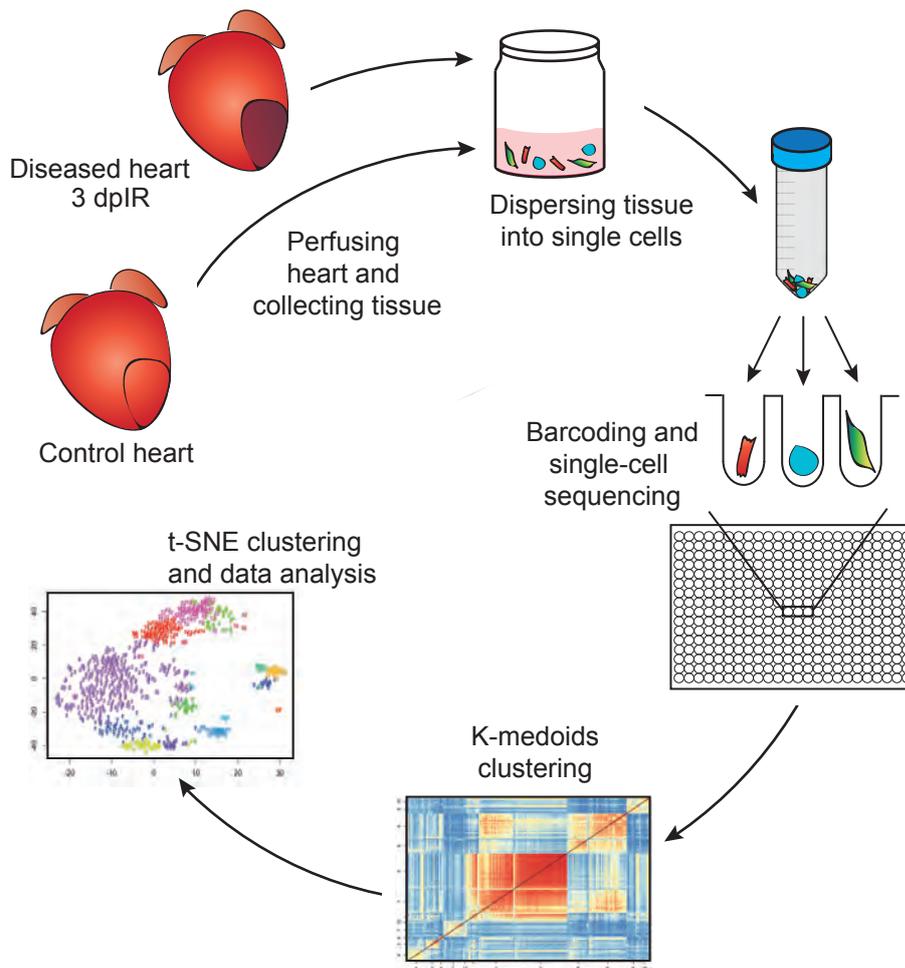


Figure S10. Schematic protocol for single-cell sequencing of the adult heart. Control or disease hearts were perfused and collected from mice, after which a selected area of tissue was washed, minced and subjected to enzymatic digestion. After dispersing the cells, the cells were sorted into 384 well plates, after which the cells were lysed, barcoded and subjected to automated single-cell sequencing. Downstream data analysis using RaceID2 revealed the intra- and intercellular changes in gene expression at a single-cell level of either a healthy or diseased heart.