**SUPPLEMENTAL MATERIAL**

**Single-cell transcriptomic profiling provides insights into
disease-related processes in human hypertrophic cardiomyopathy**

**Short title: Eding** Single-cell RNA sequencing in HCM

Joep E.C. Eding MD, MSc1#, Anne E. de Leeuw MSc1#, Maya Wright Clark, MSc 1,2#, Martijn Wehrens, PhD1, Cornelis J. Boogerd, PhD1, Bas Molenaar, MSc1, Petra H. van der Kraak, BSc3, Diederik W.D. Kuster, PhD4, Jolanda van der Velden, PhD4, Michelle Michels MD, PhD5, Aryan Vink MD, PhD3, and Eva van Rooij, PhD1,2**\***

# These authors contributed equally to this work

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

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

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

4 Department of Physiology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Cardiovascular Sciences, The Netherlands

5 Department of Cardiology, Erasmus MC, Rotterdam, The Netherlands

**Supplemental material**

* Supplemental methods
* Supplemental References
* Supplemental Tables
* Supplemental Figures

**Supplemental methods**

**Human heart samples**

Cardiac tissue from interventricular septum was obtained during myectomy surgery from patients with HCM to relieve left ventricular outflow tract obstruction (n=3 for single cell analysis, n=97 for real-time PCR (RT-PCR) analysis and n=1 for single cell/index sorting data). Approval for the use of human tissue samples was obtained from the local ethics committee (fresh myectomy samples, and bulk myectomy RNA samples) or the local scientific advisory board of the biobank of the University Medical Center Utrecht (explanted heart tissue for histology). For RT-PCR, cardiac tissue samples from non-failing donor hearts were used as control (n=9). Four of these were obtained from BioChain (Lots B607033, B711068, B711065 & A504241), the other five were obtained from the Sydney Heart Bank1.

**Histology and Immunohistochemistry**

Myectomy samples were fixed in PFA (4%) and incubated for 48 h rotating at room temperature. Next, the tissues were washed 3x for 10 min in PBS and stored in 70% ethanol (EtOH) at 4 ºC. For tissue embedding, 3 consecutive incubations were performed. Firstly, 96% EtOH for 2 h at 4 ºC, followed by 100% EtOH for 2h at 4 ºC and lastly Xylene for 2h at 4 ºC. Finally, the tissue was incubated overnight in liquid paraffin at 60 ºC and embedded in paraffin blocks. Then, 4 μm sections were cut on the HM 355S Automatic Microtome (#905200, Thermo Scientific) and placed on glass coverslips for further staining procedures. Explanted heart tissue was cut into 3 μm sections. For staining, sections were deparaffinized and rehydrated using xylene and ethanol graded series. Antigen retrieval was done by boiling the slides in EDTA buffer (pH 9.0) for 20 min and subsequently cooling them down to 37 ºC. Masson trichrome staining was performed by using the Artisan Link Pro (Agilent) stainer according to manufacturer’s protocol. For the immunostaining, sections were incubated overnight at 4 ºC with primary antibodies against titin (TTN, 9D10, Developmental Studies Hybridoma Bank, 1:400), ANP (CBL66, Millipore, 1:800) and ACTA1 (PA5-78716, ThermoFisher Scientific 1:250). For immunochemistry, after 1x PBS wash, slides were incubated with BrightVision poly-AP anti-rabbit IgG antibody (VWRKDPVR110AP, Immunologic) for 30 min at room temperature and with liquid permanent red (K0640, Agilent Dako) for 10 min at room temperature. Slides were counterstained with hematoxylin and mounted using Clearvue Mountant Xyl (ThermoFisher Scientific). Slides were digitalized using Nanozoomer XR (Hamamatsu). For immunofluorescence, slides were incubated overnight with primary antibody against ACTA1 (PA5-78716, ThermoFisher Scientific 1:250). After 3x PBS wash, slides were incubated with Alexa 647-labeled secondary antibody (A-21443, ThermoFisher Scientific, 1:500). Additionally, DAPI (#D3571, Invitrogen, 1:1000) and Wheat Germ Agglutin (WGA, W11261, ThermoFisher Scientific, 100 μg/ml) were added. Slides were incubated for 1 h in the dark at room temperature. Slides were subsequently washed 3X in PBS and mounted with prolong gold antifade (P36934, Life Technologies). Immunofluorescent imaging was done with the SPE Confocal Microscope (Leica).

**Tissue digestion**

The tissue was digested into a single-cell suspension as described before2. In short, tissue was minced into fine pieces using a scalpel and transferred into a glass vial with 1.5 mL of cold digestion buffer. Tissues were digested by shaking at 100 rpm for 15 minutes in a 37 ºC water bath. Subsequently, the solution was pipetted up and down 10 times and transferred onto a 100 μm cell strainer (EASYstrainer, #542000, Greiner Bio-One) placed on top of a 50 mL Falcon tube. The tissue was gently rubbed through the strainer using the plunger of a 1 mL syringe (#303172, BD Plastipak), after which the strainer was rinsed with 8.5 mL of DMEM (Dulbecco’s Modified Eagle Medium, high glucose, GlutaMAX Supplement, pyruvate (Gibco, #31966021)) to obtain a total volume of 10 mL. This suspension was centrifuged for 6 min at 4 ºC at 300 g. The supernatant was discarded and cells were resuspended in 1 mL fresh DMEM and kept on ice for immediate single-cell sorting.

**Flow cytometry to sort single cells**

Flow cytometry was performed on a FACS Aria III (BD Biosciences) using a 130 µm nozzle. Debris was excluded based on forward and side scatter area. Cells were selected for autofluorescence between 530 nm and 600 nm. Using FSC-W, the larger cells were selected in order to sort cardiomyocytes rather than other cell types. To show viability of sorted cells, DAPI (#D3571, Invitrogen, 1:1000) was added to a single-cell suspension. To show integrity, the cells sorted according to our protocol were collected, counterstained using DRAQ5 (65-0880-92, eBioscience, 1:1000) and re-analyzed by FACS3 to show that they are not nucleus-free cell fragments.

For sequencing, cells were single-cell sorted into 384-well plates, immediately centrifuged and frozen at -80 ºC until further processing. Additionally, 1000 cells were sorted into TRIzol reagent (Invitrogen, #15596026) for RNA quality control and 5000 cells were sorted into DMEM for imaging. Index sorting data was collected to correlate FSC (as a proxy for cell size) to gene expression on an individual cell basis.

**Imaging of single cells**

After digesting the tissue into a single-cell suspension, cells were imaged before and after sorting using Axiovert 40C (Zeiss) to visualize the morphology of the cells.

**RNA isolation and quality control**

Total RNA was isolated from 1000 cells bulk-sorted into 100μl TRIzol reagent. RNA quality, measured as RNA integrity number, was determined using a Bioanalyzer 2100 (Agilent) and RNA 6000 Pico chips (Agilent, #5067-1513). Single-cell RNA sequencing was only performed when the RIN was above 7.5.

**Library preparation and sequencing of single cells**

The SORT-seq procedure was performed by Single Cell Discoveries, Utrecht as described previously2,4,5, with minor adaptations. In short: cells were sorted into 384 well plates containing 10 μl of mineral oil and an aqueous solution of 50 nL containing primers derived from the CEL-seq2 protocol. CEL-seq2 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. Cells were lysed by 5 min incubation at 65 ºC, after which cDNA libraries were generated 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 and subsequent in vitro transcription for linear amplification as performed by overnight incubation at 37 ºC. 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 NextSeq500.

**Data analysis of single-cell RNA sequencing**

Paired-end reads from Illumina sequencing were mapped with BWA to the human reference transcriptome hg19. Read 1 was used to assign reads to libraries and cells and count UMIs while read 2 was used for mapping to the transcriptome. Reads that mapped equally well to multiple loci in the reference genome were excluded. For quantification of transcript abundance, the number of transcripts containing unique UMI’s were counted per cell-specific barcode for each gene in the reference transcriptome. Next, these 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 because of the high abundance of these transcripts in cardiomyocytes and their interference with downstream clustering. Then, all cells with a yield less than 1000 reads (after exclusion of reads mapping to the mitochondrial genome) were discarded. Next, RaceID27 was used to analyze the data with settings identical to those described in Gladka et al1. RaceID2 normalizes expression data by down-sampling all transcript counts to the lowest count (1000), uses t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the single cell gene expression profiles in a two-dimensional plot, and uses k-medoid clustering based on Pearson correlations as cell-cell distances to cluster cells. For displaying gene expression on t-SNE maps, the linear colour scales use the 1st percentile expression value as the lowest value and the 99th percentile expression value as the highest, this prevents the values at either extreme of the expression spectrum from skewing the scale too much, which would detract from readability. RaceID2 was also used to determine which genes showed enriched gene expression per cluster. Venn diagrams were generated using eulerAPE8.

**Correlation analyses**

For correlation analyses, the Pearson correlation between the parameter of interest (either expression of a gene of interest or FSC-A) and every (other) gene was calculated, provided that the gene was expressed in at least 20% of cells. This was done using the cor.test function from the R stats package, which also produces a p-value estimate based on the assumption that correlation coefficients of uncorrelated data follow a student-t distribution.

**Regulon analysis**

Clusters of genes that show significant correlations with each other were determined using a custom R script, which was applied to the gene expression data. We first calculated Pearson correlation coefficients between all genes that are expressed in at least 5% of all cells, resulting in a large gene-gene correlation matrix. Subsequently, we calculated p-values for respective correlation coefficients based on the student-t distribution, and applied the Benjamini-Hochberg correction to correct for multiple testing. We then selected genes that had a significant (p < 10-5) correlation with at least 40 other genes, resulting in the correlation matrix that was used for the regulon analysis. We applied the hclust function (using the ward.D2 algorithm) to sort rows and columns of the correlation matrix by similarity. Subsequently, we determined clusters of genes that show an enriched number of correlations within these clusters (i.e. regulons), To achieve this, we first determined the number of regulons based on the gap statistic for different k-values using k-means clustering (using the clusGap and SEMax function from the R cluster package, applying the default firstSEmax method). This procedure suggested that genes should be classified in 5 regulons; we then used the cuttree function from the R stats package (k=5) to perform this classification based on the hclust output. To determine a top-5 most characteristic genes for each of the regulons, the silhouette score was used, calculated by the silhouette function in the R cluster package.

**Identifying transcription factor motifs in the regulon gene promoters**

Transcription factor motif enrichment analysis in proximal promoters was performed using the findMotifs.pl function from the Homer suite v4.1010 using all expressed genes (expressed in at least 5% of all included cells) as background and promoter length set to -500 to +100bp.

**Gene ontology analysis**

For gene ontology (GO) analysis, gene names were mapped to Entrez Gene IDs. Mappings between Entrez Gene IDs and GO terms were then obtained from R-package *org.Hs.eg.db*11. A background set of genes was constructed by selecting all genes that were considered expressed in our data set (mean expression $\geq 1$ transcript per cell). A set of genes of interest was then determined based on genes that showed a relative expression change of 1.2x between cluster and non-cluster, a correlation coefficient > 0.1 to the gene of interest, or presence in a certain regulon. Enrichment of GO terms was then analyzed using the tools from the *GSEABase* R-package12.

**RT-PCR analysis**

Total RNA was isolated from 98 myectomy and 9 control samples using TRIzol reagent according to manufacturer’s protocol. RNA concentration was determined using Nanodrop 1000 spectrophotometer (ThermoFisher Scientific). Complementary DNA (cDNA) was synthesized from a total of 250ng of RNA using the iScript cDNA Synthesis Kit (Bio-Rad, #1708891), according to manufacturer’s protocol. RT-PCR was performed using gene specific primers (listed below) according to the instructions described by the IQ™ SYBR Green Supermix (Bio-Rad, #170-8885). The RT-PCR protocol was as follows: 95 ºC for 15min, followed by 40 cycles at 95 ºC for 15 s, 60 ºC for 30 s and 72 ºC for 30 s.

For RT-PCR of MYH6 and MYH7, cDNA was generated using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, #4368813) according to manufacturer’s protocol. RT-PCR reactions were performed using TaqMan Universal Master Mix II (Applied Biosystems, #4440040), according to the manufacturer’s instructions and using specific probes for MYH6 (ThermoFischer Scientific, catalogue number 4351372, identifier Hs01101442\_g1) and MYH7 (ThermoFischer Scientific, catalogue number 4331182, identifier Hs00165276\_m1). The RT-PCR protocol was as follows: 95 ºC for 10 min, followed by 40 cycles at 95 ºC for 15 s and 60 ºC for 60 s.

Primers:

|  |  |  |
| --- | --- | --- |
|  GENE | FORWARD SEQUENCE | REVERSE SEQUENCE |
| *ACTC1* | 5’-CCGTACCACAGGCATTGTTC | 5’-GACAAAGGAGTAGCCACGCT |
| *COL1A1* | 5’-ACAGCCTGGTGCTAAAGGTG | 5’ ACCAGGTTCACCGCTGTTAC |
| *COL2A1* | 5’-CCTGGTGTCATGGGTTTCC | 5’-GTCCTGCAGCACCTGTCTC |
| *CRYAB* | 5’-CCGACGTCTACTTCCCTGAG | 5’-CCATGCACCTCAATCACATC |
| *DSP* | 5’-GCACCAGCAGGATGTACTATT | 5’-TCAATTCAGGCTGCACGAT |
| *GAPDH* | 5’-GGGTCATCATCTCTGCCCC | 5’-GGTCATGAGTCCTTCCACGA |
| *HSPB1* | 5’-CGCGGAAATACACGCTGC | 5’-GTGATCTCGTTGGACTGCGT |
| *MYL2* | 5’-CCTTTCCACCATGGCACCTA | 5’-AAGCCATCCCTGTTCTGGT |
| *MYL3* | 5’-AAGATCACCTACGGGCAGTG | 5’-CTGGAGCATAGGCAGGAAAG |
| *MYOM1* | 5’-TTTGTTCGAGATGCTGATGC | 5’-ACGACCTTCGATCAATCCAG |
| *SYNM* | 5’-CTGGAGGATGAGAAGGCAAC | 5’-TCTGACGGCATGTTTTCAAC |
| *TTN* | 5’-GAGAGACGTCAGGCGAAAAG | 5’-TGGTCCAGGAGTGGTAAAGG |
| *XIRP2* | 5’-TAAGTCCTGCTTCCGATGCC | 5’-TGCTGCTTATGTCCAAAACCT |

The percentual contribution of MYH6 to the total cardiac myosin heavy chain mRNA content (MYH6 + MYH7) was determined by calculating the myosin ΔCt (= CtMYH6 - CtMYH7) and then applying the formula: 100% \* (1/(2ΔCt+1)). The percentual MYH7 contribution was then calculated by subtracting the percentual MYH6 contribution from 100%.

Fold change expression of COL1A1 and COL2A1 in HCM samples over control samples was calculated using the ΔΔCt method. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a housekeeping gene, to which all genes were normalized. The results are shown as mean ± standard error of the mean (SEM). Significance was tested using Students *t*-test. The number of samples (n) used in each experiment is indicated in the legend or shown in the figures. Correlation between genes was analyzed by calculating the Pearson correlation coefficient of the ΔCt-values. Plotting was done using PRISM (GraphPad Software Inc.) for pie charts and bar graphs. R and ggplot2 were used to plot correlation analyses.

**Cell size analysis**

Cell size was manually measured on sections stained for ACTA1 and labelled with WGA by using ImageJ 1.49v software. Individual cardiomyocytes were manually outlined on images using the WGA signal to identify cell boundaries, ImageJ then measured the area of each identified cell and quantified fluorescence intensity in the ACTA1 channel with that area. Total ACTA1 fluorescence was normalized to cell area. Cell size and ACTA1 were quantified in 3-5 images each for 6 myectomy samples (464-941 cells per samples, 3685 cells in total).

**Supplemental references**

1. dos Remedios CG, Lal SP, Li A, McNamara J, Keogh A, Macdonald PS, Cooke R, Ehler E, Knöll R, Marston SB, Stelzer J, Granzier H, Bezzina C, van Dijk S, De Man F, Stienen GJM, Odeberg J, Pontén F, Linke W and van der Velden J. The Sydney Heart Bank: improving translational research while eliminating or reducing the use of animal models of human heart disease. *Biophysical Reviews*. 2017;9:431-441.

2. Gladka MM, Molenaar B, de Ruiter H, van der Elst S, Tsui H, Versteeg D, Lacraz GPA, Huibers MMH, van Oudenaarden A and van Rooij E. Single-Cell Sequencing of the Healthy and Diseased Heart Reveals Cytoskeleton-Associated Protein 4 as a New Modulator of Fibroblasts Activation. *Circulation*. 2018;138:166-180.

3. Smith PJ, Wiltshire M and Errington RJ. DRAQ5 labeling of nuclear DNA in live and fixed cells. *Curr Protoc Cytom*. 2004;Chapter 7:Unit 7 25.

4. Grun D and van Oudenaarden A. Design and Analysis of Single-Cell Sequencing Experiments. *Cell*. 2015;163:799-810.

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. Grun D, Kester L and van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods*. 2014;11:637-40.

7. Grün 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.

8. Micallef L and Rodgers P. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One*. 2014;9:e101717.

9. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR and Weirauch MT. The Human Transcription Factors. *Cell*. 2018;172:650-665.

10. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H and Glass CK. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*. 2010;38:576-589.

11. Carlson, M (2019). org.Hs.eg.db: Genome wide annotation for Human. R package [computer program]. Version 3.8.2.

12. Morgan M, Falcon S, Gentleman R (2019). GSEABase: Gene set enrichment data structures and methods [computer program]. Version 1.46.0.