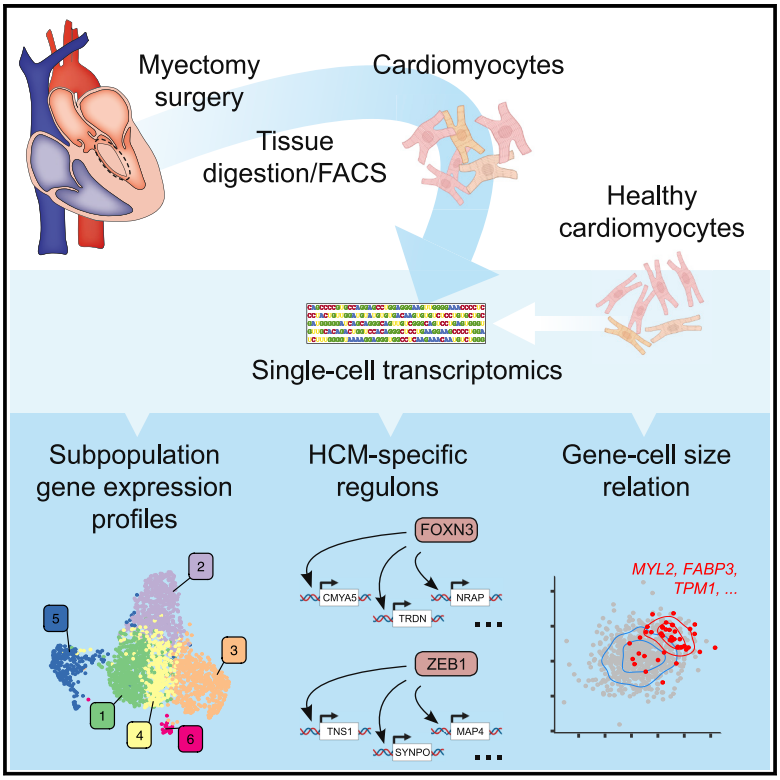
Cell Reports

Article

Single-cell transcriptomics provides insights into hypertrophic cardiomyopathy

Graphical abstract



Highlights

- We sequence single cardiomyocytes from patients with hypertrophic cardiomyopathy
- Diseased cells show diverse gene expression profiles that might be linked to HCM
- Comparison of healthy and diseased cells reveals disease-specific gene expression
- FACS-data quantify cell size and allow identification of cell-size-related genes

Authors

Martijn Wehrens, Anne E. de Leeuw, Maya Wright-Clark, ..., Michelle Michels, Aryan Vink, Eva van Rooij

Correspondence

e.vanrooij@hubrecht.eu

In brief

Wehrens et al. perform single-cell RNA sequencing on hearts from patients who suffer from hypertrophic cardiomyopathy (HCM). This reveals cardiomyocyte subpopulations, gene regulatory networks, transcription factors, and hypertrophy-related genes that might be relevant for pathogenesis of the disease.

Wehrens et al., 2022, Cell Reports *39*, 110809
May 10, 2022
https://doi.org/10.1016/j.celrep.2022.110809

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Single-cell transcriptomics provides insights into hypertrophic cardiomyopathy

Martijn Wehrens,[1,6] Anne E. de Leeuw,[1,6] Maya Wright-Clark,[1,2,6] Joep E.C. Eding,[1,6] Cornelis J. Boogerd,[1] Bas Molenaar,[1] Petra H. van der Kraak,[3] Diederik W.D. Kuster,[4] Jolanda van der Velden,[4] Michelle Michels,[5] Aryan Vink,[3] and Eva van Rooij[1,2,7,*]

[1]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center, Utrecht, the Netherlands
[2]Department of Cardiology, University Medical Center Utrecht, Utrecht, the Netherlands
[3]Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands
[4]Department of Physiology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Cardiovascular Sciences, Amsterdam, the Netherlands
[5]Department of Cardiology, Erasmus MC, Rotterdam, the Netherlands
[6]These authors contributed equally
[7]Lead contact
*Correspondence: e.vanrooij@hubrecht.eu
https://doi.org/10.1016/j.celrep.2022.110809

SUMMARY

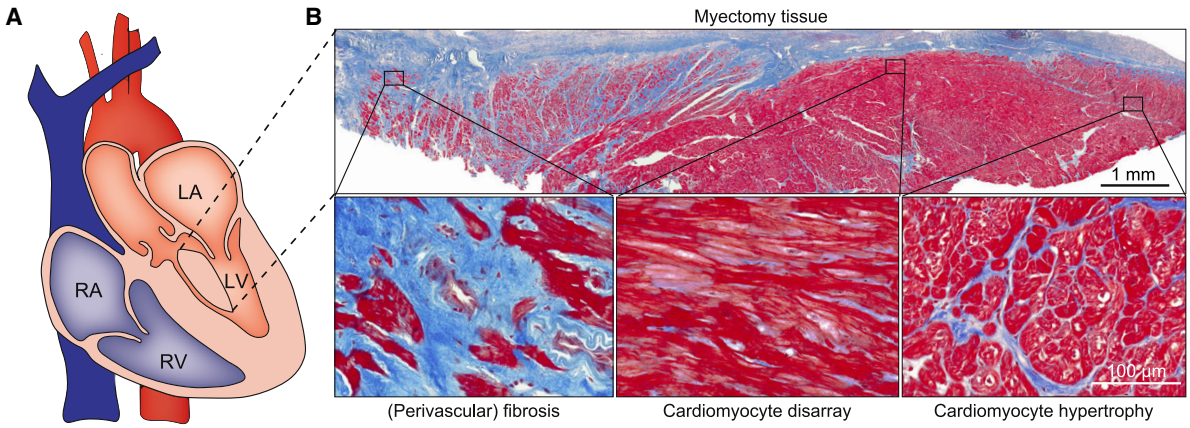
Hypertrophic cardiomyopathy (HCM) is a genetic heart disease that is characterized by unexplained segmental hypertrophy that is usually most pronounced in the septum. While sarcomeric gene mutations are often the genetic basis for HCM, the mechanistic origin for the heterogeneous remodeling remains largely unknown. A better understanding of the gene networks driving the cardiomyocyte (CM) hypertrophy is required to improve therapeutic strategies. Patients suffering from HCM often receive a septal myectomy surgery to relieve outflow tract obstruction due to hypertrophy. Using single-cell RNA sequencing (scRNA-seq) on septal myectomy samples from patients with HCM, we identify functional links between genes, transcription factors, and cell size relevant for HCM. The data show the utility of using scRNA-seq on the human hypertrophic heart, highlight CM heterogeneity, and provide a wealth of insights into molecular events involved in HCM that can eventually contribute to the development of enhanced therapies.

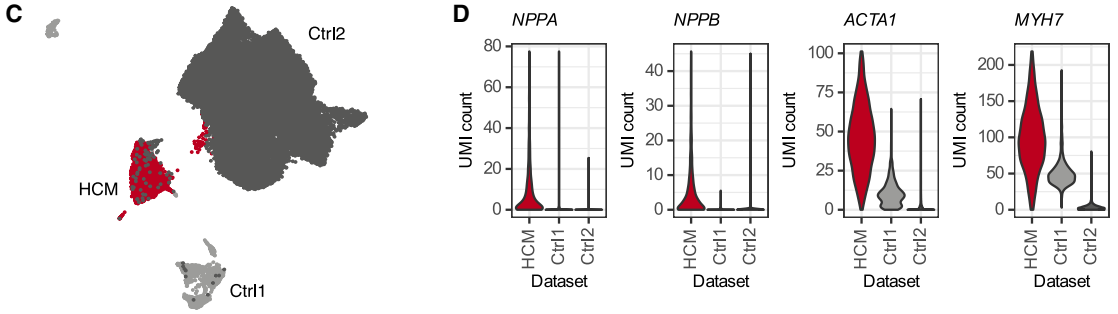
INTRODUCTION

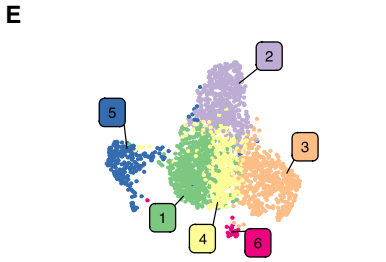
Hypertrophic cardiomyopathy (HCM) is a genetic cardiac disorder with an incidence of 1 in 200 to 500 individuals (Maron et al., 2014; Semsarian et al., 2015). The phenotype can vary from essentially asymptomatic to end-stage heart failure or cause life-threatening arrhythmias (Maron and Maron, 2013). Typically, patients carry a pathologic DNA variant in genes encoding sarcomere proteins. β-myosin heavy chain (MYH7) and myosin binding protein C (MYBPC3) are the genes most commonly involved; however, the causal genes in approximately 40% of patients with HCM remain to be identified (Marian, 2010). Clinically, HCM is characterized by unexplained segmental hypertrophy that is usually most pronounced in the basal interventricular septum (Marian and Braunwald, 2017). Myocyte disarray, a pathological hallmark of HCM, involves 5% to 40% of the myocardium, and is usually mainly present in areas of more severe hypertrophy. Other key histological features include interstitial fibrosis and vascular abnormalities (Marian and Braunwald, 2017).

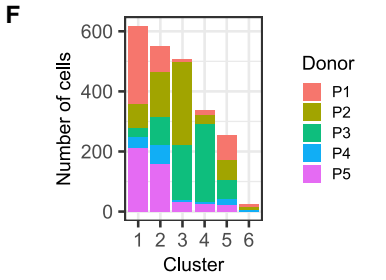
To date, the molecular mechanisms that underlie the remodeling processes in HCM remain largely unclear. While it has been suggested that myocyte disarray and hypertrophy are a direct result from changes in sarcomere function induced by the HCM-related mutations (Di Domenico et al., 2012), recent work suggests that the HCM phenotype might also be triggered by a functional imbalance among individual cardiomyocytes (CMs). Unequal force generation between adjacent CMs can initiate CM and myofibrillar disarray and trigger stretch-induced signaling leading to development of interstitial fibrosis and hypertrophy (Montag et al., 2018; Parbhudayal et al., 2018).

Single-cell RNA sequencing (scRNA-seq) provides a detailed view on gene expression differences between cell types or transcriptome heterogeneity across cells of the same type (Grun et al., 2014; Kolodziejczyk et al., 2015). Recently, we developed an approach that allows us to obtain single-cell transcriptomic data from all main cardiac cell types of the adult murine heart under both healthy and diseased conditions (Gladka et al., 2018). While several studies by now have used scRNA-seq to study adult cardiac biology in an in-depth manner (Farbehi et al., 2019; Gladka et al., 2018; Hu et al., 2018; Kretzschmar et al., 2018; Larson and Chin, 2021; Nomura et al., 2018; Ren et al., 2020; Sereti et al., 2018; Skelly et al., 2018), even on adult human CMs (Cui et al., 2019; Nomura et al., 2018; Wang et al., 2020), so far this has not been done on cells from human HCM myectomy samples.



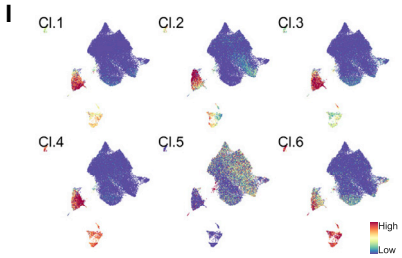


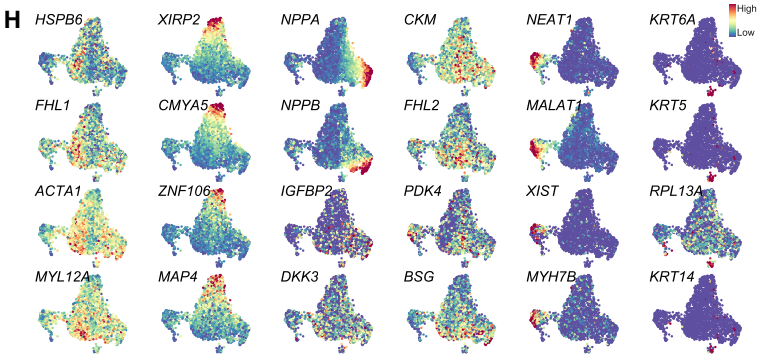




G

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	Gene	FC	Gene	FC	Gene	FC	Gene	FC	Gene	FC	Gene	FC
	HSPB6	1.67	*XIRP2*	3.36	*NPPA*	6.65	*CKM*	1.69	*NEAT1*	12.2	*KRT6A*	15.0
	FHL1	1.59	*CMYA5*	3.33	*NPPB*	4.76	*FHL2*	1.66	*MALAT1*	11.4	*KRT5*	12.5
	ACTA1	1.52	*ZNF106*	3.11	*IGFBP2*	1.38	*PDK4*	1.60	*XIST*	3.37	*RPL13A*	5.25
	MYL12A	1.46	*MAP4*	2.46	*DKK3*	1.34	*BSG*	1.56	*MYH7B*	3.05	*KRT14*	4.95
	LPL	1.38	*TCAP*	2.39	*RTN4*	1.31	*CKB*	1.52	*DDX17*	2.74	*RPS6*	4.79
	CSRP3	1.37	*NRAP*	2.32	*GAPDH*	1.31	*ATP2A2*	1.50	*PNISR*	2.43	*S100A8*	4.64
	PCDH7	1.36	*SYNM*	2.15	*PDK4*	1.30	*MYL3*	1.47	*WSB1*	2.36	*ANXA2*	4.61
	MYL2	1.36	*DES*	2.09	*XIRP1*	1.29	*TXNIP*	1.47	*CFLAR*	2.25	*ACTG1*	4.54
	HSPB8	1.34	*MYOM1*	2.03	*BSG*	1.26	*PTGDS*	1.46	*ACTB*	2.16	*RPL14*	4.30
	UBC	1.32	*HIPK2*	2.03	*S100A1*	1.24	*SLC16A7*	1.43	*PDE7A*	2.00	*KRT17*	4.28
HOMER			*MEF2C, TEAD4, MEF2B, TCF7L2, TCF3*		*GATA6, GATA3*		*MEF2D, MEF2A, HOXD13, MEF2B, SOX3*		*SRF, NF1, FOXF1, PRDM1*		*ERG, IRF8, FRA1, ZEB1*	
Lisa	*GATA4*, ZNF512, EZH2*, KLF10, BRD2, NR3C1*, TP53*, ERG**		*ARID3B, GATA4*, TBX5*, CEBPB, MEF2B, AR**		*SRPK2, ESR1*, ZNF805, TP53, SMARCC1, AR*, GATA4*, NR3C1*, EZH2**		*598-SKD, SOX9, ESR1, EGR2, LMNA*, LANA*, AR*, GATA4*, EZH2**		*HOXB13, FOXA1*, CEBPB*, GATA2*, MEF2B, GATA3*, FOS**		*WDR5*, MYC*, TP53*, EZH2*, MAFB, KAT2B**	
TRIAGE	*SFRP1, GNAS*		*DSP, MYO18B, NRAP*				*PEBP4*		*BTNL9, MIR1-1HG, CRNDE, CACNA1C, RBM20*		*CXCL14, TRIM29, DUSP4, CAMK2N1, KRT7*	





(legend on next page)

Here we used scRNA-seq to study cellular transcriptional differences between healthy and hypertrophic CMs and employed this expressional heterogeneity to link gene expression profiles to cellular characteristics related to HCM. In-depth analysis of the scRNA-seq data from HCM CMs indicated the presence of subpopulations of CMs to which each patient contributed. In addition, we were able to identify HCM-related gene correlations. Using these data, we could identify groups of genes that are co-expressed (regulons [Aibar et al., 2017]) and we were able to link these to transcription factors (TFs) that are potentially responsible for their activation. In addition, index-sorting data enabled us to correlate gene expression profiles to CM hypertrophy and confirmed myosin light chain (*MYL2*/MLC-2) expression in larger CMs. Together these data indicate CM heterogeneity in the human heart and show that scRNA-seq provides insights into cellular and molecular mechanisms that are potentially relevant for HCM.

RESULTS

Single-cell gene expression analysis of septal myectomy samples from patients with HCM

Patients with HCM show a heterogeneous remodeling response that is characterized by localized CM hypertrophy, disarray, and fibrosis (Marian and Braunwald, 2017). To explore the molecular mechanisms underlying the cellular heterogeneity observed in HCM, we aimed to examine differential gene expression between individual CMs. To this end, we collected cardiac tissue from patients with HCM who underwent a septal myectomy for outflow tract obstruction and processed the tissue for scRNA-seq (Figure 1A). Histological analysis of the myectomy samples confirmed key hallmarks of HCM (Figure 1B). To achieve this, we used our previously optimized digestion and sorting strategy (Gladka et al., 2018). We enzymatically dispersed cardiac tissue into a single-cell suspension and sorted cells into 384-well plates, where we gated for larger, single cells to enrich for CMs (Figures S1A and S1B). This gating strategy resulted in intact and nucleated cells, as resorting of this population indicated 82% of the cells to be DRAQ5 positive and DAPI negative (Figures S1C and S1D). DRAQ5 is able to enter live cells and stain nuclear DNA (Smith et al., 2004), whereas DAPI selectively stains nuclei of compromised cells. Events that are DRAQ5+ and DAPI− must therefore contain uncompromised cells with a nucleus. With each sort, we additionally collected cells for microscopy and RNA quality control. Imaging the cells after sorting visually indicated that we were sorting intact cells (Figure S1D). RNA quality from the dispersed and sorted cells was retained, as indicated by RNA Integrity Number (RIN) (Schroeder et al., 2006) (Figure S1E). Together these data showed that our protocol allowed for the isolation of good-quality RNA from individual, intact cells collected from human septal myectomy samples. To obtain single-cell transcriptomes of individual cardiac cells from myectomy samples, we used the SORT-seq protocol, as described previously (Gladka et al., 2018; Grun et al., 2015).

Transcript abundance per gene was quantified by using a custom mapping pipeline using STAR and featureCounts (see STAR Methods for additional details); 33% ± 12% of transcripts mapped to the mitochondrial genome (Figure S1F), which is consistent with results from previous studies (Kannan et al., 2019). Reads mapping to the mitochondrial genome were excluded from data, since they interfere with the downstream analysis. After filtering cells for a minimum of 1,000 transcripts mapping to the nuclear genome, a total of 2,292 cells from five different septal myectomy samples with an average number of 2,201 unique non-mitochondrial reads per cell were included for downstream *in silico* analysis (Figure S1G). A Uniform Manifold Approximation and Projection (UMAP) map for *MYH7*, a well-known CM marker, confirmed the far majority of the remaining cells to be CMs (Figure S1H). Together these data show we were able to collect reliable scRNA-seq data from HCM CMs to start exploring potential disease underlying mechanisms.

scRNA-seq identified HCM-related gene expression changes

To identify HCM-related gene expression changes, we next combined our scRNA-seq data with gene expression data from healthy adult human CMs. To do so we included scRNA-seq data from healthy left ventricular (LV) CMs from Wang et al. (Ctrl1, n = 1400) (Wang et al., 2020) and from septal CMs from Litviňuková et al. (Ctrl2, n = 27604) (Litviňuková et al., 2020) (Tables S1 and S2). To prevent batch effects stemming

Figure 1. Single-cell analysis of septal myectomy samples from patients with HCM reveals different subpopulations of CMs

(A) Schematic representation of the human heart highlighting the septal myectomy sample used for this study.

(B) Masson's trichrome staining of myectomy tissue from a patient with HCM showing fibrosis, myocardial disarray, and CM hypertrophy.

(C) UMAP showing transcriptome similarities between HCM cells acquired in this study (red dots), and cells from healthy donors from previously published studies (Ctrl1 and Ctrl2, respectively light gray and dark gray dots).

(D) Violin plots of the normalized unique molecular identifier (UMI) counts for *NPPA, NPPB, ACTA1,* and *MYH7* indicating an induction of stress markers in HCM CMs.

(E) UMAP showing transcriptome similarities between HCM cells only. The colors represent the clusters identified by Seurat.

(F) Bar graph showing the contribution of each of the five patients (P1–P5) to each of the Seurat clusters.

(G) Top 10 most enriched genes in each cluster and their FC enrichment in the respective cluster. The bottom three rows show TFs potentially driving gene expression for each cluster (HOMER and Lisa) and cluster-enriched genes ($p < 0.01$) with the largest regulatory potential (TRIAGE). Enrichment was calculated for cells in the cluster over all cells outside the cluster.

(H) UMAPs depicting the expression of the top 4 enriched genes for each cluster. Expression in UMAPs is shown as normalized transcript counts on a color-coded linear scale.

(I) Composite expression of HCM cluster enriched genes (see E–G), projected on UMAPs with pooled cells from HCM, Ctrl1, and Ctrl2 (see C). Composite expression is determined by cell-averaged *Z* scores of enriched gene sets (FC > 1.1, $p < 0.05$) per cluster (Cl.1–5). RA, right atrium; RV, right ventricle; LA, left atrium; LV, left ventricle. See also Figures S1 and S2, and Tables S1 and S2.

from bio-informatic analyses as much as possible, we downloaded raw FASTQ files and re-mapped Ctrl1 using our own pipeline. For Ctrl2, we downloaded available mapped data, as the Ctrl2 mapping pipeline closely matched our pipeline (see STAR Methods for additional details).

Next, the Seurat algorithm was applied for identification and clustering of cells (Figures S2A-S2C) (Stuart et al., 2018). Cells appeared to separate from each other based on origin, with the majority of HCM cells clustering away from cells coming from control hearts (Figure 1C). Clustering analysis identified five different CM clusters (Figure S2A), with cluster 4 mainly representing CMs from HCM hearts (Figure S2B). This HCM cluster was enriched for well-known cardiac stress marker genes, such as *MYH7*, *NPPA,* and *XIRP2* (Figure S2C and Table S3), which is in line with these cells being diseased compared with the control cells. Also, when plotting the individual expression levels of these cardiac disease-related genes, we could show a clear induction in the HCM CMs compared with both sets of control cells (Figure 1D). Functionally, the HCM cluster and cluster 5, which mainly contains Ctrl1 cells, show similar enrichment of gene ontology (GO) terms related to energy metabolism, which might be explained by the fact that HCM and Ctrl1 processed whole cells for sequencing, as opposed to Ctrl2, which mainly consists of sequenced CM nuclei (Figure S2D). Nevertheless, these data indicate the HCM CMs to express disease-related gene expression profiles that might provide insights into mechanisms relevant for cardiac remodeling during HCM.

scRNA-seq reveals different subpopulations of HCM CMs

To start exploring the gene expression profiles underlying the cellular heterogeneity in HCM, we next focused in on the CMs coming from HCM hearts. Clustering analysis on the 2,292 included cells revealed six clusters to which every patient contributed (Figures 1E–1G, S2E, and Table S4). Even though we were unable to detect a clear separation between clusters 1 to 4, we could confirm gene enrichment by UMAP (Figure 1H). This suggests gene expression differences in the clusters to be gradual instead of bimodal with heterogeneous cell-to-cell differences in gene expression. These clusters might represent differential gene regulation between different groups of single CMs, which could be related to functionally different pathogeny and/or disease progression at the single-cell level. To identify potential drivers of this regulation, we then applied HOMER, Lisa, and TRIAGE, which are software packages that can be used to identify regulatory factors based on gene subsets, to the lists of cluster-enriched genes (Figure 1G).

To explore the functional relevance of the different clusters, we performed GO analysis on the genes that showed significant enrichment per cluster (adjusted $p < 0.05$ and fold-change (FC) > 1.1) (Figure S2F). Clusters 1–2 showed GO term enrichment for terms related to the sarcomere, suggesting this subpopulation has more pronounced sarcomeric remodeling (or retainment) compared with other cells. Clusters 3 and 4 on the other hand showed GO terms related to signaling and metabolic processes, respectively. UMAPs containing both control and HCM CMs showed that genes enriched in HCM clusters 1 to 4 are HCM specific (Figure 1I).

In summary, scRNA-seq on human myectomy samples revealed the presence of functionally different subpopulations of CMs in the human HCM heart, which could be relevant for the disease.

NPPA expression is specific for a subset of CMs in patients with HCM

Classically, cardiac expression levels of Natriuretic Peptide A and B (*NPPA*/ANP and *NPPB*/NBP) have served as a hallmark for CM hypertrophy, stress or failure (Man et al., 2018).

Though almost all HCM cells are expressing higher levels of *NPPA* compared with Ctrl samples (Figure 1D), our clustering analysis indicated the presence of a subpopulation of HCM CMs that is even more enriched for *NPPA* (cluster 3) (Figure 1G). The heterogeneity in *NPPA* expression was visualized by a UMAP indicating the *NPPA* expression per cell (Figure 1H). We next used our scRNA-seq data to determine which genes show a correlation with *NPPA*. By taking a cutoff of 0.01 for the adjusted p value, we identified 83 positively correlated genes and 48 negatively correlated genes (Figure 2A and Table S5). By far the strongest positive correlation was found for *NPPB*, which is in line with expectations (Kretzschmar et al., 2018). UMAP confirmed the overlap for positively correlated genes (Figure 2B) and the negatively correlated genes (Figure 2C). Interesting to note is that genes related to muscle contraction (*TTN* and *RYR2*) and heart disease (*XIRP2* and *CRYAB*) showed a lower abundance in the cells with higher levels of *NPPA*.

Immunohistochemistry for ANP on myectomy samples indicated the ANP-positive cells to predominantly border fibrotic areas in the HCM samples (Figure 2D), while no positive cells could be detected in a control heart (Figure 2E). The HCM-specificity was further confirmed by comparing *NPPA* expression in HCM versus Ctrl cells, which showed very few Ctrl cells express *NPPA* (Figures 1D and 2F). The positive correlations for the top correlated genes were consistently found in multiple patients with HCM, indicating the correlation represents biology rather than it being by chance or dominated by data from a single patient (Figure 2G). The near absence of *NPPA* expression in the majority of CMs from control hearts made a comparable correlation analysis not applicable in those. In general, the *NPPA* positively correlated genes appeared more abundant in HCM CMs than cells coming from control hearts (Figure 2H).

These data indicate that while overall *NPPA* expression is higher in HCM CMs, there is a subpopulation of CMs that is enriched for *NPPA* and *NPPB* that are located in stressed, fibrotic regions in HCM hearts that is absent in control human heart samples. The gene expression profile of these cells could help us identify genes involved in the CM stress response.

XIRP2 correlations are more pronounced in HCM CMs

Xin Actin Binding Repeat Containing 2 (*XIRP2*) is expressed in cardiac and skeletal muscle where it interacts with actin and a-actinin. It localizes to the costamere and intercalated disks, two critical structures for cardiac function (Farrell et al., 2018). While relatively little is known about *XIRP2* function, mutations in *XIRP2* have been linked to dilated cardiomyopathy (DCM) (Long et al., 2015) and an increase in expression was observed in cardiac tissue from mice with a loss of cMyBP-C (Farrell

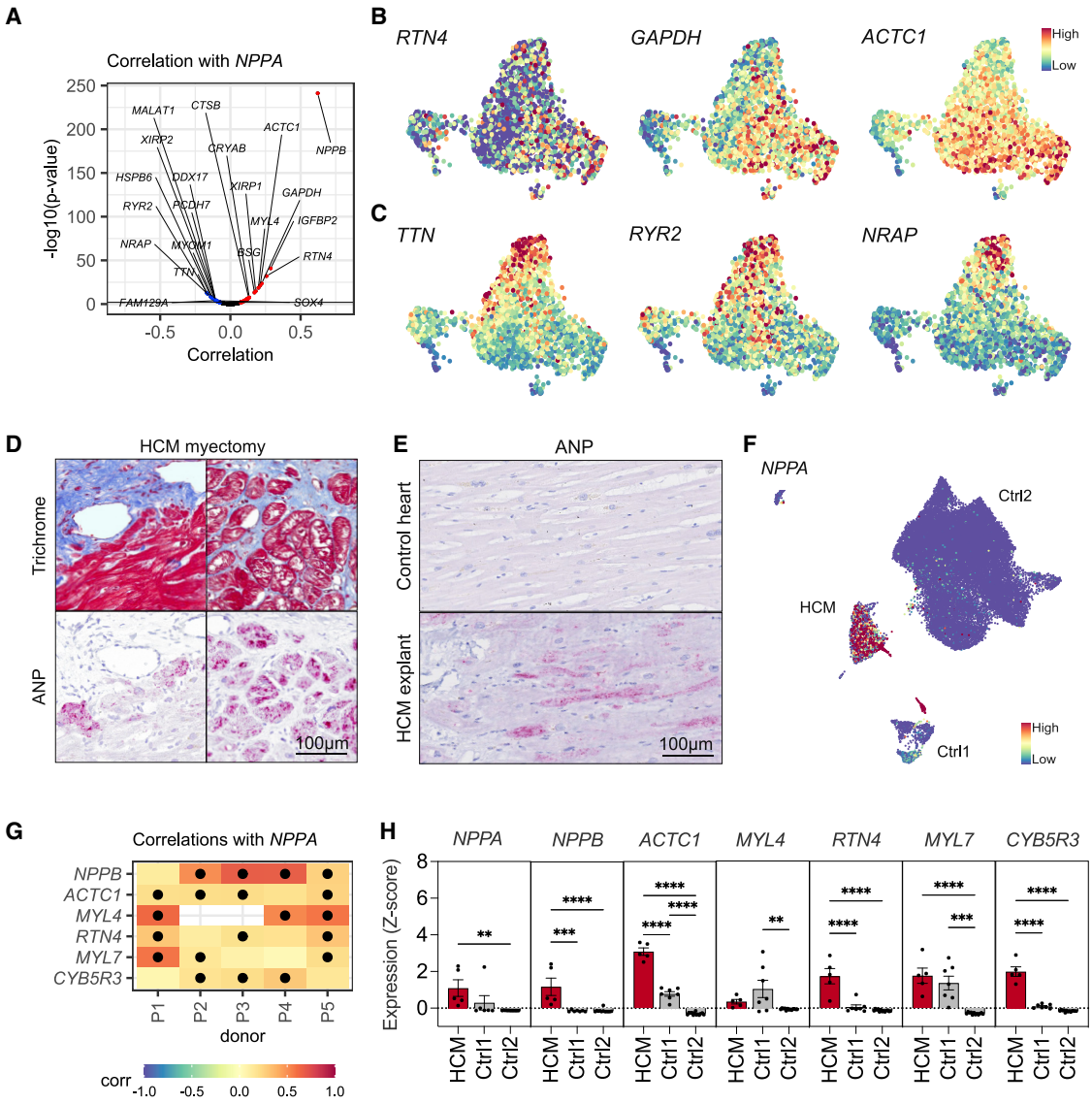


Figure 2. A subpopulation of HCM cardiomyocytes shows increased *NPPA* expression

(A–C) (A) Volcano plot depicting genes positively (red dots) and negatively (blue dots) correlated to *NPPA* with an adjusted p value < 0.01. B-C, UMAPs showing the expression level of genes positively (B) and negatively (C) correlated to *NPPA*. Expression in UMAPs is shown as normalized transcript counts on a color-coded linear scale.

(D) Representative Masson's trichrome staining (above) and ANP staining (below) on human HCM myectomy tissue showing ANP to be most expressed in CMs bordering areas of fibrosis.

(E) Immunohistochemical staining showing ANP-expression in an explanted control and HCM heart.

(F) UMAP of *NPPA* expression in all datasets showing HCM CMs to have more *NPPA* positive cells compared to healthy adult CMs.

(G) Heatmap depicting *NPPA* correlation coefficients with listed genes determined from the respective patients with HCM separately. A t test was performed with Benjamin & Hochberg correction, black dots indicate that the correlation is considered significant ($p < 0.05$).

(H) Comparison of expression of *NPPA* and *NPPA*-correlated genes between the HCM and Ctrl cells. Expression of each gene was first normalized (*Z* score), after which average expression was determined per donor (black dots). Bars indicate averages per condition. Error bars show SEM. Outliers were removed by using the ROUT test with Q = 1% and a one-way ANOVA test was performed. *$p \leq 0.05$; **$p \leq 0.01$, ***$p \leq 0.001$; ****$p \leq 0.0001$.

et al., 2018). Based on our clustering, we found *XIRP2* to be the most enriched gene for cluster 2 (Figure 1G) and negatively correlated with *NPPA* expression (Figures 2A and Table S5).

Correlation analysis for *XIRP2* showed the strongest positive correlation with *CMYA5*, *ZNF106,* and *MAP4* expression, while genes like *ACTC1* and *MYBPC3* showed a negative correlation (Figure 3A, S3A and Table S6). To assert whether the *XIRP2*-gradient gene expression (Figure 1H) was based on biology and not by chance or dominated by the data from a single patient, we looked into *XIRP2* expression for each individual

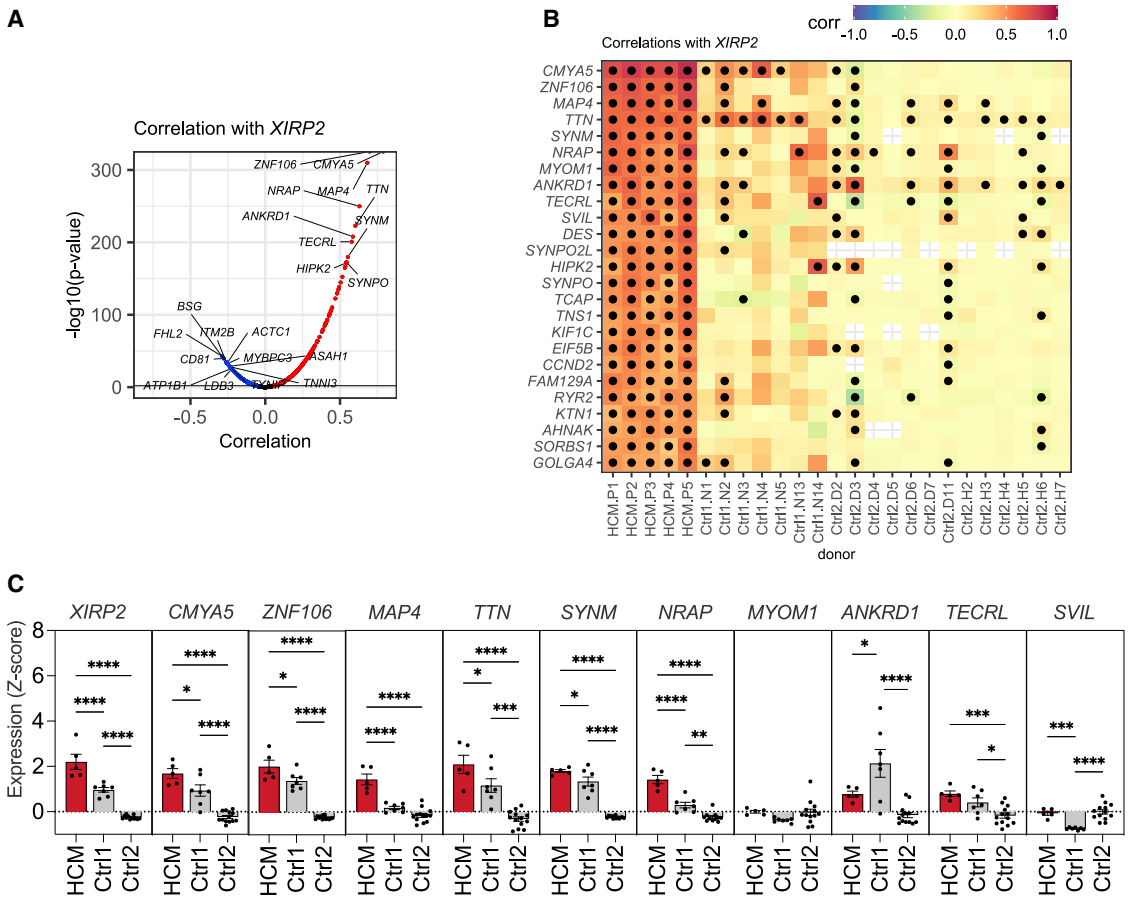


Figure 3. A subpopulation of HCM CMs shows increased *XIRP2* expression

(A) Volcano plot depicting genes positively (red dots) and negatively (blue dots) correlated to *XIRP2* in patients with HCM with an adjusted p value < 0.01.

(B) Heatmap depicting correlation coefficients between *XIRP2* and listed genes, determined in the listed donors. The top 25 genes most correlated to *XIRP2* in HCM are shown (based on average correlation coefficient in patients with HCM). A t test was performed with Benjamin & Hochberg correction, black dots indicate that the correlation is considered significant ($p < 0.05$).

(C) Comparison of expression of *XIRP2* and *XIRP2*-correlated genes between the HCM and Ctrl cells. Expression of each gene was first normalized (*Z* score), after which average expression was determined per donor (black dots). Bars indicate averages per condition. Error bars show SEM. Outliers were removed by using the ROUT test with Q = 1% and a one-way ANOVA test was performed. *$p \leq 0.05$; **$p \leq 0.01$, ***$p \leq 0.001$; ****$p \leq 0.0001$. See also Figure S3.

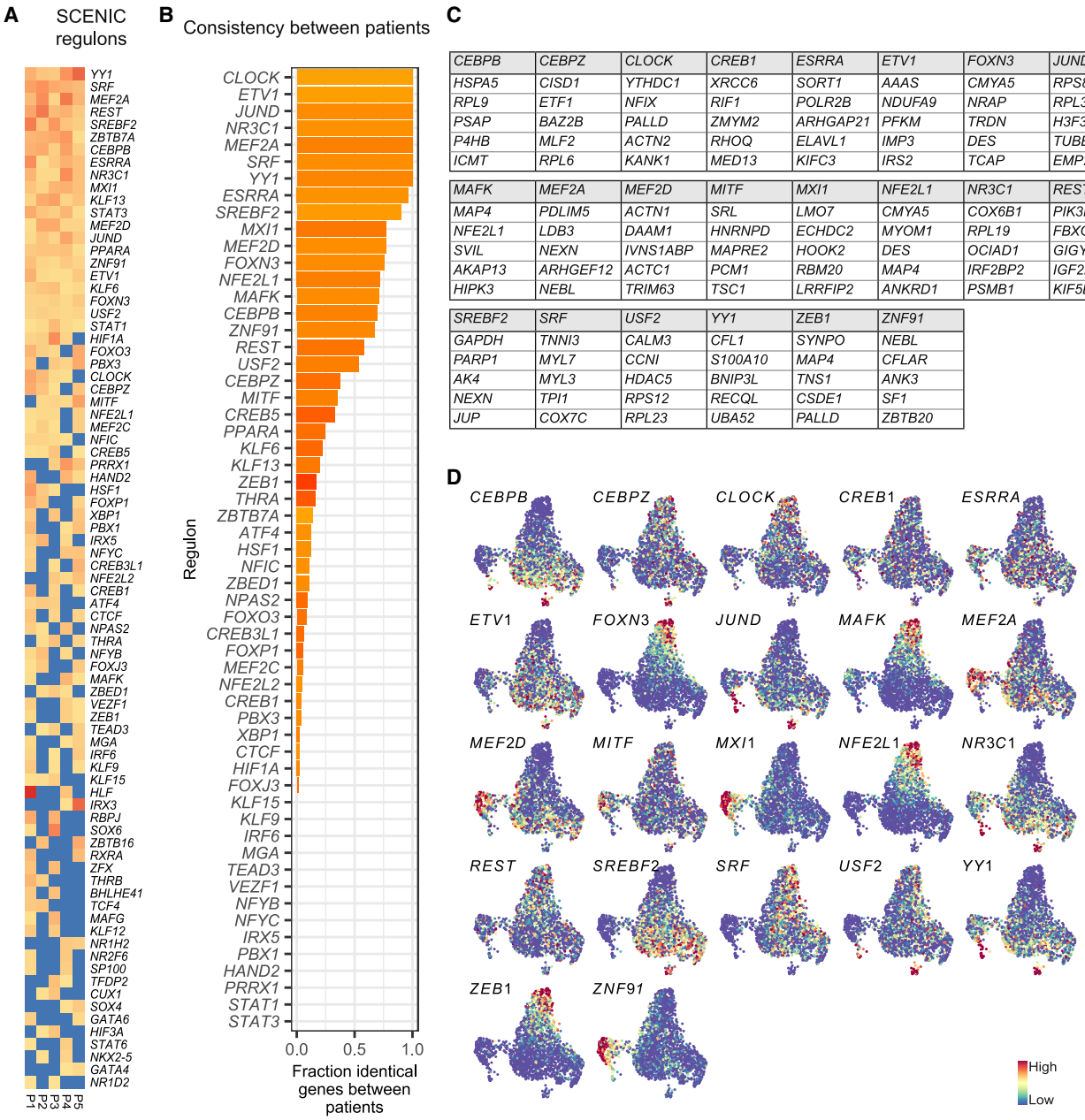
patient. Patient-specific UMAPs confirmed the presence of a *XIRP2*-gradient in each patient (Figure S3B), and a comparable expression profile for the correlated genes *CMYA5* and *TTN* (Figures S3C and S3D). Histology confirmed the heterogeneity in TTN expression at the protein level (Figure S3E). However, this heterogeneity was also observed in control tissue, indicating that healthy cells also show some degree of heterogeneity. These findings are in line with previous research showing that TTN is one of a set of proteins that have a mosaic expression pattern in the heart (Wang et al., 2018).

To determine whether these co-expression patterns are specific for HCM, we next assessed these correlations in control CMs. This indicated the co-expression of these genes to be more pronounced in cells from HCM hearts compared with control CMs (Figures 3B and S3A), making them potentially disease relevant. Based on expression levels, the majority of the XIRP2 correlated genes also appeared to be more abundantly present in HCM CMs than in CMs from control hearts (Figure 3C).

Taken together, these data underline the validity of our observed gene expression profiles, and indicate the *XIRP2* gene correlations to be more pronounced in CMs from HCM hearts, which could be relevant for the disease.

Regulon analysis reveals potential gene modules and TFs driving HCM

Clustering of cells offers a way to determine cellular subpopulations, which can be used to identify relevant gene expression patterns. However, many gene expression programs are likely to be active in the diseased heart, and they do not necessarily all need to coincide with subpopulations identified by clustering. Thus, to further identify gene expression patterns that might be relevant to HCM, we used SCENIC (Aibar et al., 2017) to look for patterns in our heterogeneous HCM single-cell expression data. The aim of this software is to identify potential regulons: groups of co-expressed genes driven by a specific TF. To achieve this, it uses machine learning methods to fit expression



CEBPB	CEBPZ	CLOCK	CREB1	ESRRA	ETV1	FOXN3	JUND
HSPA5	CISD1	YTHDC1	XRCC6	SORT1	AAAS	CMYA5	RPS8
RPL9	ETF1	NFIX	RIF1	POLR2B	NDUFA9	NRAP	RPL37A
PSAP	BAZ2B	PALLD	ZMYM2	ARHGAP21	PFKM	TRDN	H3F3B
P4HB	MLF2	ACTN2	RHOQ	ELAVL1	IMP3	DES	TUBB
ICMT	RPL6	KANK1	MED13	KIFC3	IRS2	TCAP	EMP2

MAFK	MEF2A	MEF2D	MITF	MXI1	NFE2L1	NR3C1	REST
MAP4	PDLIM5	ACTN1	SRL	LMO7	CMYA5	COX6B1	PIK3R1
NFE2L1	LDB3	DAAM1	HNRNPD	ECHDC2	MYOM1	RPL19	FBXO32
SVIL	NEXN	IVNS1ABP	MAPRE2	HOOK2	DES	OCIAD1	GIGYF2
AKAP13	ARHGEF12	ACTC1	PCM1	RBM20	MAP4	IRF2BP2	IGF2R
HIPK3	NEBL	TRIM63	TSC1	LRRFIP2	ANKRD1	PSMB1	KIF5B

SREBF2	SRF	USF2	YY1	ZEB1	ZNF91
GAPDH	TNNI3	CALM3	CFL1	SYNPO	NEBL
PARP1	MYL7	CCNI	S100A10	MAP4	CFLAR
AK4	MYL3	HDAC5	BNIP3L	TNS1	ANK3
NEXN	TPI1	RPS12	RECQL	CSDE1	SF1
JUP	COX7C	RPL23	UBA52	PALLD	ZBTB20

Figure 4. SCENIC analysis of gene expression heterogeneity reveals potential regulatory interactions

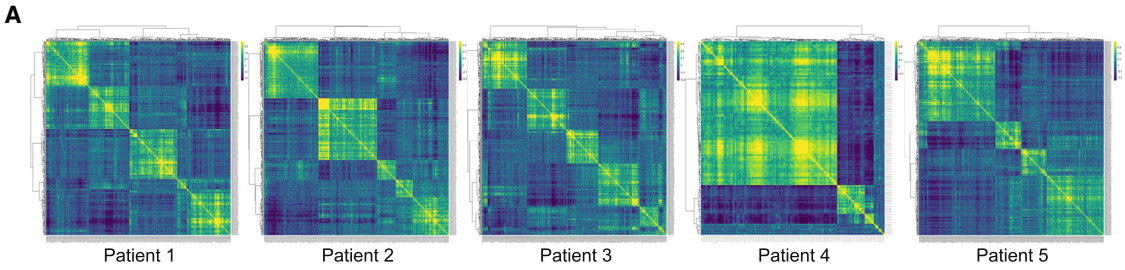
(A) Regulons determined in each patient with HCM using SCENIC. Regulons are named according to the TF suggested to regulate them. Color coding indicates the normalized area under the curve score (NES), which relates to the confidence in the link between the TF and its associated genes (yellow to red). Blue indicates a regulon was not detected in respective patient. Only regulons identified in >1 patient are listed.

(B) Graph showing the fraction of identical genes between patients.

(C) Table depicting the five top genes within each regulon based on median gene importance scores.

(D) Projection of the composite expression of indicated regulons on the HCM dataset UMAP. Composite expression is calculated by first normalizing gene expression (*Z* score), and then calculating the mean of regulon member genes expression per cell. Expression is shown by a color-coded linear scale. See also Figures S4 and S5.

patterns of genes to expression patterns of TFs. These fits are used to determine potential regulons, which are further refined through TF motif analysis. We ran SCENIC separately on data of each of the five patients. To narrow down regulons of general importance to HCM, we first selected regulons identified in three or more (of five) patients (Figure 4A), and then selected regulons most consistently linked to the same group of genes (Figure 4B). This identified 22 regulons (Figure 4C and Table S7) that showed



B

Module 1	Module 2	Module 3	Module 4	Module 5
NEAT1	*MYL2*	*RPS12*	*CMYA5*	*BSG*
MALAT1	*TNNI3*	*TMSB4X*	*XIRP2*	*CASQ2*
MYH7B	*ACTC1*	*TPT1*	*ZNF106*	*ASAH1*
DDX17	*CRYAB*	*RPL34*	*MAP4*	*PSAP*
CFLAR	*SLC25A4*	*ACTG1*	*NRAP*	*CD81*
PNISR	*TNNC1*	*RPS27A*	*ANKRD1*	*ATP1B1*
N4BP2L2	*MYL3*	*RPL23*	*TTN*	*GPNMB*
LUC7L3	*MB*	*RPS20*	*HIPK2*	*LPL*
CACNA1C	*ACTA1*	*RPL37*	*DES*	*LGALS3BP*
TNRC6A	*CKM*	*RPL32*	*TECRL*	*CTSD*
HOMER	*MEF2B, SRF, SOX9, MEF2D*	*IRF8, IRF3, PU, ERG, ER71*	*SCL, TEAD4, TCF7L2, GATA1, MEF2C*	*ERRA*

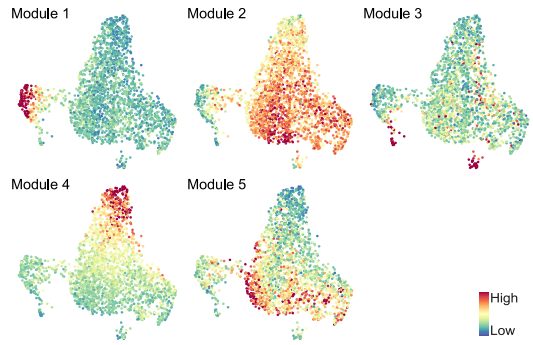


Figure 5. Gene expression heterogeneity reveals patterns of gene coregulation

(A) Patient-specific heatmaps of the Pearson correlation matrix of a subset of detected genes that are expressed in >20% of cells and have a significant ($p < 0.001$) correlation with at least 10 other genes. Hierarchical clustering was used to identify regulons in each of the patients.

(B) Top 10 genes found in each module (based on average correlation to other module member genes). The bottom row indicates TFs potentially regulating respective modules, as determined by HOMER analysis.

(C) UMAPs showing the composite module expression in the HCM dataset. Composite expression is mean *Z* score of the member gene expression levels, represented by a linear color-coded scale. See also Figures S4 and S5.

heterogeneous expression over our single cells, as determined by the average expression *Z* score of member genes (Figure 4D). GO analysis of these regulons showed functional differences between the regulons that might be related to processes relevant for hypertrophy and cell-signaling in HCM; for example, some regulons could be linked to energetic or metabolic processes, while some others were linked to sarcomere or cytoskeleton organization, and again others to membrane targeting of proteins (Figure S4A).

Gene expression is often determined by posttranscriptional activation of TFs, rather than by differential expression of TFs. As SCENIC is based on TF differential expression data (Aibar et al., 2017), we next searched for groups of co-expressed genes (modules) independently of TF expression patterns. To do so, we selected genes that were detected in all patients and expressed in at least 5% of all cells (1871 genes) and generated gene-gene correlation matrices for each patient. We then selected all genes that had a significant correlation with at least 10 other genes (respectively 356, 334, 382, 120, 389 genes per patient). Hierarchical clustering analysis on the correlation matrix for these genes resulted in patient-specific modules (Figure 5A). Strikingly, patient-to-patient comparison of the identified modules gave rise to five shared modules that were consistently identified in each of the patients (Figures 5B, S4B, and S4C and Table S8), indicating the modules to be of biological significance. The expression of these shared modules marks different cell populations, as can be seen when visualized on a UMAP (Figure 5C). Based on gene content, we again see functional differences between the modules, where terms related to mitochondrial and ATP (module 2), sarcomere organization (module 4), and calcium signaling and conduction (module 5) are consistent with HCM-related processes (Figure S4D). To identify TFs that could potentially be involved in the coregulation of the genes per module, we screened the promotor regions of these genes for binding site enrichment using HOMER. This yielded several potential transcriptional regulators (Figure 5B). Shared module 2 for example showed an enrichment for myogenic enhancer factor 2 (*MEF2*) binding sites, a well-known regulator of muscle genes (Black and Olson, 1998), *SRF* and *SOX9*, which have all been shown to regulate CM hypertrophy (Schauer et al., 2019; Zhang et al., 2001). This suggest that module 2 genes might be involved in CM hypertrophy.

Having identified regulons and modules from differential gene expression within the HCM population of cells, we next aimed to determine which of these regulons and modules show upregulation in HCM hearts compared with healthy hearts. We first applied the SCENIC algorithm to all donors from the Ctrl samples, and found all of the selected 22 regulons are also identified in healthy donor CM populations (Figures S4E and S4F). This suggests that regulatory interactions underlying this

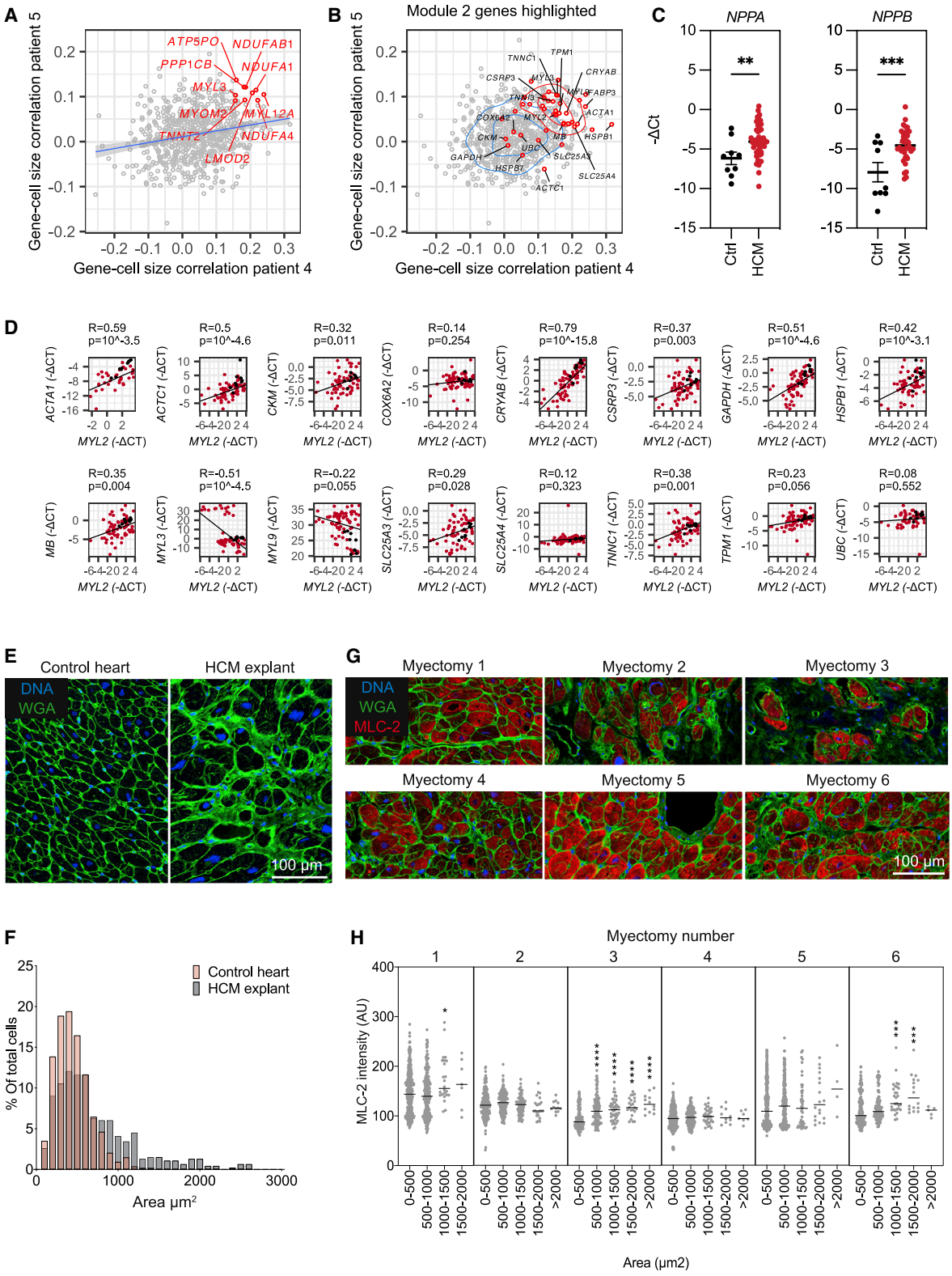


Figure 6. Index sorting provides a gene expression profile for enlarged CMs

(A and B) Correlations between FSC-A and gene expression determined from patient 4 (x axis) versus those correlations determined from patient 5 (y axis). Each gray dot corresponds to a gene. In (A), genes that occur in the top 10% FSC-A correlated genes of both patient 4 and 5 are highlighted in red, and in (B) genes that occur in module 2 are highlighted in red.

(legend continued on next page)

co-expression are also active in healthy CM populations. We additionally tested whether the co-expression groups are unique in terms of member genes, and found that there is limited overlap between the individual regulons and/or modules (Figure S4G). To more directly determine HCM-specificity of the regulons and modules, we assessed their expression in each of the patients and donors. Strikingly, this shows that the majority of the SCENIC regulons show enrichment in HCM, with the regulons *ZEB1*, *FOXN3*, *NFE2L1,* and *MAFK* showing the largest contrast between HCM and Ctrl (Figure S5A). In addition, we observe that co-expression modules 4 and 5 show a clear HCM-enrichment (Figure S5B).

Integration of cell size with scRNA-seq reveals hypertrophy-associated genes

As one of the hallmarks of HCM is CM hypertrophy, it is of great interest to integrate cell size with the transcriptomic data obtained from scRNA-seq. Previously, index-sorting data have been used to integrate morphological parameters obtained from flow cytometry with gene expression data obtained from scRNA-seq from bone marrow stem cells (Tan et al., 2017). To determine the enriched genes in hypertrophic CMs, we gathered FSC-A data in patients 4 and 5 as a proxy for cell size, which we subsequently linked to the gene expression profile of single CMs.

To first test the validity of this approach, we correlated FSC-A with total single-cell mRNA read count, since bigger cells are expected to contain more mRNA. This showed a positive correlation between mRNA read count and FSC-A values for both patients (Figures S6A and S6B). Further correlation analysis revealed genes that were both positively and negatively correlated with cell size (Figure 6A and Table S9). While some of these genes have previously been linked to hypertrophy (Lim et al., 2001), we here show the correlation of these genes to CM size on a single-cell level rather than on the organ-wide level. Interestingly, the genes constituting shared module 2 show much higher correlations to cell size than other genes ($p = 2 \times 10^{-12}$ and $p = 5 \times 10^{-8}$ for patients 4 and 5, respectively) (Figures 6B, S6C, and S6D).

In conjunction with our earlier observation that hypertrophy-related TFs are linked to this module, this suggests module 2 might be of extra importance to hypertrophy. We therefore aimed to validate the correlations between genes in this module in 97 additional septal myectomy samples from patients with HCM and nine left ventricle RNA samples from non-failing donor hearts as controls (Control). Out of the 97 patients with HCM, 42 (43%) had a known pathogenic HCM mutation (*MYBPC3,* n = 26; *MYH7*, n = 7; multiple, n = 3; other, n = 6), while the causal mutation was unknown for the rest of the patients with HCM (Table S1).

RT-PCR analysis for *NPPA* and *NPPB* validated the disease phenotype in the HCM samples compared with control (Figure 6C). RT-PCR confirmed that the majority of pairwise correlations between the top-ranked gene in module 2, *MYL2*, and other module 2 genes are positive (Figure 6D). In addition, pairwise comparison of all genes in module 2 showed 81% of correlations to be positive (Figure S6E, Table S10). These data independently confirmed the validity of the correlations in gene expression that were identified by scRNA-seq.

To visualize the distribution in cell size we performed Wheat Germ Agglutin (WGA) staining on control and HCM tissue and were able to show that CMs in HCM samples were on average larger than in control samples and showed a wider range in CM size (Figures 6E and 6F). In the sarcomere, myosin is formed by two heavy chains (encoded by *MYH6* and *MYH7*) and two myosin light chains (encoded by *MYL2* and *MYL3*). *MYL2* encodes the regulatory light chain (MLC-2) whereas *MYL3* encodes the essential myosin light chain (Sheikh et al., 2015). In both patients 4 and 5, *MYL3* and *MYL2* were present among the genes positively correlated to cell size (Figure 6A). To further examine the correlation between *MYL2* expression and cell size, we performed MLC-2 immunostainings on six septal myectomy samples (Figure 6G) and quantified both MLC-2 intensity and cell size. In three out of six samples we could confirm an increase in MLC-2 protein in the larger cells (Figure 6H).

Together these data indicate that index sorting allows us to link cell morphological parameters to gene expression profiles and enables us to identify genes linked to CM hypertrophy.

DISCUSSION

In this study, we used scRNA-seq to define gene expression patterns that are relevant for HCM. By comparing gene expression profiles from healthy and diseased subjects we were able to show an enrichment for known and markers of CM stress induced in HCM hearts. In-depth analysis of genetic profiles and gene correlations in diseased CMs revealed intercellular gene expression differences. Based on this heterogeneity, we bioinformatically clustered transcriptionally related cells or genes, each of which are likely related to different cellular functions. In addition, based on scatter properties of HCM CMs we were able to correlate gene expression to cell size, which

(C) RT-PCR showing an increase in *NPPA* and *NPPB* expression levels in patients with HCM. Outliers were removed by using the ROUT test with Q = 1%, t tests were performed. For *NPPA*, there are n = 9 Ctrl and n = 49 HCM samples, for *NPPB*, there are n = 9 Ctrl and n = 36 HCM samples, mean and SEM are indicated.
(D) RT-PCR validation of correlation between genes identified in module 2. Pearson's correlation coefficients (indicated with an R) were calculated and t tests were performed based on HCM samples. Samples were normalized to *RPL32*. Per gene, n values are as follows: *ACTA1*, 33; *ACTC1*, 65; *CKM*, 62; *COX6A2*, 68; *CRYAB*, 71; *CSRP3*, 64; *GAPDH*, 62; *HSPB1*, 61; *MB*, 66; *MYL3*, 59; *MYL9*, 74; *SLC25A3*, 59; *SLC25A4*, 66; *TNNC1*, 66; *TPM1*, 69; *UBC*, 52. In addition n = 9 Ctrl values are shown in each plot.
(E) Representative immunofluorescent images from a control heart and an explanted HCM heart stained for WGA.
(F) Histogram showing the different distributions in CM cross-sectional area between a control heart and an explanted HCM heart.
(G) Representative images of WGA (green) and MLC-2 (red) co-staining on 6 HCM myectomy samples.
(H) Quantification of CM cross-sectional area and MLC-2 fluorescence intensity (AU) in the samples from (G). Outliers were removed by using the ROUT test with Q = 1%. Differences were tested by using a one-way ANOVA followed by a Tukey's multiple comparisons test. FSC-A indicates Forward scatter area; AU, Arbitrary units; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$. See also Figure S6 and Table S1.

contributes to a better understanding of disease-driving mechanisms in HCM.

In comparing healthy versus HCM CMs, we could detect clear transcriptional differences between the two. This included an HCM-related enrichment of well-known marker genes for cardiac stress, but additionally revealed genes that previously had not been functionally linked to HCM, such as myoglobin (encoded by *MB*). Myoglobin serves as an oxygen storage unit in muscle cells and is known as a circulating biomarker for muscle injury (Berenbrink, 2021). Our data show a strong transcriptional induction of *MB* in HCM CMs compared with healthy cells. This might be because *MB* is induced under hypoxic conditions (Kanatous et al., 2009), which is a key feature of cardiac hypertrophy (Mirtschink and Krek, 2016). Whether this actively contributes to HCM pathogenesis or whether circulating levels of *MB* can serve to track HCM disease progression is currently unknown and warrants further investigation.

CM heterogeneity has been shown to occur at many levels (Montag et al., 2018; Parbhudayal et al., 2018; Remme et al., 2009; van der Velden et al., 2011). This heterogeneity in adult CMs has more recently been confirmed by single-cell studies (Gladka et al., 2018; Nomura et al., 2018; Wang et al., 2020), including the one described here. As the primary genetic origin of HCM lies within the sarcomere, and CM heterogeneity is a key hallmark in HCM, we focused on the analysis of CMs only. Based on the inter-myocyte transcriptomic heterogeneity, our algorithm clustered the CMs into six distinct subpopulations, with one cluster being highly enriched for *NPPA*. Correlation analysis revealed a positive correlation between *NPPA* and known stress marker genes, such as *NPPB*, *ACTC1,* and *XIRP1*, but also showed a positive correlation with genes with relatively unknown cardiac functions, like reticulon 4 (*RTN4*). This suggests these lesser-known genes could also play an active part in the CM stress response.

For example, *RTN4* is related to a variety of functions, it is known as a neurite outgrowth inhibitor, might play a role in the endoplasmic reticulum stress response in human DCM and cardiac ischemia (Ortega et al., 2014). In addition, it affects lipid homeostasis, AKT signaling, and cytoskeleton modulation in cancer (Pathak et al., 2018). Moreover, it is induced in mouse HCM models (Sasagawa et al., 2016) and it was recently shown that a knockdown of *RTN4* might be cardioprotective (Fan et al., 2021). However, a clear functional role in the CM stress response still remains to be defined. Nevertheless, given these previous observations, and our observation that *RTN4* expression correlates with *NPPA*, suggests that *RTN4* might play a role in HCM pathogenesis.

In addition, we noticed a positive correlation between *NPPA* expression and *GAPDH*, implying that *GAPDH* might not be suitable as A housekeeping gene when studying stressed CMs. Interesting to note is the negative correlation with key genes related to CM contractility, such as *TTN* and *RYR2*, suggesting different contractile properties in cells expressing high levels of *NPPA*.

To investigate the gene regulatory network and TFs that drive HCM, we used SCENIC. Twenty-two regulons were consistently identified in patients with HCM, of which most were also expressed at higher levels in HCM CMs. Some, like *MEF2* and *SRF*, are well-known regulators of HCM (Chai and Tarnawski, 2002; Kolodziejczyk et al., 1999; Passier et al., 2000; Zhang et al., 2001), validating our approach. However, not all of the most HCM-enriched factors, i.e., *ZEB1*, *FOXN3*, *NFE2L1,* and *MAFK*, are currently known to play a role in heart disease. To our knowledge, *ZEB1* and *FOXN3* have not yet been shown to have a clear function in HCM. *NFE2L1* has been shown to be cardioprotective (Cui et al., 2021), and we have previously identified *MAFK* to be induced in failing CMs (Vigil-Garcia et al., 2021). The role of *MAFK* might thus not be restricted to genetic forms of heart failure. UMAPs indicated a clear overlap between the cells enriched for the regulons related to *NFE2L1* and *MAFK*. This is interesting, as these TFs heterodimerize for DNA binding (Kannan et al., 2012) and have both been shown to provide cellular protection against oxidative stress through the induction of antioxidant genes (Itoh et al., 1999; Numazawa et al., 2003). Our data imply that *NFE2L1* and *MAFK* are cooperatively induced in a set of HCM CMs to drive a gene program involved in sarcomere organization and muscle contraction. While the regulons for *FOXN3* and *ZEB1* also appear enriched in the same set of cells, there is currently no known functional relationship among these four factors. However, based on the SCENIC analysis and the fact that the majority of the *XIRP2* correlated genes defined in Figure 3 are predicted to be regulated by one or more of these factors, these TFs together might drive at least part of the *XIRP2* correlated gene program.

Based on index-sorting data, we were able to show that several genes in shared module 2 were positively correlated to CM size at the single-cell level. Motif analysis of module 2 member genes indicated an enrichment for *MEF2* binding sites, which would be in line with the well-known function of *MEF2* in hypertrophic remodeling (Dirkx et al., 2013). Recently, Nomura et al. (2018) used scRNA-seq to identify modules of co-expressed genes that correlated with CM size in mice 1 week after transverse aortic constriction. The genes in our hypertrophy-associated regulon 2 match surprisingly well (57%) with the hypertrophy-associated modules (M1, M2, M5, M11, and M16) in the paper by Nomura et al. This reinforces our observations to originate from biological regulation, and confirms the validity of using index sorting to link gene expression to cell size.

It is evident that a lot of experimental choices regarding cell collection and obtaining sequencing data can influence the outcome when using scRNA-seq. This might be an explanation for the observed separation between the two sets of control CMs that we included in our analyses.

scRNA-seq of CMs is challenging simply due to the size of the cells. Several strategies are currently used to separate single cells into individual wells or droplets, some of which allow for the high-throughput processing of thousands of transcriptomes in a cost-effective manner. However, all these high-throughput strategies have physical constraints regarding the size of the cells. Commercially available single-cell sorting platforms like Fluidigm C1 and Chromium can currently only sort cells that are up to 25 to 50 μM in diameter. This is considerably smaller than adult mammalian CMs, which can be approximately 125 μM along the longitudinal axis (Sorenson et al., 1985). This challenge has been circumvented by performing single-nucleus RNA sequencing on cardiac tissue, but with the drawback that

there will be enrichment for RNAs residing predominantly in the nucleus (Selewa et al., 2020).

The percentage of mitochondrial reads is regarded as an additional indicator of CM scRNA-seq library quality, as cytoplasmic transcripts leak out of damaged cells, whereas mitochondrial transcripts do not. Mitochondrial reads are expected to fall within 30% to 50% of total reads (Kannan et al., 2019). If techniques are used that damage CMs, this can increase to even 70% (Fluidigm C1 [DeLaughter et al., 2016] or handpicking myocytes [Nomura et al., 2018]) or 90% (conventional fluorescence-activated cell sorting [FACS] [Gladka et al., 2018]). Thus, the 33% mitochondrial reads in the CMs that we sequence corroborate that we sequenced intact myocytes.

The pathogenic processes and gene programs that occur in HCM are a complex interplay between many genes and gene programs, with cross-talk between different physiological pathways. Our study identifies many correlations, co-expression patterns, regulatory factors, and relations to cell size that offer an insight in this complex world. Our study provides a valuable data source to identify and further study molecular mechanisms that might be relevant for many HCM processes. This detail in insight will lead to discoveries related to HCM and other areas of heart disease.

Limitations of the study

In our study, we collected single cardiomyocytes from patients with HCM to analyze gene expression patterns and heterogeneity. Both enzymatic digestion of patient material and FACS might introduce bias in cell recovery. For example, fibrotic tissue might be harder to digest, and certain cells might be more susceptible to FACS-induced damage. In addition, we focused on CMs in this study, disregarding other cell types in the heart. To compare diseased cardiomyocytes with healthy cells, we have included reference data from healthy cells from two other studies. While we tried to minimize batch effects by closely matching bio-informatic pipelines, variation from different experimental protocols might still introduce batch effects, as also discussed above. In addition, since the five patients analyzed in our study have either an *MYBPC3* mutation or unknown genotype, we are unable to investigate the effects of specific HCM genotypes. To analyze our single-cell data, we have used established bio-informatic analyses, and focused on effects that are observed in multiple patients and depend as little as possible on analysis settings. Nevertheless, analysis choices can potentially affect identified features and analysis outcomes, such as cluster assignments of cells. While many observations we make based on single-cell data are consistent with literature and additional experiments, further follow-up studies will be required to validate and investigate functional implications of many of our findings.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human heart samples
- METHOD DETAILS
  - Histology and immunohistochemistry
  - Tissue digestion
  - Flow cytometry to sort single cells
  - Imaging of single cells
  - RNA isolation and quality control
  - Library preparation and sequencing of single cells
  - Mapping of HCM sequencing data
  - Obtaining Ctrl1 and Ctrl2 count tables
  - Analysis of single cell transcriptomics data using seurat
  - Gene ontology analysis
  - Correlation analyses
  - Transcription and regulatory factor analyses
  - SCENIC analysis
  - Custom gene module analysis
  - RT-PCR analysis
  - Cell size analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical and quantitative analysis of single cell RNA-sequencing data
  - Statistical and quantitative analysis of scRNA-seq and cell size data
  - Statistical and quantitative analysis of RT-PCR data
  - Statistical and quantitative analysis of staining and microscopy images

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.110809.

ACKNOWLEDGMENTS

We gratefully acknowledge Roy Huurman, Sylvie Dekker, Martine de Boer, Harm de Wit, and Peter van Geel from the Erasmus Medical Center for facilitating sample collection and help with flow cytometry. We thank Stefan van der Elst from the Hubrecht Institute FACS facility for his help with the viability experiments and presenting our flow cytometry data, Jeroen Korving for his help with histology, and Anna Alemany for her help with mapping raw data. We further thank Anko de Graaff and the Hubrecht Imaging Center for supporting the imaging. Last, we acknowledge Cris dos Remedios for providing us donor heart samples from the Sydney Heart Bank. This work was supported by the Leducq Foundation (14CVD04), the European Research Council under the European Union's Seventh Framework Program (ERC Grant Agreement CoG 615708 MICARUS), and the Dutch Cardiovascular Alliance (DCVA), an initiative with support of the Dutch Heart Foundation, DCVA2017-18 ARENA-PRIME and DCVA2014-40 CVON-DOSIS.

AUTHOR CONTRIBUTIONS

J.E.C.E., A.E.L., M.W.C., and E.R. designed and performed experiments. M.W., J.E.C.E., C.J.B., and B.M. analyzed the data. P.K., D.K., J.V., M.M., and A.V. provided resources and methodologies J.E.C.E., A.E.L., M.W.C., M.W., and E.R wrote the manuscript. E.R. acquired funding.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 5, 2021
Revised: January 25, 2022
Accepted: April 21, 2022
Published: May 10, 2022

REFERENCES

Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

Berenbrink, M. (2021). The role of myoglobin in the evolution of mammalian diving capacity – the August Krogh principle applied in molecular and evolutionary physiology. Comp. Biochem. Physiol. A. Mol. Integr. Physiol. *252*, 110843. https://doi.org/10.1016/j.cbpa.2020.110843.

Black, B.L., and Olson, E.N. (1998). Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. Annu. Rev. Cell Dev. Biol. *14*, 167–196. https://doi.org/10.1146/annurev.cellbio.14.1.167.

Chai, J., and Tarnawski, A.S. (2002). Serum Response Factor: discovery, biochemistry, biological roles and implications for tissue injury healing. J. Physiol. Pharmacol. *53*, 147–157.

Cui, M., Atmanli, A., Morales, M.G., Tan, W., Chen, K., Xiao, X., Xu, L., Liu, N., Bassel-Duby, R., and Olson, E.N. (2021). Nrf1 promotes heart regeneration and repair by regulating proteostasis and redox balance. Nat. Commun. *12*, 5270–5315. https://doi.org/10.1038/s41467-021-25653-w.

Cui, Y., Zheng, Y., Liu, X., Yan, L., Fan, X., Yong, J., Hu, Y., Dong, J., Li, Q., Wu, X., et al. (2019). Single-cell transcriptome analysis maps the developmental track of the human heart. Cell Rep. *26*, 1934–1950.e5. https://doi.org/10.1016/j.celrep.2019.01.079.

DeLaughter, D.M., Bick, A.G., Wakimoto, H., McKean, D., Gorham, J.M., Kathiriya, I.S., Hinson, J.T., Homsy, J., Gray, J., Pu, W., et al. (2016). Single-cell resolution of temporal gene expression during heart development. Dev. Cell *39*, 480–490. https://doi.org/10.1016/j.devcel.2016.10.001.

Dirkx, E., da Costa Martins, P.A., and De Windt, L.J. (2013). Regulation of fetal gene expression in heart failure. Biochim. Biophys. Acta *1832*, 2414–2424. https://doi.org/10.1016/j.bbadis.2013.07.023.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Di Domenico, M., Casadonte, R., Ricci, P., Santini, M., Frati, G., Rizzo, A., Carratelli, C.R., Lamberti, M., Parrotta, E., Quaresima, B., et al. (2012). Cardiac and skeletal muscle expression of mutant beta-myosin heavy chains, degree of functional impairment and phenotypic heterogeneity in hypertrophic cardiomyopathy. J. Cell Physiol. *227*, 3471–3476. https://doi.org/10.1002/jcp.24047.

Fan, P., Zhang, L., Cheng, T., Wang, J., Zhou, J., Zhao, L., Hua, C., and Xia, Q. (2021). MiR-590-5p inhibits pathological hypertrophy mediated heart failure by targeting RTN4. J. Mol. Histol. *52*, 955–964. https://doi.org/10.1007/s10735-021-10009-x.

Farbehi, N., Patrick, R., Dorison, A., Xaymardan, M., Janbandhu, V., Wystub-Lis, K., Ho, J.W.K., Nordon, R.E., and Harvey, R.P. (2019). Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. Elife *8*, 1–39. https://doi.org/10.7554/elife.43882.

Farrell, E., Armstrong, A.E., Grimes, A.C., Naya, F.J., De Lange, W.J., and Ralphe, J.C. (2018). Transcriptome analysis of cardiac hypertrophic growth in MYBPC3-null mice suggests early responders in hypertrophic remodeling. Front. Physiol. *9*, 1–14. https://doi.org/10.3389/fphys.2018.01442.

Gladka, M.M., Molenaar, B., de Ruiter, H., Van Der Elst, S., Tsui, H., Versteeg, D., Lacraz, G.P.A.A., Huibers, M.M.H.H., Van Oudenaarden, A., and Van Rooij, E. (2018). Single-cell sequencing of the healthy and diseased heart reveals cytoskeleton-associated protein 4 as a new modulator of fibroblasts activation. Circulation *138*, 166–180. https://doi.org/10.1161/circulationaha.117.030742.

Grun, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. Cell *163*, 799–810. https://doi.org/10.1016/j.cell.2015.10.039.

Grun, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640. https://doi.org/10.1038/nmeth.2930.

Grun, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., van Oudenaarden, A., Grün, D., Lyubimova, A., et al. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255. https://doi.org/10.1038/nature14966.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Hu, P., Liu, J., Zhao, J., Wilkins, B.J., Lupino, K., Wu, H., and Pei, L. (2018). Single-nucleus transcriptomic survey of cell diversity and functional maturation in postnatal mammalian hearts. Genes Dev. *32*, 1344–1357. https://doi.org/10.1101/gad.316802.118.

Itoh, K., Ishii, T., Wakabayashi, N., and Yamamoto, M. (1999). Regulatory mechanisms of cellular response to oxidative stress. Free Radic. Res. *31*, 319–324. https://doi.org/10.1080/10715769900300881.

Kanatous, S.B., Mammen, P.P.A., Rosenberg, P.B., Martin, C.M., White, M.D., DiMaio, J.M., Huang, G., Muallem, S., and Garry, D.J. (2009). Hypoxia reprograms calcium signaling and regulates myoglobin expression. Am. J. Physiol. Cell Physiol. *296*, 393–402. https://doi.org/10.1152/ajpcell.00428.2008.

Kannan, M.B., Solovieva, V., and Blank, V. (2012). The small MAF transcription factors MAFF, MAFG and MAFK: current knowledge and perspectives. Biochim. Biophys. Acta Mol. Cell Res. *1823*, 1841–1846. https://doi.org/10.1016/j.bbamcr.2012.06.012.

Kannan, S., Miyamoto, M., Lin, B.L., Zhu, R.J., Murphy, S., Kass, D.A., Andersen, P., and Kwon, C. (2019). Large particle fluorescence-activated cell sorting enables high-quality single-cell RNA sequencing and functional analysis of adult cardiomyocytes. Circ. Res. *125*, 567–569. https://doi.org/10.1161/circresaha.119.315493.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol. Cell *58*, 610–620. https://doi.org/10.1016/j.molcel.2015.04.005.

Kolodziejczyk, S.M., Wang, L., Balazsi, K., Derepentigny, Y., Kothary, R., and Megeney, L.A. (1999). MEF2 is upregulated during cardiac hypertrophy and is required for normal post-natal growth of the myocardium. Curr. Biol. *9*, 1203–1206. https://doi.org/10.1016/s0960-9822(00)80027-5.

Kretzschmar, K., Post, Y., Bannier-Helaouet, M., Mattiotti, A., Drost, J., Basak, O., Li, V.S.W., van den Born, M., Gunst, Q.D., Versteeg, D., et al. (2018). Profiling proliferative cells and their progeny in damaged murine hearts. Proc. Natl. Acad. Sci. U S A *115*, E12245–E12254. https://doi.org/10.1073/pnas.1805829115.

Larson, A., and Chin, M.T. (2021). A method for cryopreservation and single nucleus RNA-sequencing of normal adult human interventricular septum heart tissue reveals cellular diversity and function. BMC Med. Genomics *14*, 161–168. https://doi.org/10.1186/s12920-021-01011-z.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

Lim, D.S., Roberts, R., and Marian, A.J. (2001). Expression profiling of cardiac genes in human hypertrophic cardiomyopathy: insight into the pathogenesis of

phenotypes. J. Am. Coll. Cardiol. *38*, 1175–1180. https://doi.org/10.1016/s0735-1097(01)01509-1.

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., et al. (2020). Cells of the adult human heart. Nature *588*, 466–472. https://doi.org/10.1038/s41586-020-2797-4.

Long, P.A., Larsen, B.T., Evans, J.M., and Olson, T.M. (2015). Exome sequencing identifies pathogenic and modifier mutations in a child with sporadic dilated cardiomyopathy. J. Am. Heart Assoc. *4*, 1–11. https://doi.org/10.1161/jaha.115.002443.

Man, J., Barnett, P., and Christoffels, V.M. (2018). Structure and function of the Nppa-Nppb cluster locus during heart development and disease. Cell Mol. Life Sci. *75*, 1435–1444. https://doi.org/10.1007/s00018-017-2737-0.

Marian, A.J. (2010). Hypertrophic cardiomyopathy: from genetics to treatment. Eur. J. Clin. Invest. *40*, 360–369. https://doi.org/10.1111/j.1365-2362.2010.02268.x.

Marian, A.J., and Braunwald, E. (2017). Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. Circ. Res. *121*, 749–770. https://doi.org/10.1161/circresaha.117.311059.

Maron, B.J., and Maron, M.S. (2013). Hypertrophic cardiomyopathy. Lancet *381*, 242–255. https://doi.org/10.1016/s0140-6736(12)60397-3.

Maron, B.J., Ommen, S.R., Semsarian, C., Spirito, P., Olivotto, I., and Maron, M.S. (2014). Hypertrophic cardiomyopathy. J. Am. Coll. Cardiol. *64*, 83–99. https://doi.org/10.1016/j.jacc.2014.05.003.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. *17*, 10. https://doi.org/10.14806/ej.17.1.200.

Mirtschink, P., and Krek, W. (2016). Hypoxia-driven glycolytic and fructolytic metabolic programs: pivotal to hypertrophic heart disease. Biochim. Biophys. Acta Mol. Cell Res. *1863*, 1822–1828. https://doi.org/10.1016/j.bbamcr.2016.02.011.

Montag, J., Kowalski, K., Makul, M., Ernstberger, P., Radocaj, A., Beck, J., Becker, E., Tripathi, S., Keyser, B., Muhlfeld, C., et al. (2018). Burst-like transcription of mutant and wildtype MYH7-alleles as possible origin of cell-to-cell contractile imbalance in hypertrophic cardiomyopathy. Front. Physiol. *9*, 359. https://doi.org/10.3389/fphys.2018.00359.

Morgan, M., Falcon, S., and Gentleman, R. (2021). GSEABase: Gene Set Enrichment Data Structures and Methods.

Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.P., and van Oudenaarden, A. (2016). A single-cell transcriptome atlas of the human pancreas. Cell Syst. *3*, 385–394.e3. https://doi.org/10.1016/j.cels.2016.09.002.

Nomura, S., Satoh, M., Fujita, T., Higo, T., Sumida, T., Ko, T., Yamaguchi, T., Tobita, T., Naito, A.T., Ito, M., et al. (2018). Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure. Nat. Commun. *9*, 4435. https://doi.org/10.1038/s41467-018-06639-7.

Numazawa, S., Ishikawa, M., Yoshida, A., Tanaka, S., and Yoshida, T. (2003). Atypical protein kinase C mediates activation of NF-E2-related factor 2 in response to oxidative stress. Am. J. Physiol. Cell Physiol. *285*, C334–C342. https://doi.org/10.1152/ajpcell.00043.2003.

Ortega, A., Roselló-Lletí, E., Tarazón, E., Molina-Navarro, M.M., Martínez-Dolz, L., González-Juanatey, J.R., Lago, F., Montoro-Mateos, J.D., Salvador, A., Rivera, M., and Portoles, M. (2014). Endoplasmic reticulum stress induces different molecular structural alterations in human dilated and ischemic cardiomyopathy. PLoS One *9*, e107635. https://doi.org/10.1371/journal.pone.0107635.

Parbhudayal, R.Y., Garra, A.R., Götte, M.J.W., Michels, M., Pei, J., Harakalova, M., Asselbergs, F.W., van Rossum, A.C., van der Velden, J., and Kuster, D.W.D. (2018). Variable cardiac myosin binding protein-C expression in the myofilaments due to MYBPC3 mutations in hypertrophic cardiomyopathy. J. Mol. Cell. Cardiol. *123*, 59–63. https://doi.org/10.1016/j.yjmcc.2018.08.023.

Passier, R., Zeng, H., Frey, N., Naya, F.J., Nicol, R.L., McKinsey, T.A., Overbeek, P., Richardson, J.A., Grant, S.R., and Olson, E.N. (2000). CaM kinase signaling induces cardiac hypertrophy and activates the MEF2 transcription factor in vivo. J. Clin. Invest. *105*, 1395–1406. https://doi.org/10.1172/jci8551.

Pathak, G.P., Shah, R., Kennedy, B.E., Murphy, J.P., Clements, D., Konda, P., Giacomantonio, M., Xu, Z., Schlaepfer, I.R., and Gujar, S. (2018). RTN4 knockdown dysregulates the AKT pathway, destabilizes the cytoskeleton, and enhances paclitaxel-induced cytotoxicity in cancers. Mol. Ther. *26*, 2019–2033. https://doi.org/10.1016/j.ymthe.2018.05.026.

Qin, Q., Fan, J., Zheng, R., Wan, C., Mei, S., Wu, Q., Sun, H., Zhang, J., Brown, M., Meyer, C.A., et al. (2020). Inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data. Genome Biol. *21*, 32.

dos Remedios, C.G., Lal, S.P., Li, A., McNamara, J., Keogh, A., Macdonald, P.S., Cooke, R., Ehler, E., Knöll, R., Marston, S.B., et al. (2017). The Sydney Heart Bank: improving translational research while eliminating or reducing the use of animal models of human heart disease. Biophys. Rev. *9*, 431–441. https://doi.org/10.1007/s12551-017-0305-3.

Remme, C.A., Verkerk, A.O., Hoogaars, W.M.H., Aanhaanen, W.T.J., Scicluna, B.P., Annink, C., van den Hoff, M.J., Wilde, A.A.M., van Veen, T.A., Veldkamp, M.W., et al. (2009). The cardiac sodium channel displays differential distribution in the conduction system and transmural heterogeneity in the murine ventricular myocardium. Basic Res. Cardiol. *104*, 511–522. https://doi.org/10.1007/s00395-009-0012-8.

Ren, Z., Yu, P., Li, D., Li, Z., Liao, Y., Wang, Y., Zhou, B., and Wang, L. (2020). Single-cell reconstruction of progression trajectory reveals intervention principles in pathological cardiac hypertrophy. Circulation *141*, 1704–1719. https://doi.org/10.1161/circulationaha.119.043053.

Sasagawa, S., Nishimura, Y., Okabe, S., Murakami, S., Ashikawa, Y., Yuge, M., Kawaguchi, K., Kawase, R., Okamoto, R., Ito, M., and Tanaka, T. (2016). Downregulation of GSTK1 is a common mechanism underlying hypertrophic cardiomyopathy. Front. Pharmacol. *7*, 1–13. https://doi.org/10.3389/fphar.2016.00162.

Schauer, A., Adams, V., Poitz, D.M., Barthel, P., Joachim, D., Friedrich, J., Linke, A., and Augstein, A. (2019). Loss of Sox9 in cardiomyocytes delays the onset of cardiac hypertrophy and fibrosis. Int. J. Cardiol. *282*, 68–75. https://doi.org/10.1016/j.ijcard.2019.01.078.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. Bmc Mol. Biol. *7*, 3. https://doi.org/10.1186/1471-2199-7-3.

Selewa, A., Dohn, R., Eckart, H., Lozano, S., Xie, B., Gauchat, E., Elorbany, R., Rhodes, K., Burnett, J., Gilad, Y., et al. (2020). Systematic comparison of high-throughput single-cell and single-nucleus transcriptomes during cardiomyocyte differentiation. Sci. Rep. *10*, 1535–1613. https://doi.org/10.1038/s41598-020-58327-6.

Semsarian, C., Ingles, J., Maron, M.S., and Maron, B.J. (2015). New perspectives on the prevalence of hypertrophic cardiomyopathy. J. Am. Coll. Cardiol. *65*, 1249–1254. https://doi.org/10.1016/j.jacc.2015.01.019.

Sereti, K.I., Nguyen, N.B., Kamran, P., Zhao, P., Ranjbarvaziri, S., Park, S., Sabri, S., Engel, J.L., Sung, K., Kulkarni, R.P., et al. (2018). Analysis of cardiomyocyte clonal expansion during mouse heart development and injury. Nat. Commun. *9*, 754. https://doi.org/10.1038/s41467-018-02891-z.

Sheikh, F., Lyon, R.C., and Chen, J. (2015). Functions of myosin light chain-2 (MYL2) in cardiac muscle and disease. Gene *569*, 14–20. https://doi.org/10.1016/j.gene.2015.06.027.

Shim, W.J., Sinniah, E., Xu, J., Vitrinel, B., Alexanian, M., Andreoletti, G., Shen, S., Sun, Y., Balderson, B., Boix, C., et al. (2020). Conserved epigenetic regulatory logic infers genes governing cell identity. Cell Syst. *11*, 625–639.e13. https://doi.org/10.1016/j.cels.2020.11.001.

Skelly, D.A., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., and Pinto, A.R. (2018). Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. Cell Rep. *22*, 600–610. https://doi.org/10.1016/j.celrep.2017.12.072.

Smith, P.J., Wiltshire, M., and Errington, R.J. (2004). DRAQ5 labeling of nuclear DNA in live and fixed cells. Curr. Protoc. Cytom. *Chapter 7*, Unit 7.25. https://doi.org/10.1002/0471142956.cy0725s28.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499.

Sorenson, D., Sonnenblick, E.H., Robinson, T.F., and Capasso, J.M. (1985). Size and shape of enzymatically isolated ventricular myocytes from rats and cardiomyopathic hamsters. Cardiovasc. Res. *19*, 793–799. https://doi.org/10.1093/cvr/19.12.793.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Iii, W.M.M., Stoeckius, M., Smibert, P., and Satija, R. (2018). Comprehensive integration of single cell data. Preprint at bioRxiv, 1–24. https://doi.org/10.1101/460147.

Tan, W.L.W., Lim, B.T.S., Anene-Nzelu, C.G.O., Ackers-Johnson, M., Dashi, A., See, K., Tiang, Z., Lee, D.P., Chua, W.W., Luu, T.D.A., et al. (2017). A landscape of circular RNA expression in the human heart. Cardiovasc. Res. *113*, cvw250–cvw309. https://doi.org/10.1093/cvr/cvw250.

van der Velden, J., Merkus, D., de Beer, V., Hamdani, N., Linke, W.A., Boontje, N.M., Stienen, G.J.M., and Duncker, D.J. (2011). Transmural heterogeneity of myofilament function and sarcomeric protein phosphorylation in remodeled myocardium of pigs with a recent myocardial infarction. Front. Physiol. *2*, 83. https://doi.org/10.3389/fphys.2011.00083.

Vigil-Garcia, M., Demkes, C.J., Eding, J.E.C., Versteeg, D., De Ruiter, H., Perini, I., Kooijman, L., Gladka, M.M., Asselbergs, F.W., Vink, A., et al. (2021). Gene expression profiling of hypertrophic cardiomyocytes identifies new players in pathological remodelling. Cardiovasc. Res. *117*, 1532–1545. https://doi.org/10.1093/cvr/cvaa233.

Wang, L., Yu, P., Zhou, B., Song, J., Li, Z., Zhang, M., Guo, G., Wang, Y., Chen, X., Han, L., and Hu, S. (2020). Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. Nat. Cell Biol *22*, 108–119. https://doi.org/10.1038/s41556-019-0446-7.

Wang, T.Y., Lee, D., Fox-Talbot, K., Arking, D.E., Chakravarti, A., and Halushka, M.K. (2018). Cardiomyocytes have mosaic patterns of protein expression. Cardiovasc. Pathol. *34*, 50–57. https://doi.org/10.1016/j.carpath.2018.03.002.

Zhang, X., Azhar, G., Chai, J., Sheridan, P., Nagano, K., Brown, T., Yang, J., Khrapko, K., Borras, A.M., Lawitts, J., et al. (2001). Cardiomyopathy in transgenic mice with cardiac-specific overexpression of serum response factor. Am. J. Physiol. Hear. Circ. Physiol. *280*, 1782–1792. https://doi.org/10.1152/ajpheart.2001.280.4.h1782.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Titin (TTN)	Developmental Studies Hybridoma Bank	Cat# 9D10; RRID: AB_528491
ANP	Millipore	Cat# CBL66; RRID: AB_2283096
MLC-2V	Synaptic Systems	Cat# 310-003; RRID: AB_10645889
BrightVision poly-AP anti-rabbit IgG antibody	Immunologic	Cat# VWRKDPVR110AP; Cat# DPVR110AP
Alexa 647-labeled secondary antibody	ThermoFisher Scientific	A-21443; RRID: AB_2535861
Biological samples		
Cardiac tissue from patients with HCM (HCM5, .., HCM258)	This paper	N/A
Healthy human cardiac tissue (Control1)	Biochain	B607033
Healthy human cardiac tissue (Control2)	Biochain	B711068
Healthy human cardiac tissue (Control3)	Biochain	B711065
Healthy human cardiac tissue (Control4)	Biochain	A504241
Healthy human cardiac tissue (Control5)	Sydney Heart Bank	4062
Healthy human cardiac tissue (Control6)	Sydney Heart Bank	4104
Healthy human cardiac tissue (Control7)	Sydney Heart Bank	6008
Healthy human cardiac tissue (Control8)	Sydney Heart Bank	7054
Healthy human cardiac tissue (Control9)	Sydney Heart Bank	8004
Chemicals, peptides, and recombinant proteins		
DMEM, high glucose, GlutaMAX Supplement, pyruvate	Gibco	#31966021
DAPI	Invitrogen	#D3571
DRAQ5	eBioscience	65-0880-92
HEPES	Sigma Aldrich	H3375-250G
2,3-Butanedione monoxime (BDM)	Sigma Aldrich	B0753-25G
Taurine	Sigma Aldrich	T0625-10G
Liberase TL	Roche	5401020001
Dnase I	Worthington	LK003172
Liquid permanent red	Agilent Dako	K0640
Clearvue Mountant Xyl	ThermoFisher Scientific	89207-160
Prolong gold antifade	Life Technologies	P36934
TRIzol reagent	Invitrogen	#15596026
IQ™ SYBR Green Supermix	Bio-Rad	#170-8885
WGA	ThermoFisher Scientific	W11261
Critical commercial assays		
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	#4368813
TaqMan Universal Master Mix II	Applied Biosystems	#4440040
RNA 6000 Pico chips	Agilent	#5067-1513
iScript cDNA Synthesis Kit	Bio-Rad	#1708891
Deposited data		
scRNA-seq data	This paper	GSE138262
scRNA-seq data	(Wang et al., 2020)	GSE121893 and GSE109816
scRNA-seq data	(Litviňuková et al., 2020)	ERP123138
Oligonucleotides		
Primers for RT-PCR, please see Table S11	Integrated DNA Technologies	N/A
Taqman specific probes for *MYH6*	ThermoFischer Scientific	Hs01101442_g1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Graph pad PRISM v9	Graphpad Prism Inc	https://www.graphpad.com/scientific-software/prism/
ImageJ v1.51	NIH	https://imagej.nih.gov/ij/
Adobe inDesign CC 2019	Adobe Systems Incorporated	https://www.adobe.com/
R Studio v. 4.1.0	RStudio	https://cran.r-project.org/
Cutadapt (v. 3.4)	(Martin, 2011)	https://anaconda.org/bioconda/cutadapt
Trim Galore (v. 0.6.6)	N/A	https://github.com/FelixKrueger/TrimGalore
Burrows-Wheeler Alignment Tool (v. 0.7.17-r1188)	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/
STAR (v. 2.7.8a)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
featureCounts (v. 2.0.1)	(Liao et al., 2014)	http://subread.sourceforge.net/
UMI-tools (v. 1.1.1)	(Smith et al., 2017)	https://github.com/CGATOxford/UMI-tools
Seurat (v. 4.0.3)	(Stuart et al., 2018)	https://www.satijalab.org/seurat
Homer suite (v 4.1156)	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/
Lisa (v. 2.2.5)	(Qin et al., 2020)	http://lisa.cistrome.org/
Triage	(Shim et al., 2020)	https://github.com/woojunshim/TRIAGE
GSEABase (v. 1.54.0)	(Morgan et al., 2021)	https://bioconductor.org/packages/release/bioc/html/GSEABase.html
SCENIC	(Aibar et al., 2017)	http://scenic.aertslab.org
Custom scripts	This paper	https://doi.org/10.5281/zenodo.6282659
Other		
Artisan Link Pro	Agilent	N/A
Axiovert 40C	Zeiss	N/A
Bioanalyzer 2100	Agilent	N/A
FACS Aria III	Bioscience	N/A
HM 355S Automatic Microtome	Thermo Scientific	#905200
Nanodrop 1000 spectrophotometer	ThermoFisher Scientific	N/A
SPE Confocal Microscope	Leica	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Eva van Rooij (e.vanrooij@hubrecht.eu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- RNA-sequencing data have been deposited at Gene Expression Omnibus and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. This paper also analyzes existing, publicly available data. The accession numbers for these datasets are listed in the key resources table. Microscopy data reported in this paper will be shared by the lead contact upon request.
- Analysis scripts have been deposited at GitHub and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human heart samples

Cardiac tissue from the interventricular septum was obtained from myectomy surgery in patients with HCM to relieve left ventricular outflow tract obstruction (n = 5 for scRNA-seq, of which n = 2 where used for additional index sorting analysis and n = 97 for real-time PCR (RT-PCR) analysis), see Table S1 for detailed information about the patients including their sex/age. Written informed consent was obtained from each patient before surgery, and approval for the use of human tissue samples was obtained from the Medical Research Ethics Committee of the Erasmus Medical Center Rotterdam (fresh myectomy samples, and bulk myectomy RNA samples) or the scientific advisory board of the biobank of the University Medical Center Utrecht (explanted heart tissue for histology). 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 and A504241), the other five were obtained from the Sydney Heart Bank (identifiers: 4062, 4104, 6008, 7054, 8004) (dos Remedios et al., 2017), see Table S1 for additional details.

METHOD DETAILS

Histology and immunohistochemistry

Myectomy samples were fixed in PFA (4%) and incubated for 48 h rotating at room temperature. Next, the tissues were washed 3× for 10 min in PBS and stored in 70% ethanol (EtOH) at 4°C. For tissue embedding, 3 consecutive incubations were performed. Firstly, 96% EtOH for 2 h at 4°C, followed by 100% EtOH for 2 h at 4°C and lastly Xylene for 2 h at 4°C. Finally, the tissue was incubated overnight in liquid paraffin at 60°C and embedded in paraffin blocks. Then, 5 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°C. 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°C with primary antibodies against titin (TTN, 9D10, Developmental Studies Hybridoma Bank, 1:400), ANP (CBL66, Millipore, 1:800) and MLC-2 (MLC-2V, 310-003, Synaptic Systems, 1:500). For immunochemistry, after 1× 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 MLC-2. After 3× 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 3× 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 before (Gladka et al., 2018). 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 wheeling the vial for 15 min in a 37°C 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°C 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 gating was performed according to our previously optimized protocol (Gladka et al., 2018). Briefly, cytometry was performed on a FACS Aria III (BD Biosciences) using a 130 μm nozzle. Debris was excluded based on forward (FSC-A) and side scatter (SSC-A). area. Cells were selected for autofluorescence between 530 nm and 600 nm. Using FSC-W, the larger cells were selected in order to sort CMs rather than other cell types. Cells were single-cell sorted into 384-well plates, immediately centrifuged, and frozen at −80°C 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 were collected to correlate FSC-A (as a proxy for cell size) to gene expression on an individual cell basis.

To show our sorted cells were viable, we gated for DAPI-negative cells (DAPI #D3571, Invitrogen, 1:1000). To ensure our DAPI-negative events were nucleated, we next counterstained our sorted DAPI-negative cells with DRAQ5 (65-0880-92, eBioscience, 1:1000) and re-analyzed the cells by FACS (Smith et al., 2004).

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 100 μL 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 previously (Gladka et al., 2018; Grun and van Oudenaarden, 2015; Muraro et al., 2016) 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 6 bp 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°C, 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 from 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°C. 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. 5% of the sequencing run ($15 \cdot 10^6$ reads) were allotted to each 384-well plate. Four 384-well plates were sequenced per patient, two for the patient where index-sorting was applied.

Mapping of HCM sequencing data

After sequencing, each of the HCM samples were mapped. Read 1 (R1) contains respectively the UMI (positions 1–6) and barcode (positions 7–14) followed by the poly-T sequence (theoretically 24 nt) and sample mRNA sequence, read 2 (R2) contains the mRNA transcript sequence only. UMI and cell barcode (BC) information was attached to R2 metadata, and only R2 was used for mapping. In addition, we discarded reads that did not show $\geq$10 thymine nucleotides after the BC in R1. Reads were then trimmed for sequencing primers and repeating single nucleotides using the Cutadapt (Martin, 2011) wrapper Trim Galore (default settings). We then filtered out ribosomal RNA reads by discarding reads that mapped to the human rRNA genes using Burrows-Wheeler Alignment Tool (bwa merge of bwa aln and bwa mem -h 15) (Li and Durbin, 2009). These pre-processed reads were then mapped using Spliced Transcripts Alignment to a Reference (STAR –outFilterMultimapNmax 20) (Dobin et al., 2013), assigned to genes using featureCounts (-R BAM) (Liao et al., 2014), where multi-mappers were ignored (featureCounts default behavior), and processed into UMI count tables using UMI-tools (umi_tools count; default settings) (Smith et al., 2017). As reference genome, we used the human genome (GRCh38.93) and gene annotation (gtf file) acquired from ensembl.org. The gtf file was filtered to only contain genes with a gene_biotype annotation of protein_coding, lincRNA or antisense. After mapping, HCGN symbols were linked to Ensembl IDs using the R biomaRt package.

Obtaining Ctrl1 and Ctrl2 count tables

We included cells from healthy donor hearts from two different studies, Wang et al. (Wang et al., 2020) (referred to as Ctrl1) and Litviňuková et al. (Litviňuková et al., 2020) (Ctrl2). From Wang et al., we used samples N1, N2, N3, N4, N5, N13 and N14, as they came from healthy donors and included LV cells. We obtained raw FASTQ files from Gene Expression Omnibus (GEO, accession numbers GSE121893 and GSE109816) using the NCBI SRA Toolkit (fasterq-dump command) and SRR identifiers obtained from the provided metadata files. We collected FASTQ data for cells annotated as LV cells (clusters LV1-9) into files per donor and plate (we note that these selected cells were filtered for quality already by Wang et al.). These data were collected with the iCell8 platform from Takara Bio, R2 again contained the biological information that was mapped, and R1 contained only the BC (11 nt) followed by the UMI (ranging from 10-14 nt), of which we used only the first 10 nucleotides to deal with the variable length. (A reference BC file was obtained from the Cogent NGS Analysis Pipeline, v. 1.0, Takara Bio.) We mapped the obtained FASTQ files using the same pipeline as the HCM samples. For Litviňuková et al. (who used 10× Genomics) we downloaded count tables for CMs for all available samples (D1-7, D11, H2-7) from https://www.heartcellatlas.org/ (h5ad files) and selected only cells taken from the septum. These tables were generated by the Cell Ranger pipeline (v. 3.0.2, 10× Genomics), which, identical to our pipeline, uses the STAR aligner in combination with human genome GRCh38.93 and gene annotation (ensembl.org), and only considers uniquely mapped reads (10× Genomics documentation). The downloaded Litviňuková et al. raw count tables were converted to h5seurat format and used directly for analysis in Seurat.

Analysis of single cell transcriptomics data using seurat

We used Seurat to analyze our data (Seurat v. 4.0.3, SeuratDisk v. 0.0.0.9019, SeuratObject v. 4.0.2, loomR v. 0.2.1.9000, hdf5r v. 1.3.3) (Stuart et al., 2018). Raw count tables were loaded into Seurat for all three datasets (HCM, Ctrl1, Ctrl2), and the Seurat "merge"

function was applied to perform pooled analysis. Separate Seurat objects based on this data were also created for separate donors. Similar Seurat analyses were then performed for the pooled data, the HCM, Ctrl1, Ctrl2 sets, and donors separately. Mitochondrial read counts were determined, after which mitochondrial genes were removed from the count table. Cells were then filtered based on total read count, cells with total UMI count >1000 were taken along (we note that for Ctrl1 and Ctrl2 data, cell quality filtering was already performed by respective authors). Genes detected in <5 cells were discarded. Total UMI counts were 1780041, 1614560, 2028833, 396894 and 1418990 for patients 1–5 respectively (excluding mitochondrial counts), corresponding to mean UMI counts per well of 3779, 2660, 3241, 2795 and 3182. Then, to perform the analysis, the following Seurat functions were applied with default parameters unless explicitly stated: NormalizeData (with normalization.method = 'RC', and median total cell transcript count as scale.factor), FindVariableFeatures, ScaleData (with do.scale = F, do.center = F, scale.max = Inf), RunPCA (npcs = 30, using the top 2000 variable features), RunUMAP, FindNeighbors, FindClusters (resolution = 0.1 or 0.4 for pooled and HCM datasets, respectively). We use the FindMarkers function (min.pct = 0.05) to identify differentially expressed genes for the clusters. For gene expression plots, the scale was set from 0 to the 97th or 98th percentile for UMAPs and violin plots respectively, to exclude outliers from the scale. To determine patient-level gene expression values for genes, gene expression values were first normalized to Z-scores and then averaged per patient. To determine composite gene expression values for groups of genes (such as regulons or modules, see below), we also first normalized expression to Z-scores, and then averaged over the genes belonging to the group of interest. To determine patient-averaged composite expression for groups of genes, these two strategies were combined.

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*[57]. A background set of genes was constructed by selecting all genes expressed in >5% of cells. A set of genes of interest was then determined based on genes that were significantly enriched (FC > 0, adjusted $p < 0.05$) between cluster and non-cluster, or by their presence in a gene module (see below). Enrichment of GO terms was then analyzed using the tools from the *GSEABase* R-package (v. 1.54.0) (Morgan et al., 2021). The pvalueCutoff parameter of the hyperGTest function was set to 0.05.

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 10% (gene-gene correlations) or 33% (gene-FSC-A correlations) 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 (using the Fisher z-transformation). A p value correction was applied using the p.adjust function (Benjamini & Hochberg method).

Transcription and regulatory factor analyses

To identify transcription or regulatory factors that drive expression of groups of genes (such as cluster-enriched genes or modules), we ran Homer, Lisa and/or TRIAGE. HOMER (Homer suite v4.11; human-o v6.3 and human-p v5.5) (Heinz et al., 2010) performs TF motif enrichment analysis in proximal promoters. We used the findMotifs.pl function using all expressed genes (expressed in at least 5% of all included cells) as background and promoter location set to −300 to +50 bp (default options). We then show top 5 significant results ($p < .05$ and $q < 0.9$; if available) from the "known results". Lisa (v. 2.2.5) (Qin et al., 2020) leverages existing H3K27ac ChIP-seq data from transcription regulators to identify regulators that potentially drive gene sets of interest. We ran the "oneshot" Lisa procedure (genome hg38) on gene sets of interest with the "–rp_map enhanced_10K" setting and provided a background list of genes (all genes that are expressed in 5% of cells). We then select significant hits ($p < 0.01$) and show a list that consists of both top-5 hits (by p value only) and top-5 of hits that are observed >1 cell type or CM-specific; hits occurring in both top-fives are indicated with an asterisk. TRIAGE (Shim et al., 2020) leverages H3K27me3 CHIP-seq data to calculate repressive tendency scores (RTS) and discordance scores for genes, with the aim of identifying regulatory genes that control differentiation in a set of genes. To identify potential regulators for each cell cluster, we took differentially expressed genes (FC > 0 and adjusted $p < 0.01$) with an RTS >0.03 (deemed priority genes by Shim et al.), and listed the 5 (if available) regulators with the highest RTS score.

SCENIC analysis

To identify regulons, we converted raw count tables into loom files for all of the separate patients and donors from the HCM, Ctrl1 and Ctrl2 datasets, and ran a command line interface (CLI) SCENIC pipeline using "pyscenic" (Aibar et al., 2017). First, we executed the GRN step, using the "pyscenic grn -o adj.csv" command, supplying the hs_hgnc_curated_tfs.txt TF list from the pyscenic resource directory. We then executed the "pyscenic ctx" command, where we supplied the hg19-tss-centered-10kb-7species.mc9nr.feather and hg19-500bp-upstream-7species.mc9nr.feather databases (obtained via https://resources.aertslab.org/cistarget/) and the motifs-v9-nr.hgnc-m0.001-o0.0.tbl motif file (obtained via https://resources.aertslab.org/cistarget/motif2tf/). Finally, the "pyscenic aucell" step was executed. We processed output files further in R. We extracted Normalized Enrichment Scores (NES) from the reg.csv output file, from which we determined the median NES score for each respective regulon; the NES score reflects confidence in the link between the TF and gene set. To merge regulons from multiple patients with HCM, we selected those genes that are present in 3 or more patients. Per regulon, we also extracted importance scores for each of the linked genes from the adj.csv output file, which we

use to rank the genes within the (merged) regulons (here also the median was used to determine importance scores from multiple patients/donors).

Custom gene module 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 of each patient separately. We first selected genes that are expressed in at least 20% of cells pooled for all Patients with HCM. Then, for each patient, we selected genes expressed in 5% of that patient's cells, and calculated Pearson's correlations coefficients between all remaining gene pairs, 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 ($p < 0.001$) correlation with at least 10 other genes, resulting in the correlation matrix that was used for the regulon analysis. We applied the hclust function (using the ward.D2 method) to sort rows and columns of the correlation matrix by similarity, and generate a hierarchical tree. Using this tree, we calculated the gap statistic for a range of number of clusters (k), and determined the optimal numbers of clusters (K). (We used the clusGap and maxSE function from the R cluster package, applying the Tibs2001SEmax method.) Genes were assigned to K co-expression modules based on the cutree function, resulting 6, 5, 11, 5 and 6 modules for patients 1–5 respectively. Within each module, we ranked genes according to the average correlation to other genes in each module.

To determine final modules presented in the paper, we systematically compared the patient-specific modules for overlap in member genes. We defined gene identity overlap for a pair of modules simply as the number of overlapping genes divided by the size of the smallest regulon in the pair. We used the hclust function again to generate a dendrogram, which by eye very clearly showed 5 clusters, as also confirmed by a median gene overlap of 68%, 86%, 62%, 89% and 82% between patients for each of the module groups. We determined our final five modules by joining patient-specific modules based on this clustering, and included genes into each of the final modules if they were identified in at least 3 patient-specific modules of that cluster.

RT-PCR analysis

Total RNA was isolated from 97 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 250 ng 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°C for 15 min, followed by 40 cycles at 95°C for 15 s, 60°C for 30 s and 72°C for 30 s. Ribosomal protein L32 (*RPL32*) was used as a housekeeping gene, to which all genes were normalized. Expression of *NPPA* and *NPPB* in HCM samples and control samples was calculated (-ΔCt). The results are shown as mean ± standard error of the mean (SEM). Significance was tested using Student's t test. Plotting was done using PRISM (GraphPad Software Inc.). Also for the correlation analyses, expression values normalized against *RPL32* were calculated (-ΔCt). All the PCR primers that were used are listed in Table S11. R and ggplot2 were used to plot and perform correlation analyses. Specifically, based on ΔCt-values of HCM-samples, the Pearson correlation coefficients were determined for gene pairs, and corresponding p values were calculated by using a t test (using the stats:cor.test function). The number of samples (n) used in each RT-PCR experiment is indicated in the legend or shown in the figures.

Cell size analysis

Cell size was manually measured on sections stained for MLC-2 and labelled with WGA by using ImageJ 1.49v software. Circular events were selected in the WGA channel, to quantify only cells whose cross-section is perpendicular to their long axis. ImageJ was used to measure the area of each identified cell, and fluorescence intensity in the MLC-2 channel within that area. Total MLC-2 fluorescence was normalized to cell area. Cell size and MLC-2 were quantified in 3–5 images each for 6 myectomy samples (326–669 cells per samples, 2931 cells in total).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical and quantitative analysis of single cell RNA-sequencing data

scRNA-seq was performed on n = 5 patients with HCM, and publicly available scRNA-seq data from healthy donors were included from two sources (Ctrl1, n = 7; Ctrl2, n = 14) as described in the Method details. Correlation coefficients for gene expression correlations and their significance were determined using the cor.test function in R, which calculates Pearson's correlation coefficient and a p value estimate based on a student-t distribution using a Fisher z-transformation as also described in the Method details. The signifance values were adjusted by the Benjamini & Hochberg method using the p.adjust function. To test the validity of statistical assumptions, we randomly drew series of expression values from our data, and determined correlation coefficients based on the random series; those coefficients showed a distribution consistent with the assumption that uncorrelated data follow a student-t distribution. As described in the Method details, to determine patient-level gene expression values for genes, gene expression values were first normalized to Z-scores and then averaged per patient. By average, we refer to the arithmetic average. To determine composite gene expression values for groups of genes, we also first normalized expression to Z-scores, and then averaged over the

genes belonging to the group of interest. To determine patient-averaged composite expression for groups of genes, these two strategies were combined. Patient-averages of (composite) gene expression were plotted and statistically analyzed using PRISM (GraphPad Software). To test for statistical significance of differences between average gene expression values in different conditions, outliers were first removed by using the ROUT test with Q = 1%, and subsequently a one-way ANOVA test was performed (*, $p \leq 0.05$; **, $p \leq 0.01$, ***, $p \leq 0.001$; ****, $p \leq 0.0001$). Error bars in plots show the SEM. Assumptions about the sample distribution were not tested. For the GO term analyses, the hyperGTest function from the *GSEABase* R-package was applied as described in the Method details.

Statistical and quantitative analysis of scRNA-seq and cell size data

To determine correlation coefficients between gene expression and FSC-A values, determined in n = 2 patients, we also used the cor.test and p.adjust function in R (again performing a t test, using a Fisher z-transformation, and a Benjamini & Hochberg correction for multiple testing). To determine correlation coefficients and their significance between total reads per cell and FSC-A values, values produced by the cor.test function were used without multiple sample adjustment.

Statistical and quantitative analysis of RT-PCR data

RT-PCR was applied to n = 9 control samples and n = 97 samples from patients with HCM. Not all samples gave a signal for all genes, samples that gave a signal were included in the analysis. For *NPPA* and *NPPB*, these data were plotted and statistically analyzed using PRISM (GraphPad Software). Outliers were removed by using the ROUT test with Q = 1%, and t-tests were performed (*, $p \leq 0.05$; **, $p \leq 0.01$, ***, $p \leq 0.001$; ****, $p \leq 0.0001$). For *NPPA*, n = 9 Ctrl and n = 49 HCM samples were analyzed, for *NPPB*, n = 9 Ctrl and n = 36 HCM samples were analyzed. Mean values and SEM are indicated in the plots. Correlation coefficients between pairs of genes and their significance were again determined by a t test using the cor.test and p.adjust function in R (which performed a t test, using a Fisher z-transformation, and a Benjamini & Hochberg correction for multiple testing). Assumptions about the sample distribution were not tested. For the correlation analyses between gene pairs consisting of *MYL2* and another gene, n-values for the respective pairs of *MYL2* and the listed gene were: *ACTA1*, 33; *ACTC1*, 65; *CKM*, 62; *COX6A2*, 68; *CRYAB*, 71; *CSRP3*, 64; *GAPDH*, 62; *HSPB1*, 61; *MB*, 66; *MYL3*, 59; *MYL9*, 74; *SLC25A3*, 59; *SLC25A4*, 66; *TNNC1*, 66; *TPM1*, 69; *UBC*, 52. In addition n = 9 Ctrl samples were analyzed for each pair.

Statistical and quantitative analysis of staining and microscopy images

Expression of MLC-2 was determined as described in the Method details for n = 6 patients, and plotted and statistically analyzed using PRISM. Outliers were removed by using the ROUT test with Q = 1%. Differences were tested by using a one-way ANOVA followed by a Tukey's multiple comparisons test (*, $p \leq 0.05$; **, $p \leq 0.01$, ***, $p \leq 0.001$; ****, $p \leq 0.0001$). Average values are indicated in the plots (arithmetic average).