The impact of genotype on the cellular architecture of dilated and arrhythmogenic cardiomyopathies

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von

Eric Lars-Helge Lindberg, M.Sc.

Präsidentin

der Humboldt-Universität zu Berlin

Prof. Dr. Julia von Blumenthal

Dekan der Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

Prof. Dr. Dr. Christian Ulrichs

Gutachter/innen

1. Prof. Dr. Markus Landthaler
2. Prof. Dr. Norbert Hübner
3. Prof. Dr. Christoph Lippert

Tag der mündlichen Prüfung: 23.02.2023

Selbstständigkeitserklärung

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten schriftlichen Arbeit mit dem Titel

The impact of genotype on the cellular architecture of dilated and arrhythmogenic cardiomyopathies

um eine von mir selbstständig und ohne fremde Hilfe verfasste Arbeit handelt. Ich erkläre ausdrücklich, dass ich sämtliche in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken u. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken u. Ä. anderer Autorinnen und Autoren (Paraphrasen) die Quelle angegeben habe. Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend der Prüfungsordnung und/oder der Allgemeinen Satzung fur Studien- und Prüfungsangelegenheiten der HU (ASSP) geahndet werden.

Datum .....09.11.2022..... Unterschrift .....Eric Lindberg.....

Zusammenfassung

Herzinsuffizienz ist ein klinisches Syndrom, welches durch funktionelle und strukturelle Anomalien des Herzens verursacht wird, und ist weltweit die häufigste Todesursache. Die dilatative Kardiomyopathie, welche durch eine Vergrößerung der linken Herzkammer definiert ist, und die arrhythmogene Kardiomyopathie, welche im Gegensatz durch eine Dysfunktion der rechten Herzkammer definiert ist, sind eine der häufigsten Ursachen für Herzinsuffizienz. Trotz vieler Bemühungen die molekularen Veränderungen der Herzinsuffizienz zu charakterisieren, sind Zelltypzusammensetzung, Genexpressionsänderungen, und zelluläre Interaktionen unter pathologischen Bedingungen unbekannt.

Um diese Fragen zu adressieren wurde ein Protokoll zur Isolation intakter Zellkerne entwickelt um Einzelkernsequenzierung im Herzen durchzuführen. Anschließend wurde mit dem entwickelten Protokoll die zelluläre Zusammensetzung des erwachsenen gesunden menschlichen Herzens charakterisiert. Hier war mein Fokus die Charakterisierung und Identifikation von Subformen von Fibroblasten, und deren Genexpressionsunterschiede in den linken und rechten Vorhöfen und Herzkammern. Basierend auf dieser Annotation wurden die Zelltypen und Subtypen von ungefähr 900.000 Zellkernen von 61 nicht-ischämischen Herzinsuffizienzpatienten mit unterschliedlichen pathogenen Varianten in DCM- und ACM-assoziierten Genen oder idiopathischen Erkrankungen charakterisiert und mit 18 gesunden Spenderherzen verglichen. Dieser Datensatz zeigte spezifische Unterschiede des linken und rechten Ventrikels mit differenziell regulierten Genen und Signalwegen, and Veränderungen in der Zusammensetzung der verschiedenen Zelltypen und Subtypen. Um genotyp-spezifische Antworten unabhängig zu bestätigen wurden Algorithmen des maschinellen Lernens angewendet, welche die zugehörige Genotyp-Untergruppe des Patienten mit hoher Genauigkeit vorhersagten. Zusammenfassend stellen die in dieser Arbeit veröffentlichten Daten das vorherrschende Dogma in Frage, dass Herzinsuffizienz auf einen gemeinsamen finalen Signalweg zurückzuführen ist.

Contents

List of Abbreviations

C	Celsius
4OH-P	hydroxproline
ACM	arrythmogenic cardiomyopathy
AD	adipocytes
AP	apex
ARVC	arrythmogenic right ventricular cardiomyopathy
BO	Bad Oeynhausen
cDC	classical dendritic cell
cDNA	complementary DNA
CLR	centered log-ratio
cm	centimeter
CM/vCM	cardiomyocytes/ventricular cardiomyocytes
CPU	central processing unit
CRT	cardiac resynchronisation therapy
DCM	dilated cardiomyopathy
DNA	desoxyribonucleic acid
EC	endothelial cells
ECM	extracellular matrix
FACS	fluorescence-activated single cell sorting
FB	fibroblast
FC	fold change
FDR	false discovery rate
FSC	forward scatter
FW	free wall
GEM	Gel beads in emulsion
GFR	glomerular filtration rate
h	hour
HB	homogenisation buffer
HCA	Heart Cell Atlas
HF	heart failure
HLA	human leukocyte antigens
ISG	interferon stimulated genes
LA	left atrium
LV	left ventricle
LVAD	left ventricular assist device
LVIDd/s	Left Ventricular inner diameter Diastolic/systolic
M	molar
MDC	Max-Delbrck Center for Molecular Medicine
MHC	major histocompatibility complexes
min	minutes
ml	milliliter
mm	millimeters of mercury
mM	millimolar
mmHg	millimeters of mercury
MP	Macrophage
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MY	myeloid
NC	neuronal cells
NGS	next-generation sequencing
OCT	Optimal cutting temperature compound
PBS	phosphate buffered saline
PC	pericytes
PCR	polymerase chain reaction
PV/Pvneg	Pathogenic variant/Pathogenic variant negative
Q3	third quartile
RA	right atrium
RIN	RNA integrity number
RNA	ribonucleic acid
RV	right ventricle
RVEDD	Right Ventricular End Diastolic Diameter
SB	storage buffer
SCD	sudden cardiac death
scRNAseq	single-cell RNA sequencing
SMC	smooth muscle cells
snRNAseq	single-nucleus RNA sequencing
SP	Septum
SSC	sideward scatter
U	Units
ul	microliter
uM	micromolar
umap	uniform manifold approximation and projection
UMI	unique molecular identifier
vFB	ventricular fibroblast

List of Figures

List of Tables

1 Graphical Abstract

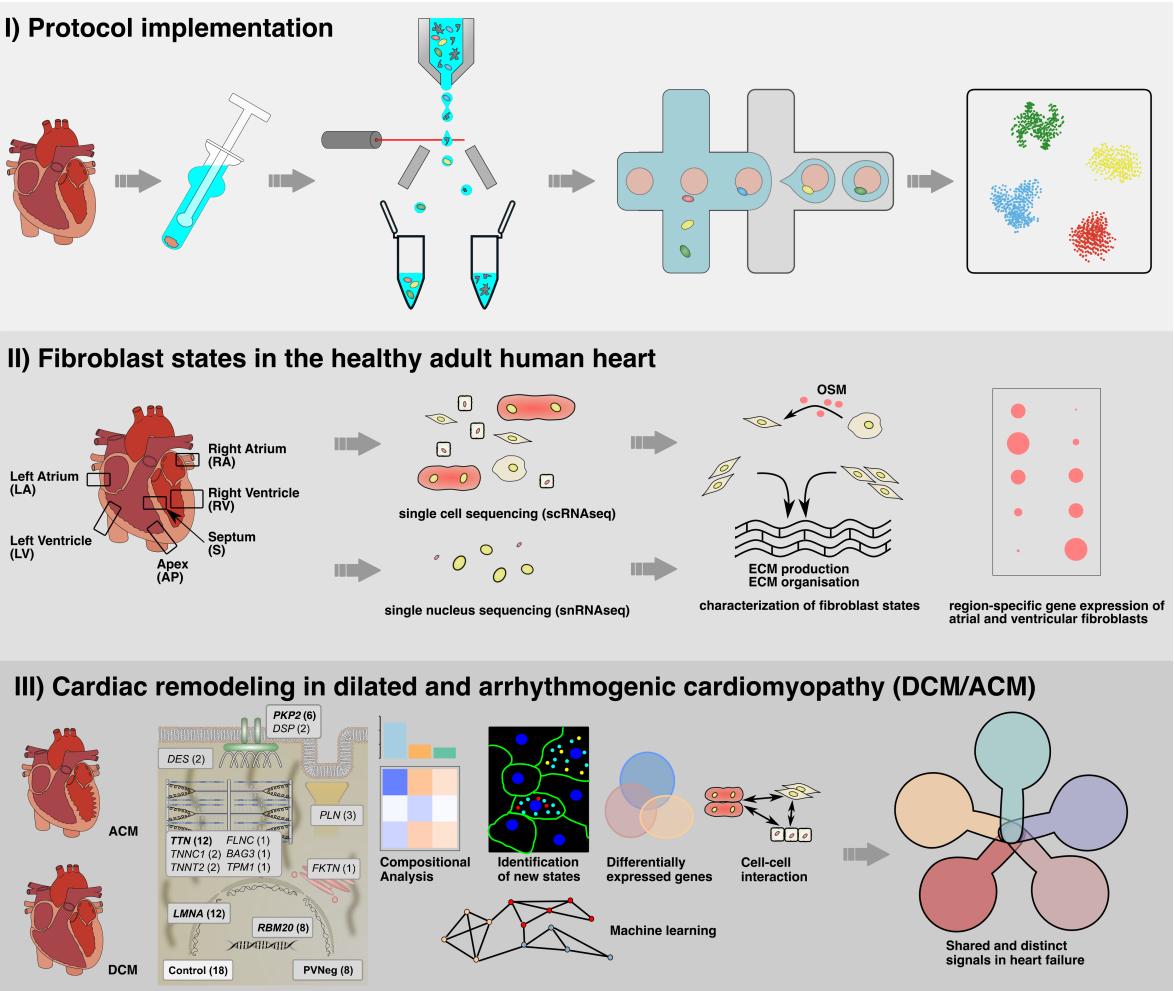

Figure 1: Graphical abstract

2 Summary

Heart failure is a clinical syndrom and leading cause of death worldwide, caused by functional and structural abnormalities of the heart. Dilated Cardiomyopathy, defined by a left ventricular enlargement, and arrhythmogenic cardiomyopathy, defined by a right ventricular dysfunction, are leading causes of heart failure. Despite previous efforts to characterise molecular changes in the failing heart, little is known on cell-type specific abundance and expression changes under pathological conditions, and how individual cell-types interact during heart failure and cardiac remodelling.

To address this question, a protocol for the isolation of intact nuclei was firstly established to perform robust single-nucleus RNA sequencing in the heart. Next, the cell-type composition of the healthy adult human heart was characterised. Here my focus was on the fibroblast nieche by characterising fibroblast states, their composition and their atria- and ventricle-specific expression patterns. Cell type and state annotation was then used to characterize the transcriptome of roughly 900,000 nuclei from 61 failing, non-ischemic human hearts with distinct pathogenic variants in DCM and ACM genes or idiopathic disease and compared those to 18 healthy donor hearts. This dataset revealed distinct responses of the right and left ventricle with differently regulated genes and pathways, and compositional changes across cell types and states. To independently confirm genotype-specific responses, machine learning approaches were applied, predicting genotype subgroups with high accuracy. Taken together, the findings published in this thesis upend the prevalent dogma that heart failure results in a final common pathway.

3 Introduction

3.1 Anatomy of the human heart

The heart is a multi-chamber organ pumping blood through the circulation (Figure 2). Within the four chambers of the human heart, biophysical stimuli vary dramatically during a heart beat for example due to large blood pressure differences between the chambers. Already 130 years ago, the surgeon Robert H. Woods reported a relationship between the density of muscle fibres, heart chamber volume, and cellular stress, relating to the law of Laplace, linking cellular architecture and gene activity with heart function (Woods, 1892). 23 years later, the first reports on pressures and volumes of the four chambers were reported by the American physiologist Carl J. Wiggers (Wiggers, 1915). The right atrium (RA) is the chamber with the thinnest wall (around 2mm) and lowest systolic pressure (2-6 mmHg). The LV in contrast has a reported wall thickness of 8-10mm and systolic pressure can reach up to 130 mmHg. The concept that the normal function of the human heart relies on highly heterogeneous cell populations with specialized functions that are governed by differential gene expression was then addressed multiple times in the past (Katz and Katz, 1989). The precise cellular composition in the adult healthy human heart and disease-associated changes remained incompletely understood. Despite previous efforts to characterize molecular changes in the failing heart, little is known on cell-type specific abundance and expression changes under physiological conditions, and how cell types interact during heart failure and cardiac remodelling.

3.2 Heart failure and cardiomyopathies

Heart failure (HF) is a clinical syndrome of different etiologies with symptoms caused by functional or structural abnormality, elevated blood levels of natriuretic peptide, and systemic congestion (Bozkurt et al., 2021). The estimated number of people living with heart failure worldwide is 64.3 million (James et al., 2018), with an estimated prevalence of known heart failure of 1-2% in the adult population of developed countries (Groenewegen et al., 2020). HF encompasses a broad spectrum of cardiac disorders, of which, among others, cardiomyopathies can be an underlying cause. Cardiomyopathies describe structural and functional heart muscle dysfunction and are subdivided using different criteria (Kumar et al., 2017). Broadly, cardiomyopathies are divided into ischemic and nonischemic cardiomyopathies. Ischemic cardiomyopathies are caused by the lack of oxygen supply, as observed in patients

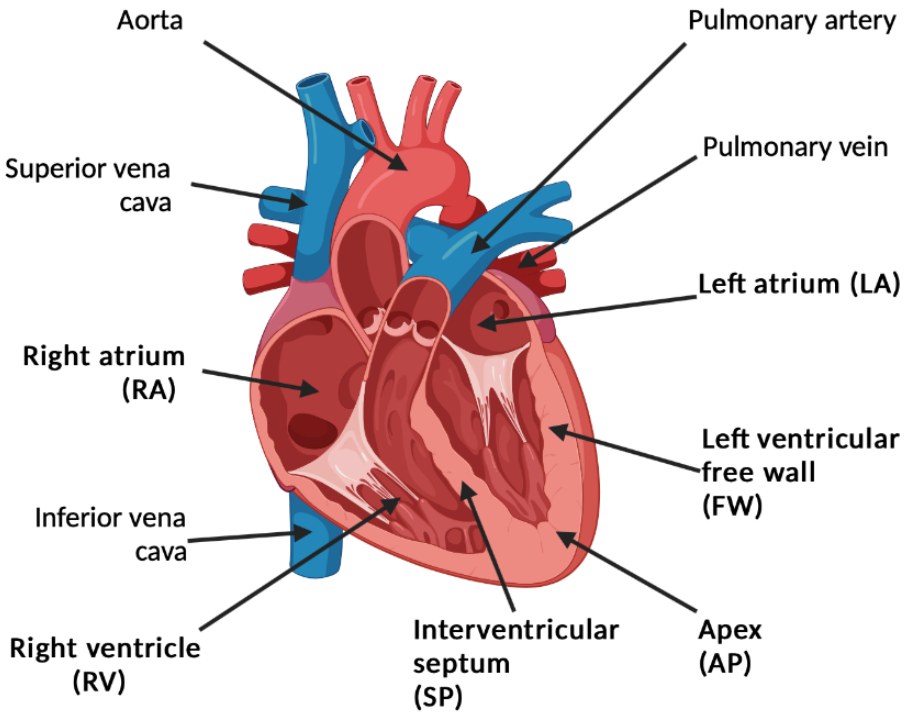

Figure 2: Cross-section of the adult human heart showing both atria, ventricles, and the interventricular septum. Created with BioRender.com

with coronary artery diseases (Felker et al., 2002), comprising about half of all heart failure cases (Adams Jr et al., 2005; Felker et al., 2002). As ischemic cardiomyopathies arise from secondary myocardial damage, for example, systemic hypertension or atherosclerosis in coronary arteries, the disease is not emerging from the myocardium itself (Maron et al., 2006). Two often occurring forms of nonischemic cardiomyopathies, dilated cardiomyopathy (DCM) and arryhthmogenic cardiomyopathy (ACM) were studied in detail in this work.

3.2.1 Dilated Cardiomyopathies (DCM)

DCM is characterized by a left- or biventricular chamber enlargment, systolic dysfunction with abnormal loading conditions (Maron et al., 2006; Pinto et al., 2016b) (Figure 3). Commonly used diagnostic methods include Echocardiography or cardiac MRI (Hershberger and Jordan, 2021). In the majority of cases DCM manifests in the fourth to sixth decade. Previous studies highlighted the higher occurance in males compared to females (with varying ratios of 3-4:1) (Lyden et al., 1987; Fairweather et al., 2013).

The reported prevalence estimate worldwide for DCM is 1:250 (Goldman and AI, 2019). Reported cases are not only high for the western world, but also for other continents. One example is the African continent, where DCM accounts for 17-48% of all cardiac conditions

observed at autopsies (Maharaj, 1991; Hakim and Manyemba, 1998; Amoah and Kallen, 2000). Causes for DCM are various, including genetics (McKenna and Judge, 2021), drugs such as alcohol and cocaine (Awtry and Philippides, 2010), or certain infectious diseases such as Coxsackievirus B and parvovirus B19 (Kindermann et al., 2008). The fraction of genetic DCMs was reported to be 25-35% (Kumar et al., 2017). Understanding the genetics of DCM is at the moment of high interest to understand predisposition and the epidemiology of the disease. In the past genetic studies focused on Mendelian patterns typically affecting younger patients, as interpreting the clinical genetics in adults is often more complex due to environmental factors, especially for genetic variants of lower penetrance. In a recent study adult and pediatric DCM were reported to have distinct features such as increased sarcomere thickness in adult DCM, absence of myocardial fibrosis in pediatric DCM, and microvascular alterations (Patel et al., 2017).

DCM associated genes encode a hetergenous group of proteins involved in contractile force transmission and generation, cytosceletal architecture, nuclear scaffolding, transcription, splicing, and electrolyte homeostasis (Burke et al., 2016) (Figure 4). Notably, most cardiomyopathy genes are important for biological processes in cardiomyocytes with few exceptions.

Despite first reports that there are differences between DCMs arising from different pathogenic variants, a non-specific diagnosis of genetic DCM is commonly made, instead of a specific, such as *LMNA*-DCM, *TTN*-DCM.

3.2.2 Arrythmogenic Cardiomyopathy (ACM)

Arryhthmogenic Cardiomyopathy (ACM) is characterized by electrical and myocardial abnormalities, arryhthmias and fibrofatty replacement of the apoptotic areas (Marcus et al., 1982; Elliott et al., 2019) (Figure 3). Commonly used diagnostic methods include echocardiography, cardiac MRI or electrocardiography. In the majority of cases ACM manifests in early adulthood. ACM accounts for 10% of sudden cardiac death (SCD) cases and is the leading SCD cause in athletes (D'Ascenzi et al., 2018). Higher mortality rates have been reported for males compared to females (Lin et al., 2017).

The reported prevalence estimate worldwide for ACM is 1:5000 (Goldman and AI, 2019). In more than 50% of the patients, a mutation in desmosomal proteins is detected (Groenewegen et al., 2020): *PKP2*, *DSG2*, and *DSP* (Figure 4).

Arryhthmogenic right ventricular cardiomyopathy (ARVC) is the most common form of inherited ACM, often without LV dysfunction until the late stages of the disease (McKenna

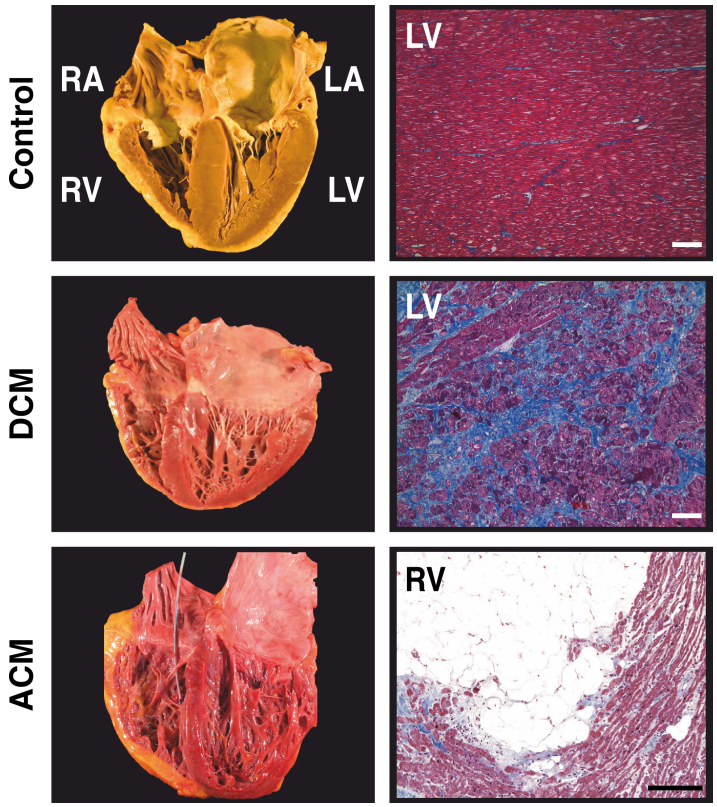

Figure 3: Comparison of morphology and pathology of healthy control, DCM and ACM hearts. (Left) Cross-sections show left ventricular dilation in DCM hearts and fibrofatty degeneration in the RV in ACM hearts. (Right) Masson trichrome staining. Cardiomyocytes are colored red, while fibrotic areas are shown in blue. Fibrofatty plagues are white (magnification 100x, bar 10m). Figure from Reichart et al. (2022).

and Judge, 2021). Although very rare, forms with LV- and biventricular involvement have been previously reported (Coats et al., 2009). Biventricular involvement has recently been incorporated in the 2020 international diagnostic criteria for ACM ('Padua criteria') (Corrado et al., 2020).

ACM, despite many similarities, has, in contrast to DCM, a more right ventricular phenotype, hence it is often referred to as arryhthmogenic right ventricular cardiomyopathy (ARVC). Due to left and right ventricular abnormalities described in this work, ACM is the preferred terminology. Molecular profiling of patients with ACM helps to further characterize differences in disease onset and progression.

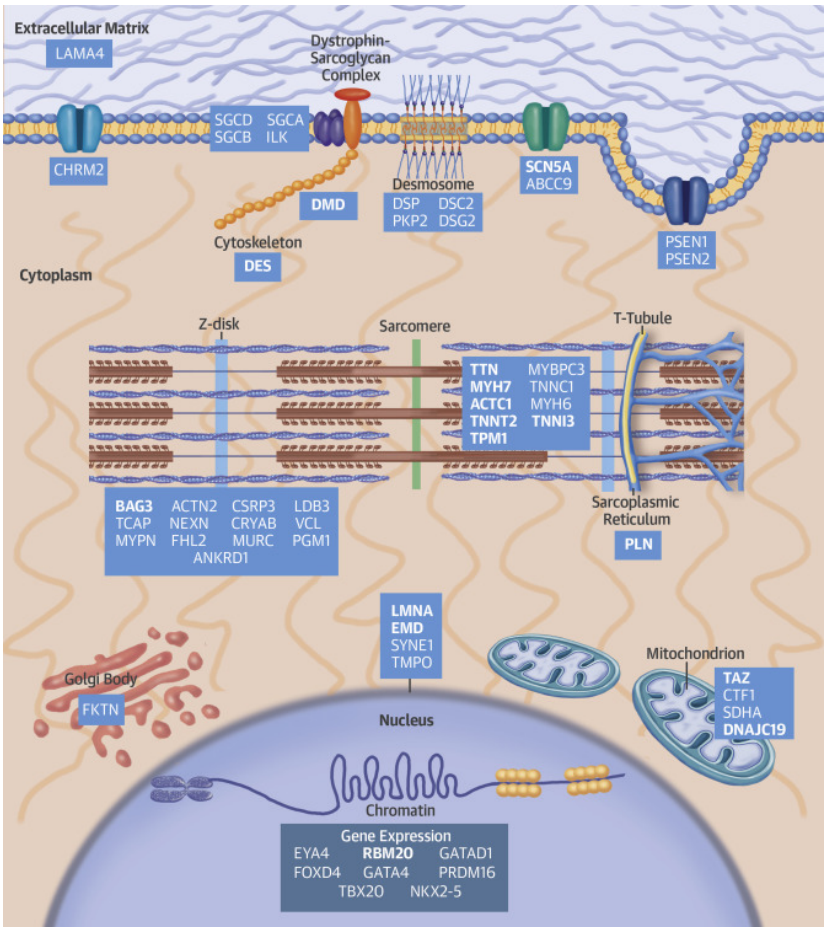

Figure 4: Schematic of genes harboring mutations associated with DCM and ACM. Genes are grouped by the encoding proteins' subcellular location. Definitive DCM-associated genes are highlighted in bold. Figure from Burke et al. (2016).

3.2.3 Current treatment strategies in dilated cardiomyopathy

Current treatment strategies rarely improve heart conditions, but are shown to dampen disease progression. A Japanese long-term study reported significant improvement in therapies over the last 20 years, associated with the increased use of angiotensin-converting enzyme inhibitors (ACEi), angiotensin receptor blockers (ARBs), and B-blockers (Matsumura et al., 2006). Depending on comorbidities such as congestion (peripheral edema), mineralocorticoid antagonists (MRA) or loop diuretics are administered. Statins are often applied to prevent atherosclerotic cardiovascular disease. In case arryhthmias are observed, antiarryhthmic drugs such as amiodarone are prescribed.

Previous studies have shown that cardiomyopathy treatment outcomes differ on the disease origin. One early reported example was amiodarone (Singh et al., 1995), for which higher survival rates are reported for patients with DCM in contrast to ischemic cardiomyopathy.

More recent clinical studies now include the genotype information in the study design, such as the Phase 3 ARRY-371797 study from Pfizer. The study was first submitted in 2018, with an estimated primary completion date in 2024. In this study, specifically LMNA A/C mutants with symptomatic DCM are included. The primary outcome measure is better performance in the 6MWT (6-minute walk test). A recent study with human induced pluripotent stem cells furthermore suggested the increased benefits of statins in DCM-laminopathies (Sayed et al., 2020). Although the outcome and patient benefits of the suggested treatments are uncertain, investigating genotype-stratified studies of heart failure patients is mandatory to improve molecular interventions. Due to the different functions and localizations of affected genes we hypothesized that individual pathogenic variants in mutated genes evoke distinct single-cell molecular phenotypes.

If drug therapies don't improve the patient's conditions or can't prevent disease progression, device therapies or cardiac transplantation are the last options. Heart transplantation is the method of choice, which is however limited by the availability of suitable donor hearts. Alternatively, an LV assist device, (LVAD) is implanted (Ammirati et al., 2014), serving as a bridge to cardiac transplantation. The inflow cannula of the assistance device is placed through the apex, blood from the left ventricle then enters the pump chamber and is then pumped to the aorta into the systemic circulation. The obtained apical core, despite its small size, is large enough for pathology or sensitive assays (Maybaum et al., 2007). Cardiac improvement was reported in some patients with LVAD support, however, after 4-5 years 50% of the LVAD patients need to receive a transplant (Miller et al., 2019).

3.2.4 Cardiac Fibrosis accompanies heart failure

A common feature of cardiomyopathies is the loss of cardiomyocytes due to apoptosis and necrosis. Upon injury, cardiac fibroblasts are activated to initiate wound healing. Fibroblasts, an essential cell type in every organ, and the myocardium, synthesize extracellular matrix and play a pivotal role in structural and mechanical homeostasis (Souders et al., 2009). This process is in interplay with other cell types, such as macrophages. As macrophages and fibroblasts are not highly abundant cell types in the heart, single-cell technologies are valuable to study those populations, as their transcriptional signatures are overlayed in bulk RNA-seq data with that of more abundant cell types.

In normal wound healing, fibroblast activation is downregulated over time. This regulatory

mechanism is dysfunctional in cardiomyopathies, where fibroblasts are frequently activated. The continuous deposition of ECM impacts tissue function, stiffness and oxygenation (Beyer et al., 2009). Cardiac fibrosis is often subdivided into three subtypes, replacing, interstitial and perivascular fibrosis (Thomas and Grisanti, 2020). Replacing fibrosis fills gaps after cardiomyocyte loss with extracellular matrix, while interstitial and perivascular fibrosis refers to increased ECM deposition without cardiomyocyte loss. Due to the high incidence of fibrotic diseases, especially cardiac fibrosis, many scientific studies deal with elucidating disease mechanisms to identify new druggable mechanisms (Zhao et al., 2020).

3.3 Molecular profiling of tissues and single-cells

Next-generation sequencing technologies are the current state-of-the-art technology to measure gene activity based on mRNA abundance with a high dynamic range (Lindberg and Hübner, 2021). Previous bulk RNA studies have defined the underlying molecular mechanism of heart failure (Heinig et al., 2017; van Heesch et al., 2019; Ramirez Flores et al., 2021). However, these studies cannot assign disease-associated transcriptional and proportional changes to specific cell types in the heart. Recently emerging single-cell sequencing technologies are promising to define cell type-specific compositional and transcriptional changes in a high-throughput and unbiased manner.

3.3.1 Profiling of transcriptomic responses on the single-cell level

The workflow starts with fresh tissue, which is enzymatically digested to obtain a single-cell suspension. Single cells are then loaded on a microfluidics chip, where single cells are encapsulated in aqueous droplets with sequencing reagents and barcoded beads capturing mRNA transcripts. Although protocols to profile diverse organisms and organs are available, application in adult cardiac tissue is challenging (Gladka, 2021).

For example, cardiac cell types are very heterogenous in size, which makes sorting- or microfluidics-based single-cell sequencing strategies rather challenging. Cardiomyocytes for example are around 60-140 μm with a length-to-diameter ratio of 5:1 (Tracy and Sander, 2011), accompanied by multinucleation and therefore don't fit through the channels of a standard microfluidics chip (Bensley et al., 2016). In contrast, vascular endothelial cells are around 30μm in width and size, which varies along the vascular tree (Krüger-Genge et al., 2019).

One option is to enrich particular cell types, with the disadvantage of losing information on the responses of other neighboring cell types. In the early phases of cardiac single-cell sequencing, primarily the non-myocyte fraction was analyzed in the adult heart (DeLaughter et al., 2016; Skelly et al., 2018). As cardiomyocytes are the primary cell type expressing sarcomeric and DCM-associated genes, this is the primary affected cell type and hence of significant interest.

Optimization of dissociation protocols is challenging, as cell types require different digestion conditions. For example, large cardiomyocytes are rather sensitive to digestion, while obtaining a single-cell suspension out of a fibrotic plague requires rather harsh dissociation conditions (Skelly et al., 2018; Wang et al., 2020). A previous study furthermore provided evidence for the induction of transcriptomic stress signatures during warm dissociation (van den Brink et al., 2017). Tissue dissociation furthermore requires the availability of fresh tissue, where the logistical set-up from the hospital and to the lab is complex and requires staff being available round the clock. The selection of a protocol that applies to frozen or fixed tissue is preferred, which also enables studies on pre-established biobanks to have access to large cohorts.

An alternative to single-cell sequencing technologies is single-nucleus sequencing (snRNAseq), which is applicable to frozen tissue and overcomes the size restriction to specific cell-types (Krishnaswami et al., 2016; Lake et al., 2016).

3.3.2 Single-cell sequencing technologies

In recent years, single-cell sequencing platforms based on different technologies have entered the market. A broad separation can be done between microfluidics- and well-based platforms (Kolodziejczyk et al., 2015; Svensson et al., 2018). The microfluidics-based platform by 10x Genomics is currently one of the most commonly used single-cell platforms due to its reproducible chemistry, and high capturing and recovery rate relative to loaded nuclei (Figure 5). The platform is furthermore compatible with single nucleus RNA-sequencing.

For the first steps, Gel Beads, a master mix containing cells, template switch oligos, reagents and enzymes for reverse transcription, and Partitioning Oil are required (Figure 6). The 10x system works with the formation of Gel Bead-in-Emulsion (GEMs), in which one barcoded bead is encapsulated into a nanoliter-sized droplet. In total, more than 100,000 GEMs are formed by the Chip while being in the 10x Chromium Controller. Depending on the loading, only 1-10% of GEMs normally contain a cell. Each bead-immobilized oligo contains

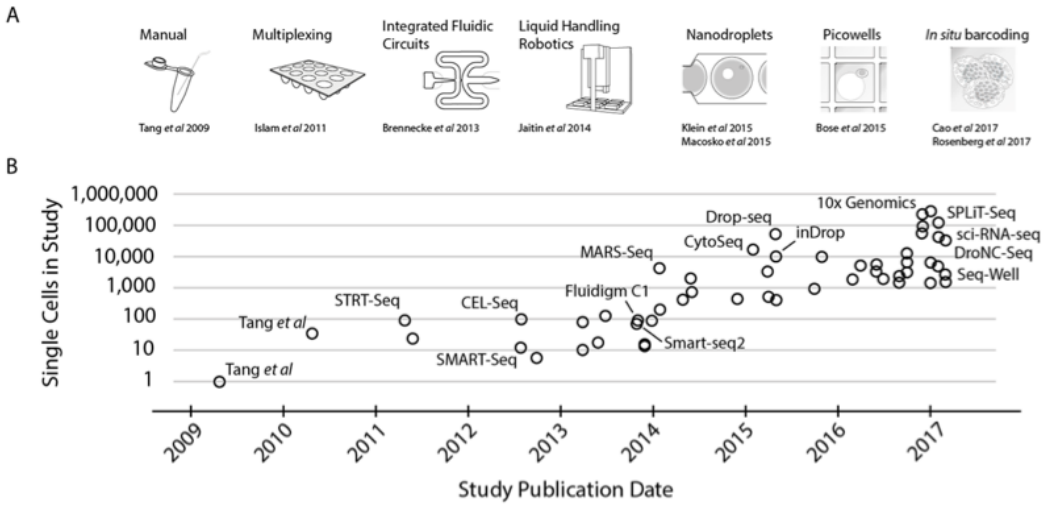

Figure 5: Scale of sequenced cells per single-cell sequencing platform. A) Key single-cell sequencing technologies ordered by year of publication and amount of cells which can be sequenced. B) Cell numbers reported in publications using different platforms. Figure from Svensson et al. (2018).

the bead-barcode, an unique molecular identified (UMI) to detect PCR duplicates, and a Poly-dT tail to capture the 3' end of polyadenylated transcripts. After reverse transcription and cDNA pooling, the libraries are sequenced on an Illumina sequencer, which is the currently most often used NGS platform (Lindberg and Hübner, 2021).

3.3.3 Processing of single-cell sequencing data

After sequencing, read mapping to the genome, and PCR duplicate removal, a count matrix (cell barcodes by genes) is generated and stored in the h5 or sparse matrix format. Multiple single-cell analysis pipelines are well established in the field with similar working principles, of which scanpy was chosen for downstream analysis, a python implementation of Seurat (Satija et al., 2015; Butler et al., 2018; Wolf et al., 2018). In short, the overall goal of this pipeline is to identify nuclei with high similarity in their gene expression profile and manually assign their cell type based on statistically significantly enriched or depleted marker gene signatures (clustering). This is done by embedding high-dimensional data, here nuclei, defined by their gene expression profile, into a low-dimensional space in multiple steps, such as the selection of highly variable genes, and principal component analysis followed by a nonlinear dimension reduction method. Currently, uniform manifold approximation and projection (uMAP) is the most often used low-dimensional projection method, as this algorithm preserves local and global distances of nuclei in the original transcript space (McInnes et al., 2018). At the

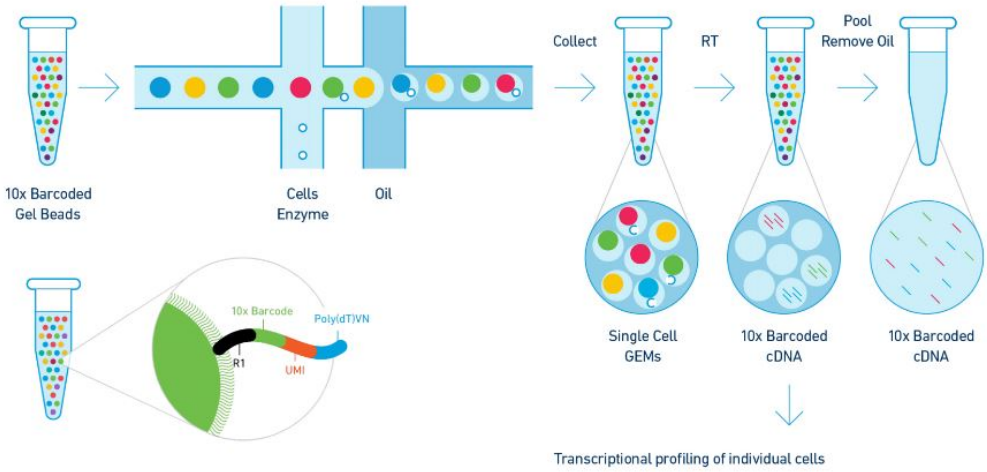

Figure 6: Schematic representation of the 10x Genomics workflow. Barcoded beads are required for Gel Bead-in-Emulsion (GEM) formation. Each GEM contains one cell, of which a cDNA library is generated. Figure from 10x Genomics (2019a).

beginning of this project, no standardized annotation strategy or comprehensive marker list for all cell types of the adult human heart was established. The first studies on cardiac tissue were published in 2016 and 2018 by multiple labs (DeLaughter et al., 2016; Pinto et al., 2016a; Skelly et al., 2018; Gladka et al., 2018). Due to technical challenges, only a small number of samples were reported, limiting the focus to highly abundant cell types. Furthermore, many studies were only done with murine tissue. Before the characterization of heart failure-specific signatures, a reference list with marker genes per cell types was needed.

Large single-cell datasets not only allow the dissection of different cardiac cell types but also have the resolution to robustly identify different states of each cell type. Nuclei in the dataset which are annotated as the same cell type are therefore subsetted and are reclustered. This allows the separation of smaller transcriptional differences within one cell type, which are not picked up during the pre-dimensional reduction, such as the selection of highly variable genes, defined by their mean expression and dispersion, or running a principal component analysis.

4 Methods

4.1 Data reporting

Power calculation was performed on murine data to estimate a suitable sample size for the disease cohorts. A sample size of 8 allows the identification of a high fraction of differentially expressed genes with high confidence without exceeding experimental budget and exhausting experimental resources (Figure S1). Calculation was performed with power 80% ($\beta = 0.2$) and $\alpha = 0.05$. Power calculation of compositional differences remained challenging in the pilot project due to low effect sizes due to large standard deviations of cell-type abundance between individuals. Processing of samples obtained from the Bad Oeynhausen Heart and Diabetes Center was blinded to avoid batch effects.

4.2 Ethics statement

Experiments on murine tissue in the initital trial phase were conducted under killing license X 9006/13.

For the healthy heart study, 14 unused transplant organs were collected at the NIHR Cambridge Biomedical Centre Hospital, UK, and Mazankowski Alberta Heart Institute, Canada (Litviňuková et al., 2020). Donor information was anonymized and used after Research Ethics Committee Approval:

a) East of England Cambridge South Research Ethics Committee (ref 15/EE/0152)

b) Mazankowski Alberta Heart Institute (MAHI, Edmonton, Canada); Human Organ Procurement and Exchange Program (HOPE, Pro00011739)

For the comparative analysis with DCM tissue, 12 previously published non-failing control hearts were included, due to library quality (D11) and elevated stress signature identified in D3). Six additional donors from the Bad Oeynhausen Heart Center were included (Reichart et al., 2022), which were processed together with heart failure samples. Discarded disease heart samples were obtained in the context of clinical patient care after heart explantation. All cardiac tissues were anonymized by their study centers and used with approved protocols reviewed by the ethics committees listed below:

a) Bad Oeynhausen Heart Center; Ethics Board of the Ruhr-University Bochum (Approvals 2020-640-1; 21/2013)

b) Mazankowski Alberta Heart Institute (MAHI, Edmonton, Canada); Human Explanted

Heart Program (HELP, Pro00011739)

c) Cardiovascular Research Centre Biobank at the Royal Brompton and Harefield Hospitals, Guys and St. Thomas NHS Foundation Trust (EC reference 09/H0504/104 +5)

d) Imperial College (REC reference16/LO/1568)

e) Mass General Brigham Human Research Protection Committee (Protocol 1999P010895)

f) Harvard Longwood Campus Institutional Review Board (Protocol M11135)

4.3 Cohort samples and patient inclusion criteria

For the Heart Cell Atlas Study, transmural tissue pieces of six different cardiac regions (left and right atria and ventricles, left ventricular apex and intraventricular septum) have been sampled. For the Heart Failure Atlas Study, biobanked full-thickness myocardial specimens obtained either during ventricule assist device (VAD) implantation (n=15, only apical core sample available) or after heart explantation (n=31, left and right ventricular free wall and septum) were collected. Regions of high epicardial fat deposits or macroscopic areas of high fibrosis were excluded. Further information on tissue availability per patient is provided in table S1 of the Heart Failure Atlas publication (Reichart et al., 2022).

4.4 Patient genotyping

Disease patients have been genotyped using assays applied before genetic counseling of patients in the hospital. For the cohort collected at the HDZ NRW and processed at MDC, DNA from blood samples from patients was isolated using the High Pure PCR Preparation Kit (Roche Diagnostics GmbH) or Genomic DNA Extraction Kit (Qiagen). Sequencing libaries were generated using the TruSightTM Rapid Capture Sample Preparation Kit (Illumina), and sequencing was performed on a MiSeq. Libraries were enriched for protein-coding exonic reads using the TruSightTM Cardio (174 genes) or the TruSightTM Cardiomyopathy (46 genes) Sequencing Panel (Illumina) and variants were called using the VariantStudioTM Software (Illumina). Classification and interpretation of variants were done according to the ACMG criteria, if no information was available via Clinvar (Landrum et al., 2018). Information on all genotyping assays per patient is provided in table S1 of the publication (Reichart et al., 2022). Patient genotype information was collected during the course of patient treatment at the Heart and Diabetes Center (HDZ) in the lab of Prof. Dr. Hendrik Milting.

4.5 Isolation of single nuclei for single-nucleus RNA-seq

For this protocol 20-50 mg tissue pieces were used, depending on the sample source. Apical cores obtained from LVAD implantation were smaller than those tissue pieces obtained after heart explantation.

The following buffers are prepared in advance:

Reagent	Final Concentration [mM]	Volume [μl] (for 3 samples)
1.5M Sucrose	250	2500
2M KCl	25	187.5
1M Magnesium chloride	5	75
1M Tris HCl (pH 8)	10	150
NFW	-	12087.5
Total	-	15000

Table 1: NIM1 buffer recipee

Reagent	Final Concentration [mM]	Volume [μl] (for 3 samples)
NIM1	-	4895
1mM DTT	1 μM	5
50x Protease Inhibitor	1x	100
Total	-	5000

Table 2: NIM2 buffer recipee

Reagent	Final Concentration [mM]	Volume [μl] (for 3 samples)
NIM2	-	4850
40U/μl RNaseIn Plus	0.4 U/μl	50
20U/μl SuperaseIn	0.2 U/μl	50
10% (v/v) Triton X-100	0.1%	50
Total	-	5000

Table 3: HB buffer recipee

Reagent	Final Concentration	Volume (for 3 samples)
PBS (–)	-	1492.5 μl
BSA	4%	60 mg
40 U/μl Protector RNAseIn	0.2 U/μl	7.5 μl
Total	-	1500

Table 4: SB buffer recipee

Dounce homogenizer and liquids need to be consistently on ice. The tissue piece was then transferred into a precooled 7ml Dounce homogenizer filled with 3 ml of HB buffer and homogenized with 7-12 strokes of the loose, and 7-12 strokes of the tight pestle. Tissue homogenate was then filtered through a 40μl cell strainer into a 50ml Falcon tube, followed by 2x 1ml washing of the homogenizer with HB buffer and centrifuged for 5min, 4 C at 500g to pellet the nuclei. The supernatant was removed without rupturing the unstable pellet. The pellet was then resuspended in 500μl SB buffer. From this step, three methods of purification where compared.

Reagent	Volume [ml]
Percoll	4.5
PBS (10x)	0.5
Total	5000

Table 5: Percoll solution

Additional 9.5ml SB was added to the resuspended pellet, and resuspended with Percoll solution to obtain a final Percoll concentration of 20%, here 2.86 ml. The solution was centrifuged for 15,000g for 20min at 4C. Nuclei were collected from the cell-containing phase using a wide-opening 1ml pipette tip. Nuclei were washed 3x 500 μl SB buffer.

For the filtration protocol, additional 500μl SB buffer were added to the resuspension. 3x 50ml Falcon tubes with a 70μm, 40μm and 10μm strainer were prepared, the filters were wetted with SB to avoid excessive loss of nuclei. The resuspension was filtered through all 3 filters. After filtration, the collection tube was centrifuged for 5min, 4C at 500g to pellet the nuclei and resuspended in 500 μl SB buffer.

Nuclei undergoing purification were transferred to a FACS sorting tube. 50μl per suspension was set aside as a negative unstained control. The other 450μl were stained with 1 drop of

NucBlue for sorting and stained for approximately 10min. A third Eppendorf tube was filled with 100μl SB buffer and serves as a collection tube for nuclei. Nuclei are then FACS sorted into the pre-chilled collection tube until 150,000 events are sorted, which took around 20min per sample. After sorting, the collection tube is centrifuged for 5min, 4C at 500g to pellet the nuclei, supernatant was removed and resuspended in SB in a ratio of 1.000 FACS events per 1μl, with 70μl minimum resuspension volume.
Nuclei quality was qualitatively assessed under a brightfield microscope. The number of nuclei in suspension was automatically counted using a Countess II. Nuclei suspension was diluted to target recovery of 5,000 nuclei per run as described in the 10x V3 3' RNA sequencing protocol (10x Genomics, 2020). Recovered nuclei per run are shown in Table S2 of the publication (Reichart et al., 2022).
The protocols were published for open access (Litviňuková et al., 2020; Nadelmann et al., 2021).

4.6 Library preparation using the 10x 3' chemistry

Single-cell libraries for murine tissue were prepared using the V2 in the pilot phase according to the manufacturer's protocol (10x Genomics, 2019a). Chromium Single Cell Reagent Kits V3 has been used for all disease and control samples generated in our lab (10x Genomics, 2020). Chromium Single Cell Reagent Kits V3.1 were used for three samples to confirm compatibility with data analysis workflow for future projects (10x Genomics, 2019b). Fragment size and libraries were quantified using the Bioanalyzer High Sensitivity DNA Analysis (Agilent) and the KAPA Library Quantification kit. Analysis was done according to the manufacturer's protocol (Technologies, 2013; KAPABiosystems, 2017). Single-cell libraries were sequenced using an Illumina HighSeq 4000 with a targeted read number of 30,000-50,000 reads per nucleus.

4.7 Preprocessing of sequencing data, mapping, and generation of count matrix

Bcl files from the Illumina sequencer were converted to fastq files using bcl2fastq. Each sample was mapped to the human reference genome GRCh38 with a modified pre-mRNA gtf file of Ens84 using the *CellRanger* package. Reads mapping into exonic and intronic regions were counted, but were discarded when overlapping multiple sequence features. The *CellRanger* suite was used with default parameters and expected cell count of 5,000. Each

library's mapping quality was assessed and resequenced if a high number of low-confidence base callings were identified. Empty droplets as identified by *Emptydrops*, integrated into the *CellRanger* pipeline, were removed from the raw_feature_bc_matrix.h5 and stored as filtered_feature_bc_matrix.h5. Two external datasets exceeded the read number per nucleus of 200,000 reads and were downsampled to 100,000 reads per nucleus on the molecule_info.h5 file to remove technical noise and overamplification of empty droplets. For this, the *downsampleReads* function available in the *DropletUtils* package in R was used. After importing all Filtered_feature_bc_matrix.h5 files, adata objects were generated, aggregated, and metadata was merged to the .obs data frame.

4.8 Doublet prediction using scrublet and solo

The aggregated anndata object was split into individual adata files by sample and doublet scores with the *scrublet* package were computed on the unprocessed UMI counts stored in adata.X (Wolock et al., 2019). Scores have been computed on a log- (scrublet_score_log) and z-transformed (termed scrublet_score_z) UMI counts.

For *solo* score computation, sample-specific h5ad files were used as input (Bernstein et al., 2020). The model_json_file for the learning contained the following parameters *n_hidden*: 384, *n_latent*: 64, *n_layers*: 1, *cl_hidden*: 128, *cl_layers*: 1, *dropout_rate*: 0.2, *learning_rate*: 0.001, *valid_pct*: 0.10. The softmax score was stored as output and stored in the column solo_score.

4.9 Quality control filtering, batch correction and low dimensional manifold embedding

For the murine pilot project, the R *Seurat* environment was used (Satija et al., 2015; Butler et al., 2018). For the analysis of healthy and diseased human hearts, pythons *scanpy* was used due to improved CPU usage (Wolf et al., 2018; Zappia and Theis, 2021).

For the aggregated anndata object, the number of detected UMIs (n_counts), genes (n_genes), and fraction of UMIs mapping to the ribosomal (percent_ribo) and mitochondrial genes (percent_mito) per nucleus were computed. Based on the distribution of quality parameters, filtering criteria were applied: $300 \leq n\_counts \leq 15000$, $300 \leq n\_genes \leq 15000$, $percent\_mito \geq 1\%$, $percent\_ribo \geq 1\%$, $solo\_score \geq 0.5$.

At first, UMI counts were normalized to a total size of 10,000 using *sc.pp.normalize*, 1 was added and values were logarithmized using the *sc.pp.log1p* function. The normalized counts are then stored in adata.raw.X. To reduce lowly expressed or equally distributed genes for

clustering (such as housekeeping genes), genes were filtered based on dispersion and mean expression, resulting in a set of highly-variable genes (HVG). For pre-dimensional reduction and initial denoising principal component analysis was applied. For the Heart Cell Atlas, cell and nuclei-separated adata objects were first independently batch-corrected with BBKNN with "patient" as batch prior to umap embedding (McInnes et al., 2018; Polański et al., 2020). Due to better performance in benchmarking experiments and cross-platform compatibility, sample integration for failing hearts was done using harmonypy, followed by construction of neighboring graph and umap embedding (Korsunsky et al., 2019; Tran et al., 2020). Leiden clustering via *sc.tl.leiden* was used for graph-based clustering to detect cellular communities (termed cell-type or -states) based on optimizing modularity (Traag et al., 2019). If the spearman correlation of distinct cellular communities was very high and no high number of distinguishing marker genes was identified, highly-correlating cellular communities were merged to avoid over-clustering. SCCAF was used to additionally evaluate clustering robustness (Miao et al., 2020).
For the Heart Cell Atlas, samples from different modalities (cells or nuclei) were subsequently integrated in a supervised manner using scGen (Lotfollahi et al., 2019).

4.10 Differential gene expression analysis

Marker genes for cell types and states have been computed using a Wilcoxon rank sum test, as implemented in scanpy. p-values were corrected for multiple testing by false discovery rate (FDR) controlling using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Fold changes are reported on log2 scale. Only genes with FDR ≤ 0.05 are reported.
To compute pairwise gene expression changes between control and genotype groups in the DCM cohort, edgeR was redesigned to be compatible with single-nucleus RNA-seq data (Robinson et al., 2010). Only genes with a minimum mean expression of 0.0125 in the control and disease groups were tested. Raw UMI counts were summed up per cell type or state per patient per region and considered as one sample. Only genes with absolute $\log_2$FC of 0.5 and FDR ≤ 0.05 were considered as significantly changed. To receive robust results from edgeR, only cell types or states with ≥ 5 nuclei were detected within ≥ 3 patients. Using this approach allows for the inclusion of more covariates in the model and gives patients equal weight in downstream analysis independent of subgroup size or recovered nuclei number. Intersections of differentially expressed genes were visualized using the UpSetR package (Conway et al., 2017).

4.11 Gene set score enrichment analysis

Enrichment of pathways and custom-defined gene sets was calculated using *sc.pp.score_genes* on library-size normalized, log-transformed and scaled UMI counts (Wolf et al., 2018). The following gene sets were used: OSM pathway (Dey et al., 2013; Abe et al., 2019; Litviňuková et al., 2020), TGFB-stimulation (curated from Schafer et al. (2017) using $ldFC > 0.7$ and FDR of 0.05), Antigen presentation (MHCII) score (Shiina et al., 2009), and all expressed collagen genes. Cell-cycle scoring was done using the scanpy function *sc.pp.score_genes_cell_cycle* according the manual.

4.12 Differential abundance analysis

Compositional data analysis was performed to determine genotype-specific differences in cell type abundances, excluding unassigned nuclei. To account for the compositional nature of the data, raw cell type counts were transformed using the centered log ratio (CLR) transformation per cell-type or -state C (Van den Boogaart and Tolosana-Delgado, 2008), as shown in equation 1:

$$\mathrm{clr(c)} = \left[\ln \frac{c_i}{g(\mathrm{c})}; ...; \ln \frac{c_D}{g(\mathrm{c})}\right] \tag{1}$$

where $g(\mathrm{c}) = \sqrt[D]{c_i ... c_D}$ was the geometric mean of the proportions vector. c was the proportions vector and D the total number of cell types or cell states. To assess statistical differences between the control vs. genotype group, a linear model and t-test were performed to determine the significance of the regression coefficient. The abundance of cell types or states was assessed between genotypes and controls and compared among each other. P-values were adjusted for multiple testing using the Benjamini and Hochberg method for all performed tests. Differential abundance of cell types and states was computed separately by region. Only samples with $\geq$ 10 nuclei per sample were included in the cell type and state-specific analysis.

In addition to CLR values, percentages of abundance differences were reported for improved interpretability. The proportional changes in mean percentages of control and disease samples were reported. Reported positive values indicate higher abundance in the heart failure group. In addition to CLR values, log ratios of abundances for cell type (or state) pairs between genotypes and controls were ascertained and reported.

4.13 Collagen quantification via hydroxyproline measurement

Extracellular matrix content was measured using hydroxyproline (4OH-P) as a surrogate using a previously published protocol (Stegemann and Stalder, 1967; Kassner et al., 2021). In short, 10 mg transmural tissue pieces were incubated at 6M HCl at 110C for 16h. Next, 4OH-P in the hydrolysate was converted with Chloramine T to pyrrole. Pyrrole was converted to a chromophore using p-di-methylaminobenzaldehyde at 60C, which was measured by a photometer at 550nm. This assay was performed at the Heart and Diabetes Center (HDZ) in Bad Oeynhausen in collaboration with the lab of Prof. Dr. Hendrik Milting.

4.14 Validation of differential gene expression and cell-states using single-molecule fluorescent in-situ hybridization with RNAscope probes

Fresh-frozen tissue pieces were fixed overnight in 4% paraformaldehyde solution and then placed in 30% sucrose in PBS until submerged. Fixed tissue pieces were embedded in OCT using a cryo mold and sliced to 5 μm thickness on a cryotome. Tissue slices were mounted on a Superfrost Plus slide and incubated for at least 1h at -20C in horizontal position. Slides were stained with probes according to the manufacturer's protocol for fixed-frozen tissue (RNAScope Multiplex Fluorescent Reagent Kit V2, ACDBiotechne). Protease IV showed the best results for digestion by measuring signal intensity. For each tissue positive and negative control staining was performed using the accompanying control probes. Probe multiplexing was done according to the manufacturer's protocol. Sections were counterstained with DAPI (Wavelength: 358/461 nm) and WGA (Biotium CF633 WGA Wavelength: 630/650nm). Imaging was done using an LSM710 confocal microscope (Zeiss, for samples collected in North America) with 20x or 40x oil immersion objectives (1.3 oil, DIC III) or an SP8 confocal microscope (Leica, for samples obtained and processed in Germany) with a NA 1.4 63x oil immersion objective. Following fluorophores were used and conjugated to the probes: Opal 520 (Wavelength: 494/525 nm), Opal 570 (Wavelength 550/570 nm), Opal 620 (Wavelength 588/616 nm), Opal 690 (Wavelength 676/694 nm) dyes (Akoya Bioscience). Spectral bleed-through was corrected via subtraction using Fiji. WGA was diluted in PBS instead of HBSS as recommended, as HBSS diminished the fluorescence signal of opal dyes or probe binding to its RNA target.

4.15 Computing differential cell-cell signaling using Cellchat

Cellchat was used to compute cell-cell interactions between cell states in LV and RV separately using the database provided at: http://www.cellchat.org/cellchatdb/ (Jin et al., 2021). If cell types are shown, communication probabilities across all cell states per cell type are aggregated using the mergeInteractions() functionality. Differences in signalling were inferred by comparing control samples with genotype subgroups separately. Cellchat was applied on library-size normalized and log-transformed counts, and accounting for population size as suggested by the developers. The output from the rankNet() function was used to generate interaction heatmaps. P values were adjusted for multiple testing using false discovery rate (FDR) controlling using the Benjamini-Hochberg method. Fold changes are reported on the log2-scale.

4.16 Models of Genotype classification using traditional machine learning models

4.16.1 Using traditional machine learning models for genotype subgroup classification

Machine learning models were trained on the raw counts with library size normalization, and logarithmized with adding prior one pseudocount. The training was done on LV and RV data separately, as joint training lead to reduced model performance. Model performance was furthermore decreased when applying the model across all cell types, hence the anndata object was split by cell type before training. The training was furthermore done using the k-fold cross-validation. To avoid overfitting to patient-specific transcriptional signatures, training was conducted by removing all nuclei from one patient, whose nuclei are then assigned the most-likely genotype class (Leave one out cross-validation policy). At first, multiple logistic regression, LASSO, and linear kernel SVM were applied due to the high level of interpretability, which however showed low performance.

4.16.2 Using graph attention networks for genotype subgroup classification

The first graph attention model was published in 2020 Ravindra et al. (2020), whose model architecture was here further developed. The described GAT architecture is based on the PyTorch framework. This developed GAT model furthermore implements knowledge derived from the modelling with traditional machine learning models, such as the cell type and region separated models and leave-one-out-cross-validation training policy. Details on model

architecture are published in Reichart et al. (2022). The model was developed in collaboration with Nikolay Shvetsov and Prof. Dr. Christoph Lippert at the Hasso Plattner Institute.

5 Materials and Software

5.1 Reagents and Equipment

Reagent/Equipment	Producer	Catalogue number
BSA	Sigma-Aldrich	A3059-50G
Chromium Next GEM Chip G Single Cell Kit, 48 rxns PN-1000120	10x Genomics	PN-1000120
Chromium Next GEM Single Cell 3 GEM, Library & Gel Bead Kit v2.0, 16 rxns	10x Genomics	PN-120237
Chromium Next GEM Single Cell 3 GEM, Library & Gel Bead Kit v3.0, 16 rxns	10x Genomics	PN-1000075
Chromium Next GEM Single Cell 3 GEM, Library & Gel Bead Kit v3.1, 16 rxns	10x Genomics	PN-1000121
DNA LoBind Tubes, 1.5 ml	Eppendorf	022431021
DNA LoBind Tubes, 2.0 ml	Eppendorf	022431048
DTT	LIFE Technologies	P2325
DynabeadsTM MyOneTM SILANE	Thermo Fisher Scientific	PN-2000048
EB Buffer	Qiagen	19086
Ethanol	Millipore Sigma	E7023-500ML
Glycerin (glycerol), 50% (v/v) Aqueous Solution	Ricca Chemical	3290-32
High Sensitivity DNA Kit	Agilent	5067-4626
High Sensitivity D5000 ScreenTape/Reagents	Agilent	5067-5584
ImmEdge Hydrophobic Barrier Pen	ACD Biotechne	310018

Reagent	Producer	Catalogue number
KAPA Library Quantification Kit for Illumina Platforms	KAPA Biosystems	KK4824
Low TE Buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA)	Thermo Fisher Scientific	12090-015
Magnesium chloride (MgCl, 1M)	LIFE Technologies	AM9530G
NucBlue	LIFE Technologies	R37605
Nuclease-free Water	Thermo Fisher Scientific	AM9937
OCT-Compound/Tissue-Tek	Sakura Finetek Germany	sa-4583
Opal 520 Reagent	Akoya Biosciences Inc.	FP1487001KT
Opal 570 Reagent	Akoya Biosciences Inc.	FP1488001KT
Opal 620 Reagent	Akoya Biosciences Inc.	FP1495001KT
Opal 690 Reagent	Akoya Biosciences Inc.	FP1497001KT
PCR Tubes 0.2 ml 8-tube strips	Eppendorf	951010022
Percoll	VWR	17-0891-02
PI pellets	Sigma-Aldrich	11873580001
pluriStrainer 40μm	pluriSelect	43-50040-50
pluriStrainer 20μm	pluriSelect	43-50020-50
pluriStrainer 5μm	pluriSelect	43-50005-50
Polypropylene Conical Tube (50ml)	VWR International	352070
Protector RNAseIn 40U/ul	Sigma-Aldrich	03335402001
ProLong Gold Antifade Mountant-10ml	LIFE Technologies	P36930
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
RNAscope 4-Plex Ancillary Kit for Multiplex Fluorescent Kit v2	ACD Biotechne	23120

Reagent	Producer	Catalogue number
RNAscope 4-plex Negative Control Probe	ACD Biotechne	321831
RNAscope 4-plex positive Control Probe - Hs	ACD Biotechne	321801
RNAscope Multiplex Fluorescent Reagent Kit v2	ACD Biotechne	323100
RNAscope Probe Diluent	ACD Biotechne	300041
RNAscope Wash Buffer Reagents	ACD Biotechne	310091
RNAseIn Plus 40 U/ul	Promega	N2611 (2500U) and N2615 (10000 U)
Round bottom Polystyrene Test Tube (5ml)	Neolab	352235
Safe-Lock Tubes PCR clean, DNA low bind	Sarstedt	SARS72.706.700
Safe-Lock Tubes Ambra, 1.5ml	VWR International	0030120.191
Single Index Kit T Set A, 96 rxns	10x Genomics	PN-1000213
Sodiumchloride (KCl, 2M)	LIFE Technologies	AM9640G
SPRIselect Reagent Kit	Beckman Coulter	B23318
Sucrose	Sigma-Aldrich	S0389
SuperaseIn 20 U/ul	LIFE Technologies	AM2696
Superfrost Plus microscope slides	Thermo Fisher Scientific Inc., Waltham, MA	J1800AMNZ
TempAssure PCR 8-tube strip	USA Scientific	1402-4700
Tween 20, 10%	Bio-Rad	1662404
Tris buffer, pH 8 (1M)	LIFE Technologies	AM9855G
Triton X-100 10% (v/v)	Sigma-Aldrich	T8787

Reagent	Producer	Catalogue number
UltraPure SSC, 20x-1 L	LIFE Technologies	15557044
Wheat Germ Agglutinin, CF 405M Conjugate	LINARIS	29028
Wheat Germ Agglutinin, CF 568 Conjugate	Biozol Diagnostica	BOT-29077-1
Wheat Germ Agglutinin CF 633	BioTrend Chemikalien	b-29024-1
X50 BLADE MB35 PREMIER LP	Fisher Scientific	11992345

Table 6: Reagents and Equipment

5.2 RNAscope probes

Target gene	Catalogue number
C1QA	485451-C2
CCL2	423811-C4
DAAM1	1047161-C2
DCN	589521
IL11	425281-C2
POSTN	409181-C3
TOP2A	470321-C3

Table 7: RNAscope probes

5.3 Instruments and Pipettes

Machine	Manufacturer
HyBEZ Hybridisierungssystem with Batch Slide System	ACD Biotechne
Avanti J 26XP	Beckmann Coulter
CM3050S	Leica
FACSAria I	BD
FACSAria II	BD
FACSAria III	BD
FACSAria Fusion	BD
Centrifuge 5417R	Eppendorf
Centrifuge 5810 R	Eppendorf
Chromium Controller	10x Genomics
Chromium Next GEM Secondary Holder	10x Genomics
Magnetic Separator	10x Genomics
Pipette LTS L1000	Rainin
Pipette LTS L200	Rainin
Pipette LTS L20	Rainin
Pipette LTS L2	Rainin
Qubit Flourometer (4.0)	Thermo Fisher Scientific
TapeStation	Agilent
Vortex Adapter	10x Genomics
Vortex Mixer	VWR
Waterbath 1003	GFL

Table 8: Instruments and Pipettes

5.4 Software

Package name	Version
Agilent Bioanalyzer 2100 Laptop Bundle	
bcl2fastq	2.19.0
CellRanger pipeline	3.0.2
Fiji Image software	2.1.0, v1.53c
Jupyter Notebook	6.0.3
Leica Application Suite X software	3.5.2

Table 9: Software

Package name	Version
anndata	0.7.3
bbknn	1.3.6
Cellchat	1.1.0
harmonypy	0.1
Keras	2.3.1
leidenalg	0.8.0
pandas	1.0.3
scanpy	1.5.1
SCCAF	0.0.10
scGen	1.1.4
scikit-learn	0.22.2
scipy	1.4.1
scrublet	0.2.1
seaborn	0.10.1
solo	0.1
tensorflow	1.13.1
umap-learn	0.4.2

Table 10: List of main Python packages and version number. Python 3.9 was used for this project (Van Rossum and Drake, 2009). GPU-based tools were run on a Tesla-T4 with CUDA 11.0.

Package name	Version
BiocManager	1.30.10
BioBase	2.46.0
BioGenerics	0.32.0
CellChat	1.1.1
DropletUtils	1.6.1
dplyr	1.0.6
edgeR	3.28.1
ggplot2	3.3.3
igraph	1.2.6
IRkernel	1.1.1.9000
limma	3.42.2
magrittr	1.5
plyr	1.8.6
Seurat	3.2.2
SeuratData	0.2.1
SeuratDisk	0.0.0.9013
UpSetR	1.4.0

Table 11: List of main R packages and version number. R 3.6.3 was used for this project (R Core Team, 2021).

Python and R environments are available as .yml files. Anaconda (4.8.3) was used as package managers during this PhD project (ana, 2020) on the High Performance Cluster of the Max-Delbrück Center for Molecular Medicine, Berlin. Scripts and functions established during are published on github ($https://github.com/heiniglab/DCM_heart_cell_atlas$).

5.5 Data availability

Processed and annotated data were deposited on CZIs cellxgene (https://cellxgene.cziscience .com/ collections/e75342a8-0f3b-4ec5-8ee1-245a23e0f7cb) and raw count matrices were uploaded on Zenodo (https://zenodo.org/record/6962685# .YuyuihxBzKE). Raw fastq files were deposited on EGA under accession number EGAS00001006374. The supplementary tables of this thesis are provided in a zenodo repository under doi 10.5281/zenodo.7312018.

6 Results

6.1 Protocol optimization for isolation of intact nuclei in murine and human tissue

6.1.1 Purification and integrity of isolated nuclei

Due to the large rod-shape of Cardiomyocytes, isolation and sequencing of nuclei is the preferred method of single-cell sequencing in the heart. A hypotonic homogenization buffer (here HB) is used together with manually induced shear stress to rapture the outer cellular membrane, leading to the separation of cytoplasmic content and intact nuclei. Due to the excessive amount of cellular debris as observed under the microscope, purification is suggested prior to loading on the 10x Chromium Controller. Different purification strategies for tissue homogenates have been suggested, such as gradual filtration, density gradient centrifugation and FACS sorting (for example Lake et al. (2019); Grindberg et al. (2013); Krishnaswami et al. (2016); Lake et al. (2016)). One method is based on density gradient centrifugation using a Percoll gradient. Purification by filtration uses multiple iterations of filteration with reducing filter pore size to step-wise remove debris. For FACS sorting, previously NucBlue-stained nuclei are sorted into a separate tube. Details on gating strategy are shown below. To test effects of the purification method on nuclei quality and the amount of spurious signals (ambient RNA), perfused murine left-ventriclar tissue were processed in the Dounce homogenizer followed by three different purification protocols, loading on the 10x Chromium controller and sequencing (Figure 7)). To evaluate nuclei quality, detected transcripts of all genes per droplet are summed-up and plotted in numeric order, x- and y-axis are logarithmised ("Barcode-rank plot"). Transcripts are quantified with a 10 nucleotide long unique molecular identified (UMI), which is a primer bound to the 10x Gel bead. For the barcode-rank plot, a double plateau-shaped curve is expected. The first plateau represents nuclei containing droplets, the second showing the levels of ambient RNA. Only one sample was tested per purification method.

The FACS purified nuclei showed the highest average number of UMIs compared to the other methods, despite increased number of ambient RNA-derived UMIs detected in empty droplets. As the average detected UMIs per nuclei are 5-fold higher in the FACS-sorted group, with clearer separation between the both plateaus, FACS purification shows higher quality nuclei yield by gating-out small sized particles.

FACS sorting furthermore allows the counting of isolated nuclei and recovery of low numbers

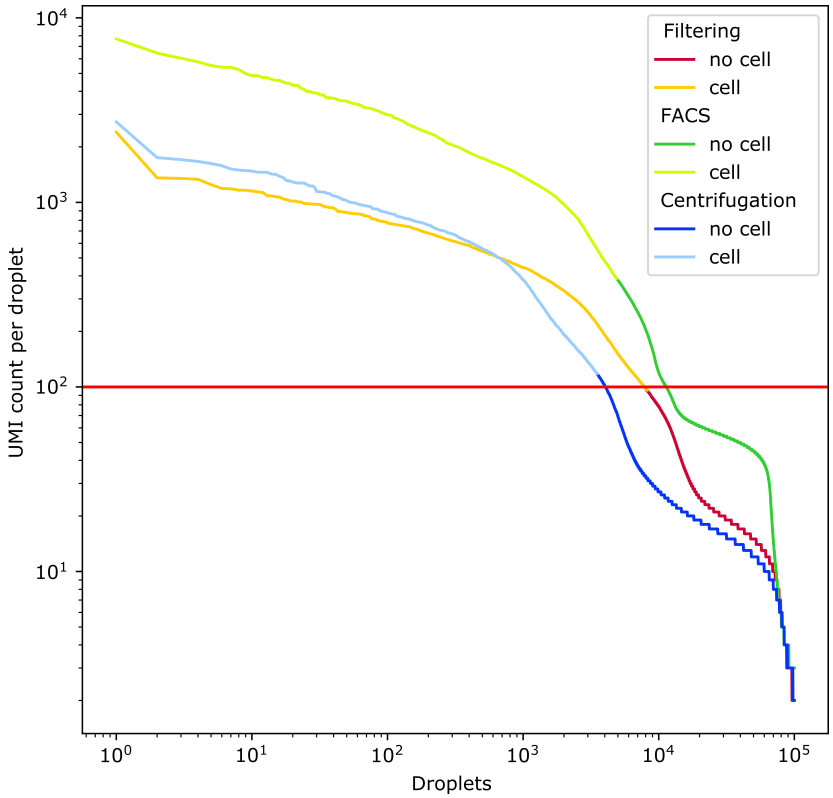

Figure 7: Barcode-rank plots for three samples purified using filtration (red), FACS (green) and density gradient centrifugation (blue). The detected UMIs per droplet are shown on the y-axis. Droplets determined to contain a cell according to Cellranger V2 are highlighted in light colors, droplets with no cells are shown in dark colors. The red horizontal line demarks 100 UMIs. Adult murine cardiac tissue was used for this pilot.

of nuclei in high volume, which is more difficult to achieve with the other methods. Gradual filtration for example has high nuclei loss in each filtration step, while density gradient centrifugation is error-prone in a setting of low nuclei recovery. We consider FACS especially suited for processing of smaller tissue pieces, such as apical cores, which are normally smaller compared to tissue pieces obtained during heart explantation. Although the filtration and centrifugation allow increased sample multiplexing, FACS sorting takes around 20 minutes per sample, FACS sorting was chosen as the method to proceed for processing failing human heart samples. Mild FACS sorting preserved nuclei integrity and yielded intact nuclei (Figure 8).

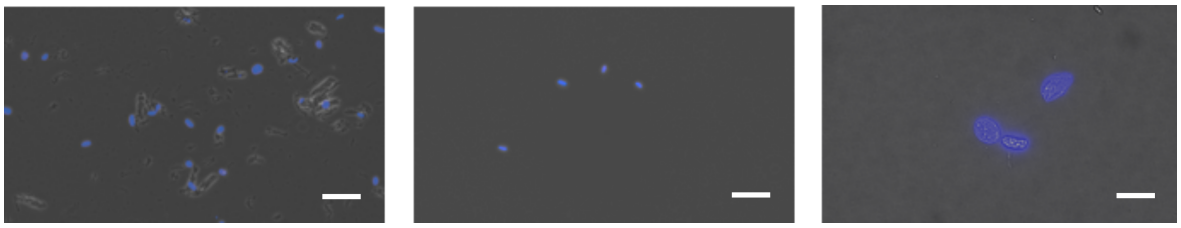

Figure 8: Isolation of intact nuclei from adult human cardiac tissue. A) Cellular homogenate before purification. Scale bar: 50μm. B) Homogenate after FACS purification. Scale bar: 50μm. C) Assessment of nuclear blebbing after nuclei isolation and FACS purification. Scale bar: 10um.

6.1.2 FACS purification strategies for unbiased cell type recovery

While FACS sorting nuclei, different bands are observed on the GFP-A vs. DAPI-A (Figure 9). The GFP-A measurement is used as a negative control for autofluorescence, while the DAPI-A gating was used to isolate NucBlue stained nuclei. Proportions of sorted particles are measured using the forward (FSC) and sidewards scatter (SSC). The P1 gate (determined via SSC-A and FSC-A) is used to remove very small or oversized particles and select the particles of interest based on size and granularity. P2 (FSC-W and FSC-H), measuring disproportion of forward scatter or particle size, and P3 gates (SSC-W and SSC-H), measuring disproportions in particle complexity, are commonly used to exclude doublets without the need for area scaling. Here gates are used to remove extreme outlier particles. Most particles (here 99.5%) are retained with the P2 and P3 gates.

Normally during FACS sorting, size gating is used to remove doublets, which are two particles stuck to each other. To investigate the implications of such a gating procedure, nuclei falling into the distinct DAPI-A bands were sorted into different tubes, processed and sequenced individually. The recovered cellular populations were investigated using snRNAseq (Figure 10).

Marker genes used for annotation are listed in chapter 2 of this monogram. While most non-myocyte nuclei are in the P5 gate (yellow), most isolated cardiomyocyte nuclei are in gate P6 (green). Previous reports have shown polyploidy in adult cardiomyocytes, suggesting that additional bands reflect polyploid nuclei, in contrast to diploid nuclei, which were possibly captured in gate P5 (Liu et al., 2010; Derks and Bergmann, 2020). Only nuclei from cardiomyocytes showed a significant increase of detected genes and UMIs when comparing

the P5 and P6 gates (Figure 11, 12). No distinct marker genes were identified separating nuclei from gates P5 and P6.

This dataset helped to optimise the FACS gating strategy. No strict doublet filtering on the FACS machine was performed. This allowed unbiased recovery of all cardiac cell types, such as cardiomyocytes.

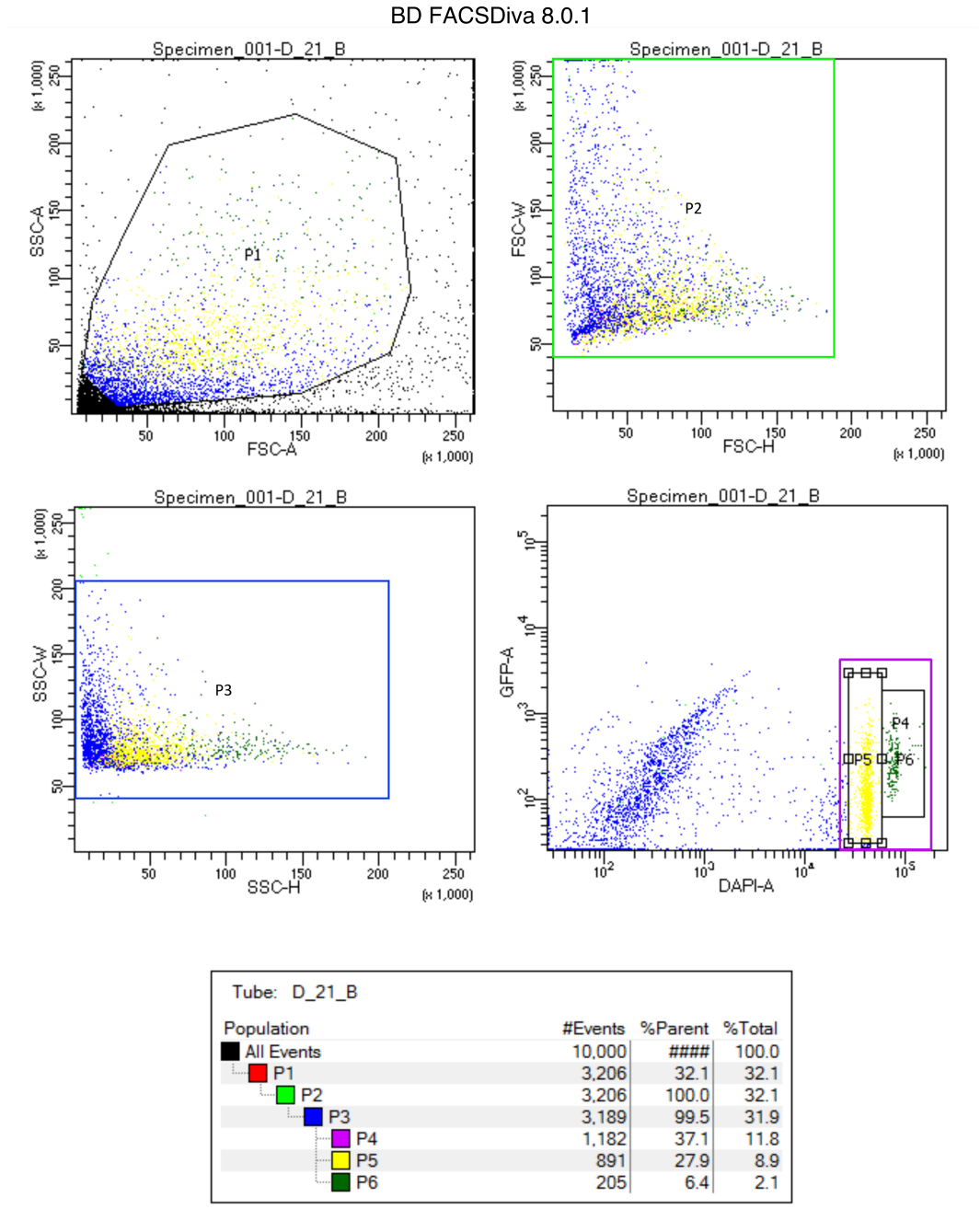

Figure 9: FACS purification of nuclei from human heart tissue homogenate. The P1 gate is used to remove very small particles representing cell debris. P4 is used to sort out NucBlue-stained nuclei. The settings used to purify all samples are shown in Figure S2. The sample used for this analysis was D21 (Sample ID BO H61 S0), a septal tissue piece from a patient with DCM without known pathogenic mutation (PVneg).

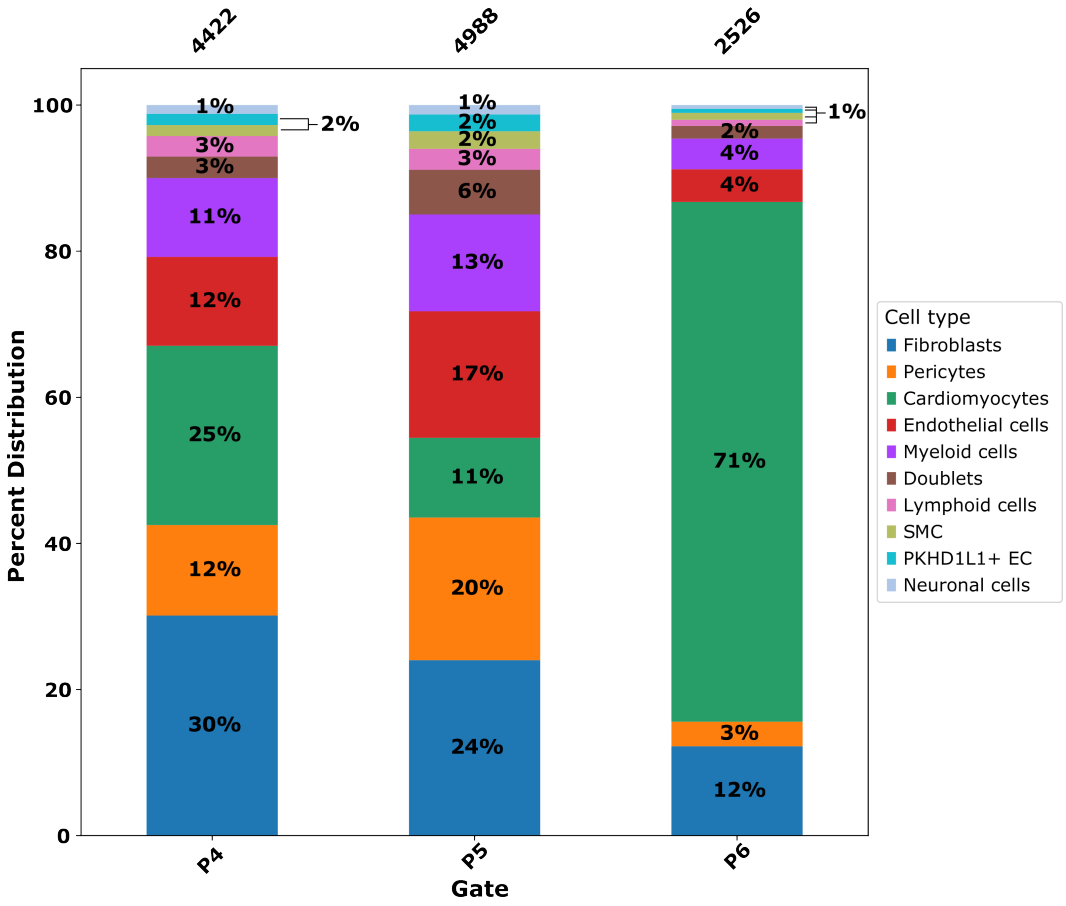

Figure 10: Cellular composition observed per FACS gates P4, P5, and P6. The cellular composition is shown as stacks, with proportions per cell type. The absolute number of nuclei included in this analysis shown on top of the bar plot. The gates from the FACS sorting are shown in Figure 9. SMC: Smooth muscle cells. EC: Endothelial cells.

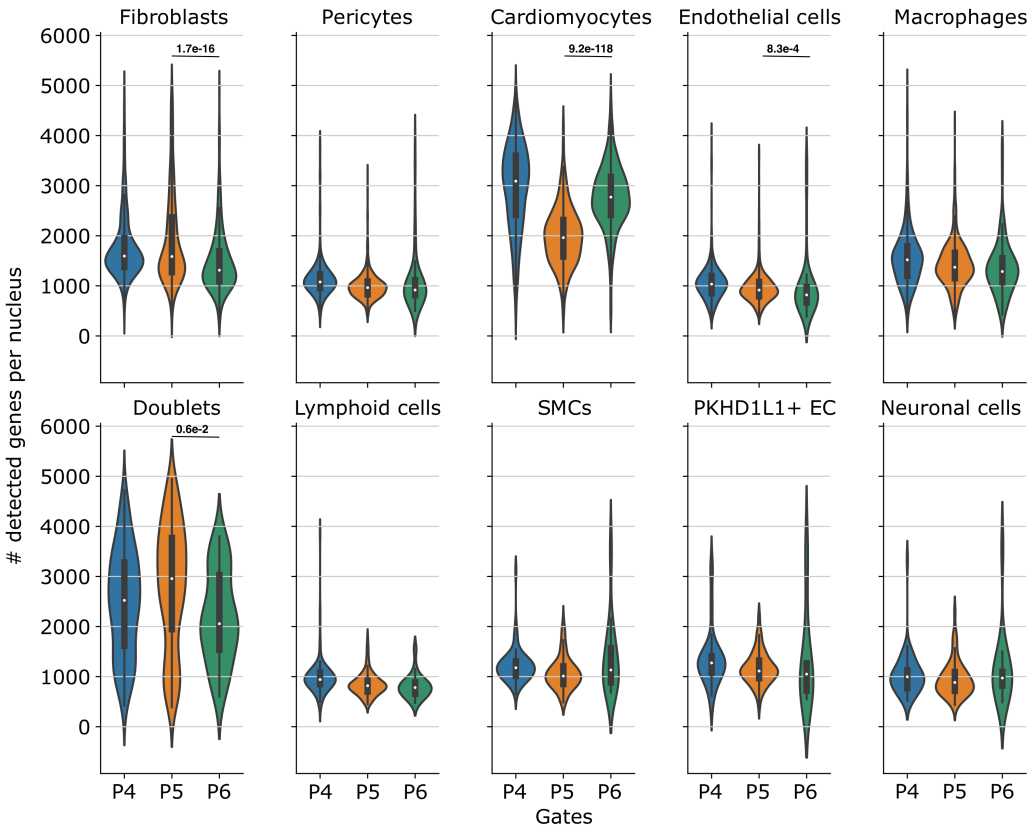

Figure 11: Violin plots of numbers of detected genes per nuclei gated out in gates P4, P5, and P6. The plot is split by identified cell type. Multiple testing adjusted p-values are shown, if ≤ 0.05. SMC: Smooth muscle cells. EC: Endothelial cells.

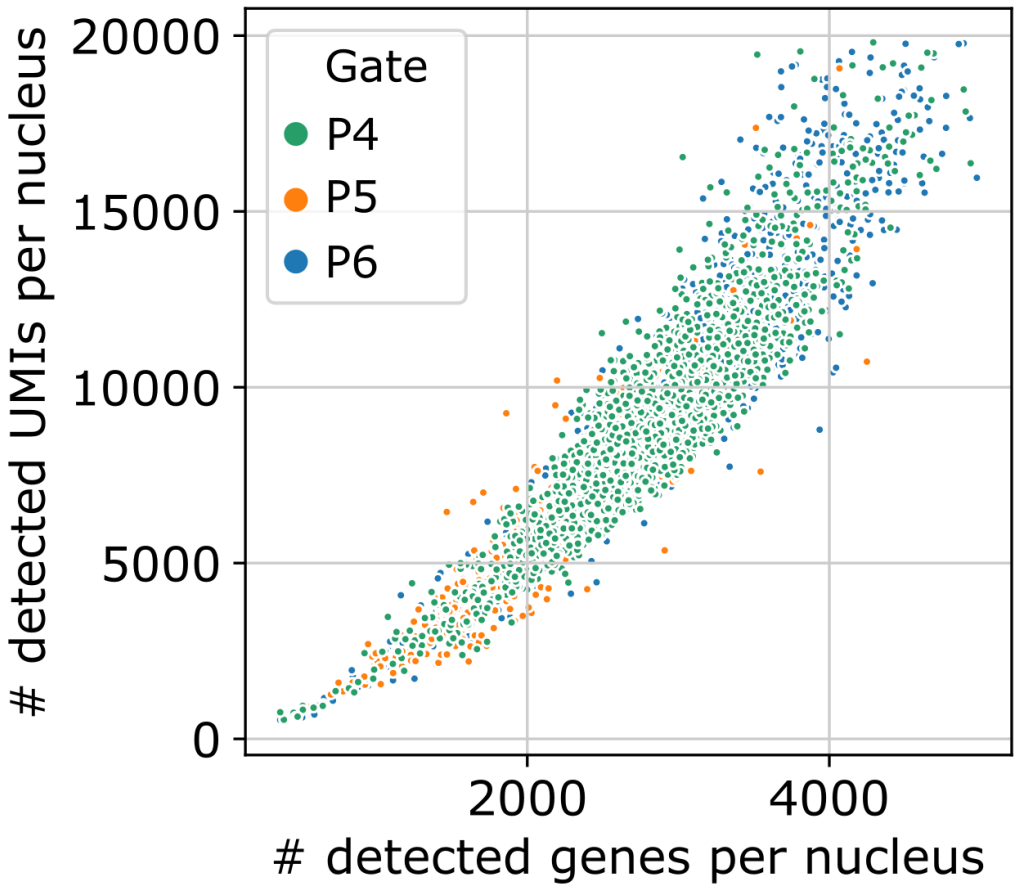

Figure 12: Scatterplot of numbers of detected genes and counts per nucleus. Data points are colored by the gate.

6.2 The fibroblast population of the healthy adult human heart revealed by single-cell sequencing

At the beginning of this project, no standardized annotation strategy or comprehensive marker list for all cell types of the adult human heart was established. The first studies on the cardiac cell composition were published in 2016 and 2018 by multiple labs (DeLaughter et al., 2016; Pinto et al., 2016a; Skelly et al., 2018; Gladka et al., 2018). Due to technical challenges, the mentioned studies report only a small number of samples, limiting the focus to highly abundant cell types. Furthermore, studies with low sample sizes do not have the statistical power to differentiate between states of the same cell type. Especially the heterogeneity of the cardiac non-myocyte fraction remained elusive until 2020, whose evaluation was part of the Healthy Human Heart Cell Atlas Consortium (Litviňuková et al., 2020). Single-cell and single-nucleus RNA-seq data of 6 regions of 14 donor hearts were generated in the course of this project and jointly analyzed to identify cardiac cell types and states. My focus within this Consortium was the analysis of the fibroblast population and region-specific transcriptional signature. This was addressed by analyzing approximately 75000 fibroblast nuclei and cells from multiple donors. Annotation results were then further refined when comparing healthy human hearts to heart failure patients, which is summarized in chapter 3 of this thesis.

The results presented in this chapter are published in Litviňuková et al. (2020); Barallobre-Barreiro et al. (2021); Quaife et al. (2022).

6.2.1 Marker genes to identify fibroblasts and other cell types in the heart

A marker gene list with specific cell type markers was collected based on literature research and analysis of the generated scRNAseq and snRNAseq data.

Differentially expressed genes per cell type are returned when testing for significant enrichment or depletion of a gene with a Wilcoxon rank-sum test (see Methods). Enriched Gene-Ontology terms were then computed for the top 100 upregulated genes using GProfiler2 (Kolberg et al., 2020). Although this approach can often give quick functional insights into identified clusters, no cell-type/-state label is returned. A far more efficient approach for annotation was to extract membrane-bound and secreted proteins from the list of upregulated genes and continue with literature research. Lists of membrane-bound and secreted proteins were obtained from the Human Protein Atlas (HPA) (Uhlén et al., 2015; Thul et al., 2017; Uhlén et al., 2019). On top of that, pathway databases such as KEGG, Reactome or Wiki

Pathways contain manually curated gene lists, which can then be used for gene set enrichment analysis (Kelder et al., 2012; Fabregat et al., 2018; Kanehisa et al., 2021). Future single-cell projects, especially in the heart, will benefit from the annotation efforts described in this thesis and other projects published since 2017. A list of literature-known marker genes used to annotate cell types is shown below (Table 12 and Figure 13).

Cell type	Marker gene
Adipocytes	ADIPOQ, GPAM, LEP
Cardiomyocytes	TNNT2, MYH7, MYL2
Endothelial cells	VWF, PECAM1, CDH5
Fibroblasts	DCN, GSN, PDGFRA
Lymphoid cells	THEMIS, CD8A, IL7R
Myeloid cells	F13A1, C1QA, CD68
Neuronal cells	NRXN1, NRXN3, PLP1
Mural cells	
Pericytes	KCNJ8, CD38, ABCC9
Smooth Muscle Cells	MYH11, SLIT3, ITGA8
Mast cells	KIT, CPA3, TPSB2

Table 12: Cardiac cell type marker genes. Mast cells here are part of the myeloid cell population, but split up as a separate cluster in the Failing Heart Project described in chapter 3.

It is worth highlighting, that fibroblasts showed similarities to adipocytes, endothelial cells, mural cells, and neuronal cells (spearman rho=0.96 (all genes), 0.93 median correlation with all cell types). Several in the literature described marker genes for fibroblasts don't show a specific expression exclusive for fibroblasts, but are also shared between multiple cell types. Two examples are Cx45 (GJC1) and Cx40 (GJA5), which have both been described to be specific for cardiac fibroblasts, but are detected with higher RNA levels in Mural cells, endothelial cells and adipocytes.

6.2.2 scRNAseq and snRNAseq of fibroblast in the healthy adult human heart

In total, nearly 500,000 cells and nuclei were analyzed, delineating 9 different cell types (Figure 14A). Some cell types show distinct regional profiles, such as cardiomyocytes or fibroblasts (Figure 14B). Fewer fibroblast cells were observed in scRNAseq data with a total yield of

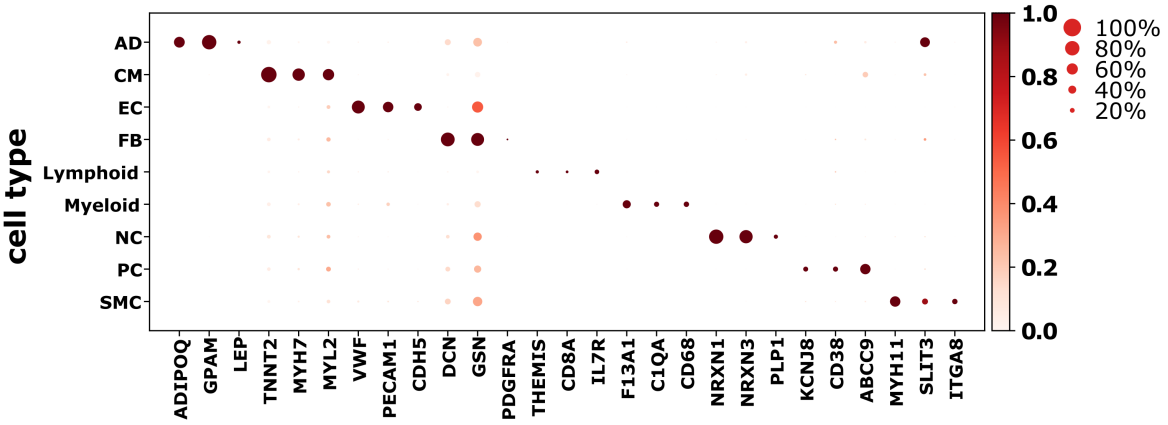

Figure 13: Dotplot shows selected marker genes of cardiac cell types. Dot size represents fraction of expressing cells/nuclei within a cell type; color, mean expression. Expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states).

2857 fibroblast cells across all donors (Figure 14C), indicating loss of FBs during isolation or insufficient digestion. On average, 1% of the isolated ventricular cells are FBs compared to 18% abundance in the snRNAseq data. This shows that snRNAseq is better suited for studying cardiac fibrosis than scRNAseq. Transcriptomes of cells and nuclei correlated with a spearman correlation of 0.75 across all fibroblast states. This difference is due to many factors, such as differences in transcriptional complexity (detected UMIs per gene), enriched identification of mitochondrial-genome-encoded and ribosomal genes, and nuclear export dynamics vs. cytosolic RNA stability (Figure 14D, E).

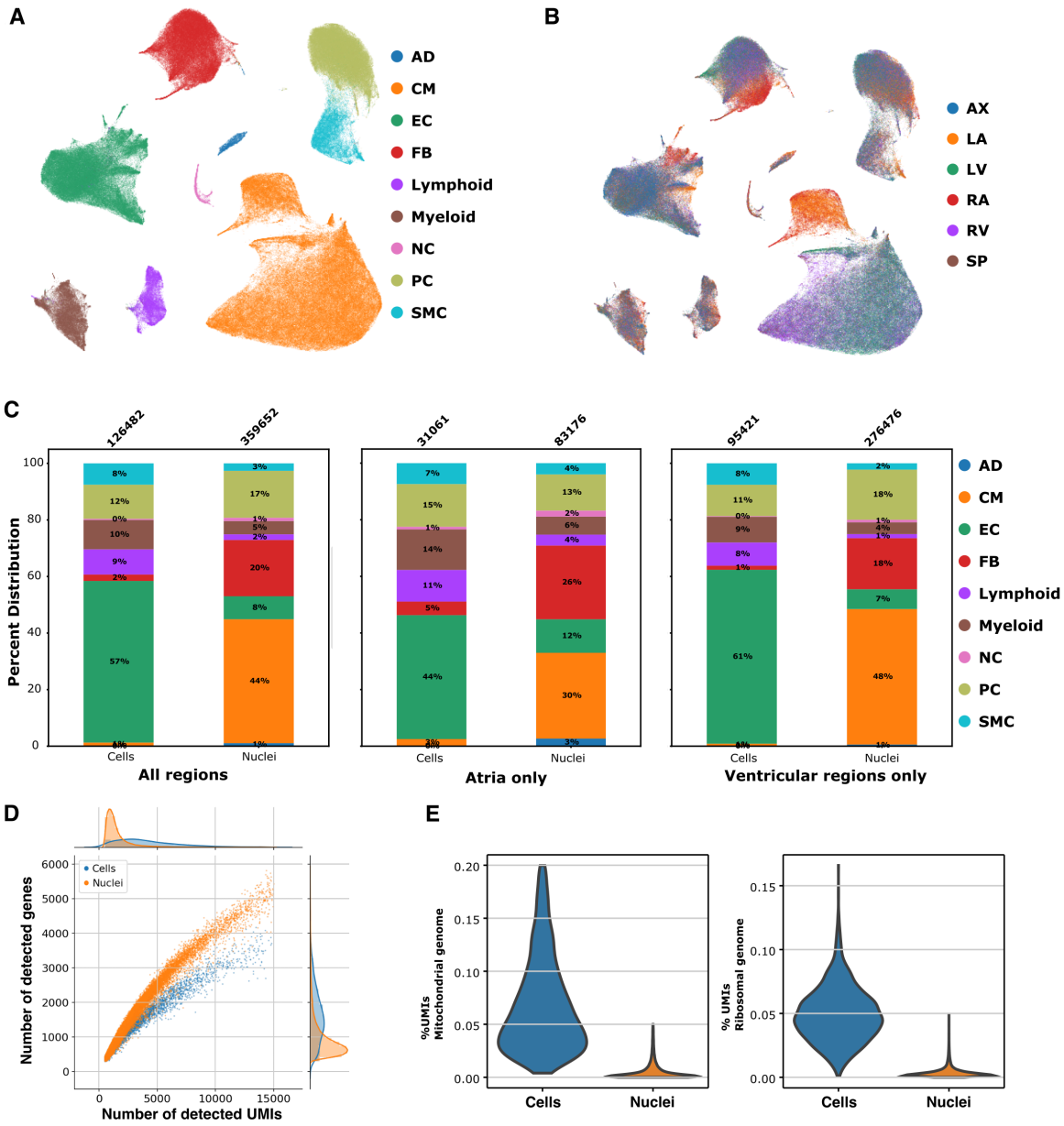

Figure 14: Regional abundance of fibroblasts and quality metrics of fibroblast cells and nuclei. A) UMAP embedding delineated 9 cell types. B) UMAP embedding of the major cell types colored by region. Notably, atrial and ventricular cardiomyocytes show distinct transcriptional signatures. C) Stacked bar plots show cell type distribution across regions in scRNAseq (Cells) and snRNAseq (Nuclei). D) Scatterplot shows number of genes (n genes) and number of UMIs (n counts) per fibroblast. On top and right probability densities are shown for n counts and n genes. For scRNAseq more UMIs are detected per gene. E) Violin plots show percent UMIs mapping to mitochondrial (left) and ribosomal genes (right) for fibroblasts only. This figures were part of the publication Litviňuková et al. (2020).

6.2.3 Fibroblast heterogeneity in the healthy adult human heart

To evaluate similarities among different fibroblasts and identify fibroblast states, subclustering was performed. Five distinct fibroblast states were identified (FB1-5), plus two clusters (FB6 and 7) resembling chimeric profiles with fibroblasts and vascular cells and cardiomyocytes (Figures 15A-G). Whether these clusters represent true biological states, elevated background RNA noise (soup), or multiplets is unclear. Using RNA in situ hybridization, no overlap of CM and fibroblast marker genes in the same cell was identified, rather indicating that those states represent multiplets.

FB1 showed only a low number of enriched markers, such as SCN7A, BMPER, or ACSM1 with no detection of specific marker genes. Due to their low number of discriminative markers and increased proportion in the four ventricular regions (AP, S, LV, RV), this population was termed canonical ventricular fibroblasts. In contrast, FB2 showed enriched expression for CFH, ID4 and KCNT2. As for FB1 and FB2, canonical genes were similarly expressed in all FB from the respective chambers, I propose that these genes define a basal, chamber-specific FB expression program. This observation is in line with previously described differences between atrial and ventricular cardiac fibroblasts, such as stronger profibrotic response in the atria, for example in the context of atrial fibrillation (Burstein et al., 2008) (Figure 15D).

FB3 significantly upregulated cytokine receptors such as OSMR and IL6ST with decreased expression of extracellular matrix proteins. Gene set enrichment analysis using the Oncostatin-M pathway gene set showed elevated expression of genes involved in this pathway (Figure 15C) (Dey et al., 2013; Abe et al., 2019). Two fibroblast populations were identified with upregulation of genes, involved in extracellular matrix production. FB4 showed upregulation of TGFβ activated genes such as POSTN, TNC, and FN1, resembling activated fibroblasts. In a collaboration with the Barton lab the physiological role of the microprotein encoded by LINC01013 was studied, a gene that was so far annotated as non-coding. FB5, in contrast, upregulated genes involved in extracellular matrix organization, remodeling, or cleavage, such as matrix metalloprotease 2 (MMP2), PCOLCE2, or Fibulin 2 (FBLN2). FB3 were less abundant in the left ventricle, while FB4 and FB5 are less abundant in the right atrium (Figures 15E). FB3-5 have been validated by using fluorescence in situ hybridization using RNAscope v1 probes (Figures 15F, G).

The transcriptional signature per fibroblast cell state provides valuable information on studying cell state-specific gene expression. For example, Quaife et al. have studied the microprotein-encoding lncRNA LINC01013, which is higher expressed in TGFβ stimulated fibroblasts

(Quaife et al., 2022). This Healthy Human Heart Cell Atlas allowed analysis of cell state-specific expression of LINC01013, which was found to be enriched in fibroblasts and specifically expressed in ECM-producing, TGFβ stimulated FB4.

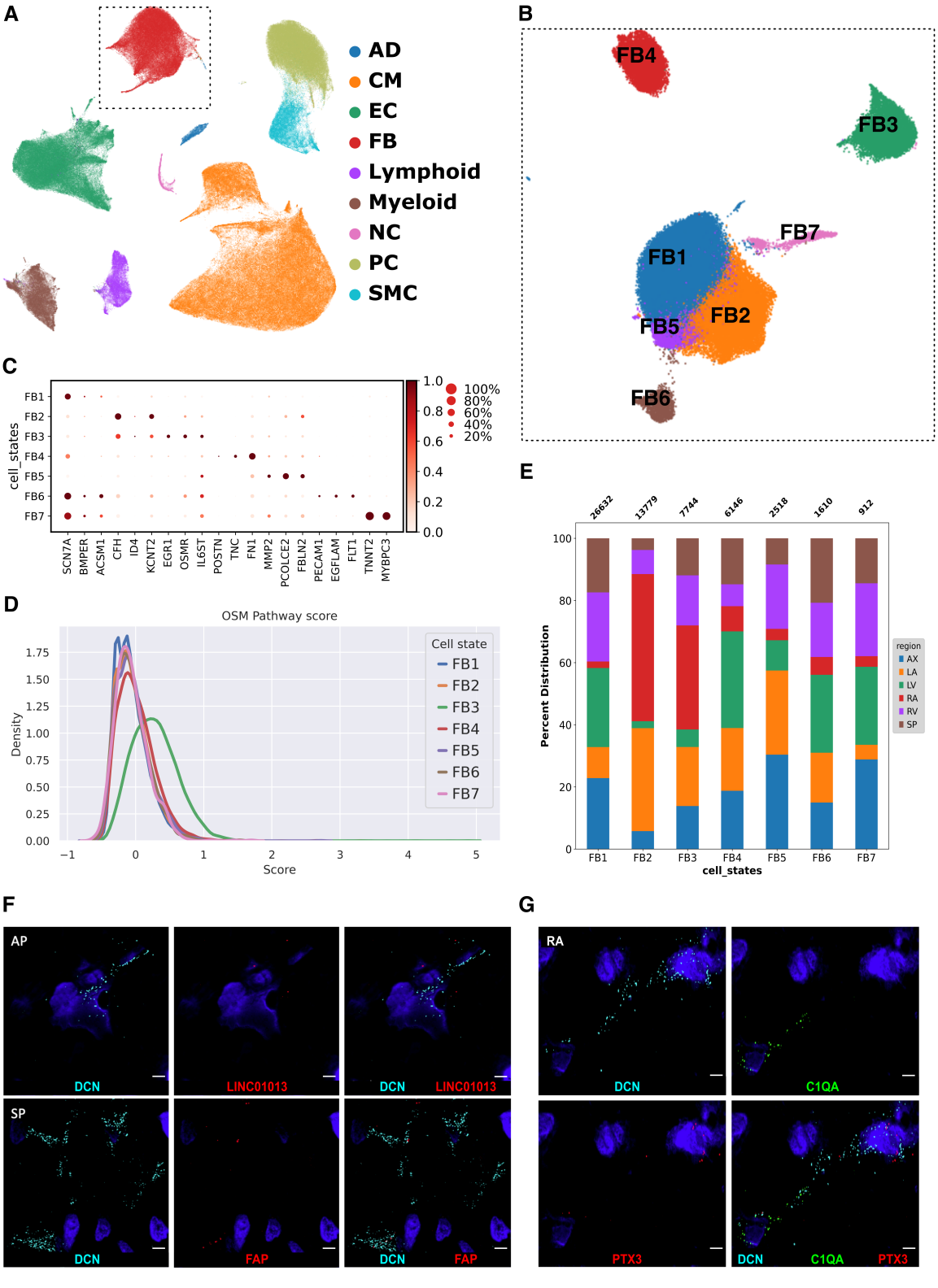
A
AD
CM
EC
FB
Lymphoid
Myeloid
NC
PC
SMC
B
FB4
FB3
FB7
FB1
FB2
FB5
FB6
C
cell_states
FB1
FB2
FB3
FB4
FB5
FB6
FB7
SCN7A
BMPER
ACSM1
CFH
ID4
KCNT2
EGR1
OSMR
IL6ST
POSTN
TNC
FN1
MMP2
PCOLCE2
FBLN2
PECAM1
EGFLAM
FLT1
TNNT2
MYBPC3
1.0
0.8
0.6
0.4
0.2
0.0
100%
80%
60%
40%
20%
D
OSM Pathway score
Density
Score
Cell state
FB1
FB2
FB3
FB4
FB5
FB6
FB7
E
26632
13779
7744
6146
2518
1610
912
Percent Distribution
cell_states
region
AX
LA
LV
RA
RV
SP
F
AP
DCN
LINC01013
SP
FAP
G
RA
C1QA
PTX3

Figure 15: Fibroblast states in the healthy adult human heart. A) UMAP embedding delineates 9 cell-types. For subclustering, only those nuclei annotated as fibroblast (red) were used. B) UMAP depicting FB states in all tissue samples. C) Dotplot shows selected marker genes of FB states. Dot size represents fraction of expressing cells within a cluster; color, mean expression. Expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states). D) Oncostatin M patway was enriched in FB3. The gene list used for scoring can be found in Supplementary Table 12 of (Litviňuková et al., 2020). E) Regional distribution per FB state. FB2 and FB3 are enriched in atria (left and right), while FB1, FB4 and FB5 are enriched in ventricular samples (left, right, apex and interventricular septum). Single-molecule fluorescent in situ hybridization targeting LINC01013, fibroblast activated protein (FAP) and PTX3 confirmed F) FB4, FB5 and G) FB3. DCN is used as a FB marker, C1QA for macrophages, nuclei are DAPI-stained (blue). Scale bars, 5um. This figures were part of the publication Litviňuková et al. (2020).

6.2.4 Regional differences of fibroblast gene expression

Separate clustering of atrial and ventricular FBs recapitulated the populations described above, such as an OSM-stimulated population in each chamber (aFB4 and vFB3) (Figure 16A-D). In addition, distinct chamber-specific extracellular matrix (ECM) producing FBs were identified (aFB2 versus vFB2) (aFB1 versus vFB4) (Figure 16E). For example, while COL15A1, COL6A6, and COL4A4 were higher expressed in the atria, COL8A1, COL6A3, and COL1A1 showed elevated expression in the ventricles. Similarly, ECM modulators showed chamber-specific expression patterns. Furthermore, other region-enriched gene signatures were identified. For example, APOD, which is responsible with the lecithin:cholesterol acyltransferase (LCAT) for esterification of cell-derived cholesterol (Francone et al., 1989), was significantly higher abundant in the ventricles. CFH, a protein important for complement activation and wound healing (Wu et al., 2009; Argenziano et al., 2019), was atrial enriched. This was validated using single-molecule RNA fluorescent in situ hybridization (Figure 16G-H).

Together with the lab of Manuel Mayr, the regional specificity of ECM genes was investigated (Barallobre-Barreiro et al., 2021). ADAMTS5, the highest expressed ADAMTS gene in cardiac fibroblasts cleaving proteoglycans such as versican, was higher expressed in the ventricles than in atria. This finding was independently confirmed using proteomics data (Figure S8C from Barallobre-Barreiro et al. (2021)). VCAN expression was downregulated in

ventricles and higher expressed in the atria. Taken together, this dataset shows differences in ECM composition in the different heart chambers suggesting different mechanical properties.

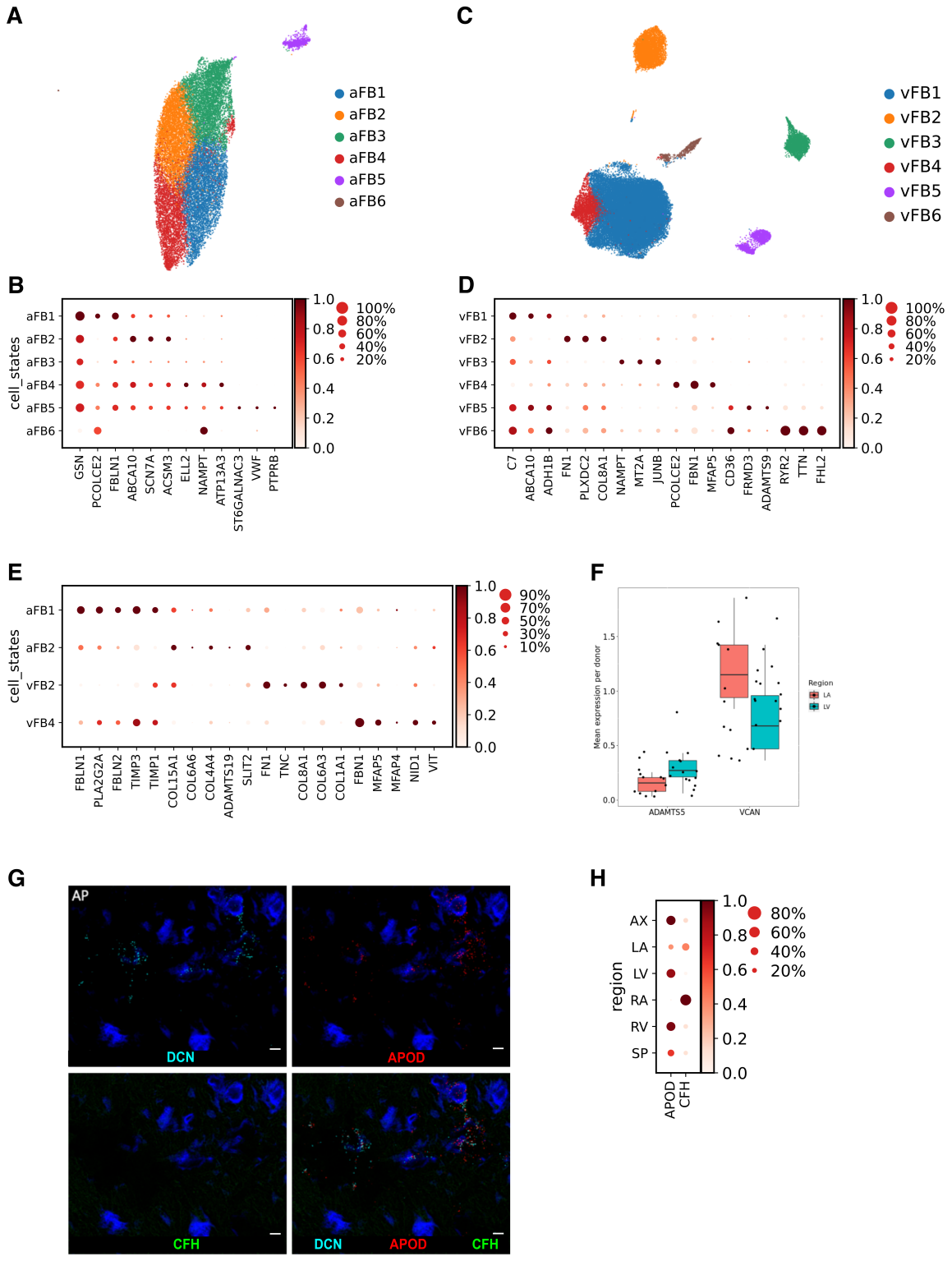

Figure 16: Fibroblast states in the atria and ventricles. A) UMAP embedding delineates 6 fibroblast states in the atria. B) Dotplot shows selected marker genes of atrial FB states. Dot size represents fraction of expressing cells within a cluster; color, mean expression. Expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states). C) UMAP embedding delineates 6 fibroblast states in the ventricles. D) Dotplot shows selected marker genes of atrial FB states. Dot size represents fraction of expressing cells within a cluster; color, mean expression. Expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states). E) Dotplot shows selected atrial and ventricular-enriched ECM genes. Dot size represents fraction of expressing cells within a cluster; color, mean expression. Expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states). F) Pseudobulked average expression for ADAMTS5 and VCAN across cardiac fibroblasts in the left atrium and ventricle per donor. G) Single-molecule RNA fluorescent in situ hybridization targeting ventricle enriched APOD and CFH (atrial enriched) in an apical sample. DCN is used as a FB marker, nuclei are DAPI-stained (blue). Scale bars, 5um. H) Dotplot shows expression of ventricle enriched APOD and atrial enriched CFH. This figures were part of the publication Litviňuková et al. (2020).

6.3 From the Healthy Heart Cell Atlas to understanding heart failure

The Healthy Heart Cell Atlas allowed studying transcriptional and compositional changes of previously characterized cell types and states in genetic cardiomyopathies, a common cause of end-stage heart failure. Here in total 196 snRNAseq libraries from ventricular samples of 61 heart failure patients and 18 healthy controls were analyzed for changes in cellular composition and transcription. Common heart failure and genotype-specific signatures were quantified, upending the current dogma that heart failure results in a final common pathway. The results of this chapter have been published in Reichart et al. (2022), of which I am a shared first author and co-corresponding author.

6.3.1 Patient samples

Heart failure patient samples were collected for snRNAseq from biobanks located at the Heart and Diabetes Center (HDZ) in Bad Oeynhausen (Germany), the Harvard Medical School (USA), the Mazankowski Hospital in Alberta (Canada), and the Imperial College London (UK). Cardiac tissue sections from in total 61 genotype-stratified patients were available (Figure 17), with an additional six healthy donor hearts obtained from the HDZ in Bad Oeynhausen. The six healthy donors were combined with ventricular tissues from healthy donors, which have been processed and published during the Heart Cell Atlas project, described in chapter 4.2 of this thesis. In total 61 patients with pathogenic mutations in different known DCM- and ACM-associated genes were included and compared to 8 patients with no identified pathogenic mutation (Figure 17A). Major genotype subgroups ($\leq$6 patients) of DCM comprised patients with mutations in LMNA, TTN, and RBM20 genes, and for ACM patients with a mutation in PKP2. More males than females were in the study cohort (Figure 17B), which is due to increased male prevalence of DCM and ACM (Lyden et al., 1987; Fairweather et al., 2013). The majority of available cardiac tissue was obtained from explanted ventricular tissues, both LV and RV samples were available and sequenced, allowing comparison of left and right ventricular changes (Figure 17C). For tissue material obtained during left ventricular assist device (LVAD) implantation, only the apical core, a piece of the left ventricular free wall, was available. Samples of patients with rare mutations or low sample size have been also sequenced but were not as deeply analyzed as the other genotypes due to the lack of statistical power to define disease mechanisms (Figure 17D). All sequencing libraries are now publically available to the scientific community to further study genotype-specific differences in heart failure.

6.3.2 Differences in clinical metadata between patients

Patients in individual genotype subgroups were analyzed for differences in clinical records. This information is independent of the sequencing results and was evaluated after patient sample collection from the biobanks. For patient characterization, hard data (measurable parameters) and soft data (qualitative information) were collected, but only hard data were statistically evaluated. Clinical information per patient were provided in T1 of the online supplement.

Left ventricular inner diameter diastole (LVIDd, mm) and systole (LVIDs, mm) were highest in RBM20 patients, with differences to TTN and LMNA patients (Figure 18A, B). RBM20 mutated patients showed the lowest age at transplantation and highest kidney function, quantified by the estimated glomerular filtration rate (eGFR), compared to the other genotype groups (Figure 18C, D). For LMNA and TTN patients, a Spearman correlation of -0.6 (p-value=0.009) was observed for heart failure duration and eGFR. PKP2 patients showed the highest right ventricular end-diastolic diameter (RVEDD) (Figure 18E). RBM20 patients were the only DCM genotype where no cardiac resynchronization therapy (CRT) was reported, while 4/12 in the LMNA group, 5/12 in the TTN group and 2/8 in the PVneg group were reported. An implanted CRT device stimulates both left and right ventricles and thereby helping to recover normal heart rhythm, but also fulfills the same functionality as a implantable cardioverter defibrillator (ICD).

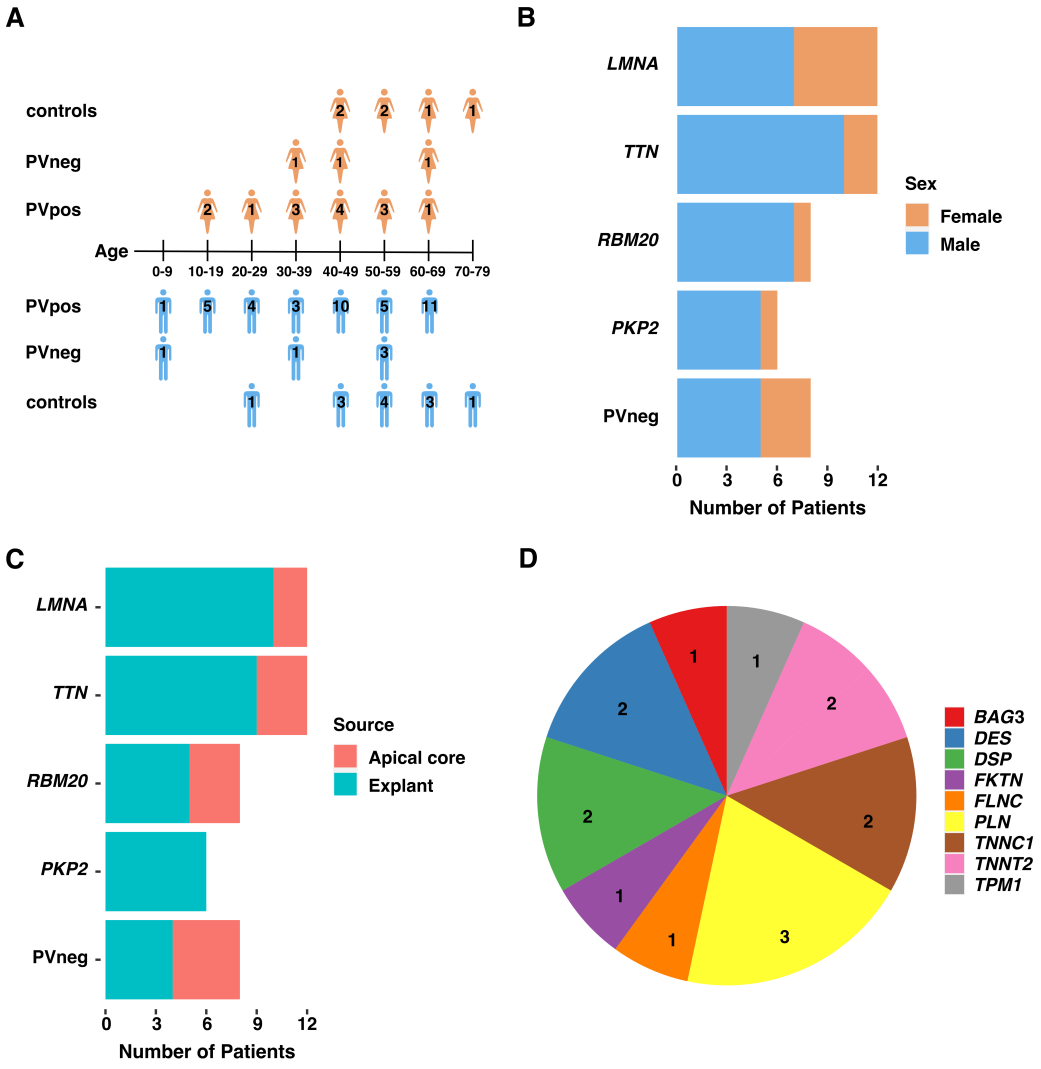

Figure 17: Patients included in the heart failure cohort. A) Age and sex distribution of patients with pathogenic variant (PVpos), no identified pathogenic variant (PVneg), and healthy controls. The number of patients and donors per age bin are shown on the patients. B) Number of males and females in the main genotype classes, genotypes with more or equal to 6 patients. C) Tissue sources per genotype group. For explant tissue, multiple regions are available, while for patients undergoing LVAD implantation, only apical cores were obtained. D) For some genotypes only low number of patients were available, but have not been analysed in depth. Figure A) was part of the publication Reichart et al. (2022).

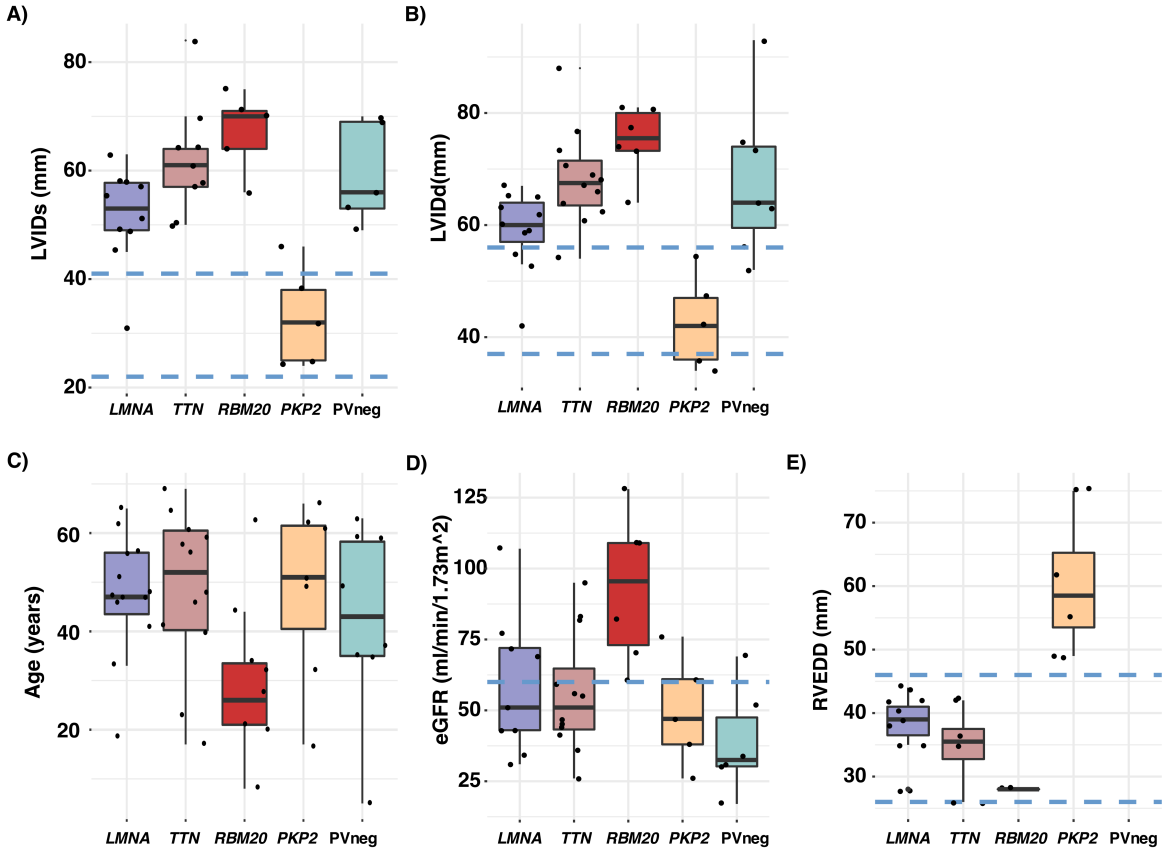

Figure 18: Evaluation of clinical parameters between genotypes. A) Left ventricular inner diameter in systole and B) diastole (LVIDd, LVIDs), C) patient age at transplantation, D) GFR and E) RVEDD distribution plotted across genotypes. Dotted lines indicate normal ranges as reported in the clinical literature (Harkness et al., 2020; Mewis et al., 2006). A GFR of 60 or higher is considered as healthy (Wetzels et al., 2007). Clinical information per patient were provided in T1 of the online supplement

6.3.3 Quality of the human heart failure samples

Patient samples were sequenced using snRNAseq. Criteria, which were determined to be essential for determining library quality, are summarized in Figure 19: First, the barcode-rank plot indicates droplets containing cells and the relative proportion of ambient RNA levels in the dataset. Second, the cDNA yield (ng/mul) was determined to be an important measure to exclude samples of low quality. This measure was returned as the most dominant feature in a logistic regression model, trained on metrics obtained during and after library preparation to classify libraries of high and low quality (data not shown). Figure 19A shows, that samples with low UMI counts tend to have a low cDNA yield, which can be explained by increased RNA degradation, comparable to the RIN value in bulk RNAseq. This value is rarely reported despite its significant meaning for sample and library quality. Third, the estimated number of nuclei in relation to the originally targeted nuclei number, which was for this study always 5000 nuclei (Figure 19B). Sequencing libraries were generated in-house and in two additional centers (Harvard Medical School and Imperial College London), with 12 healthy controls from Cambridge (CB) and Harvard Medical School/the Mazankowski Hospital in Alberta being published in Litviňuková et al. (2020) as the Healthy Human Heart Cell Atlas (HCA) (Figure 19C). Despite the large range of storage times, we have not identified this as a predictor of poor quality, suggesting that preprocessing before freezing and interruption of the cold chain are more determining factors of sample quality.

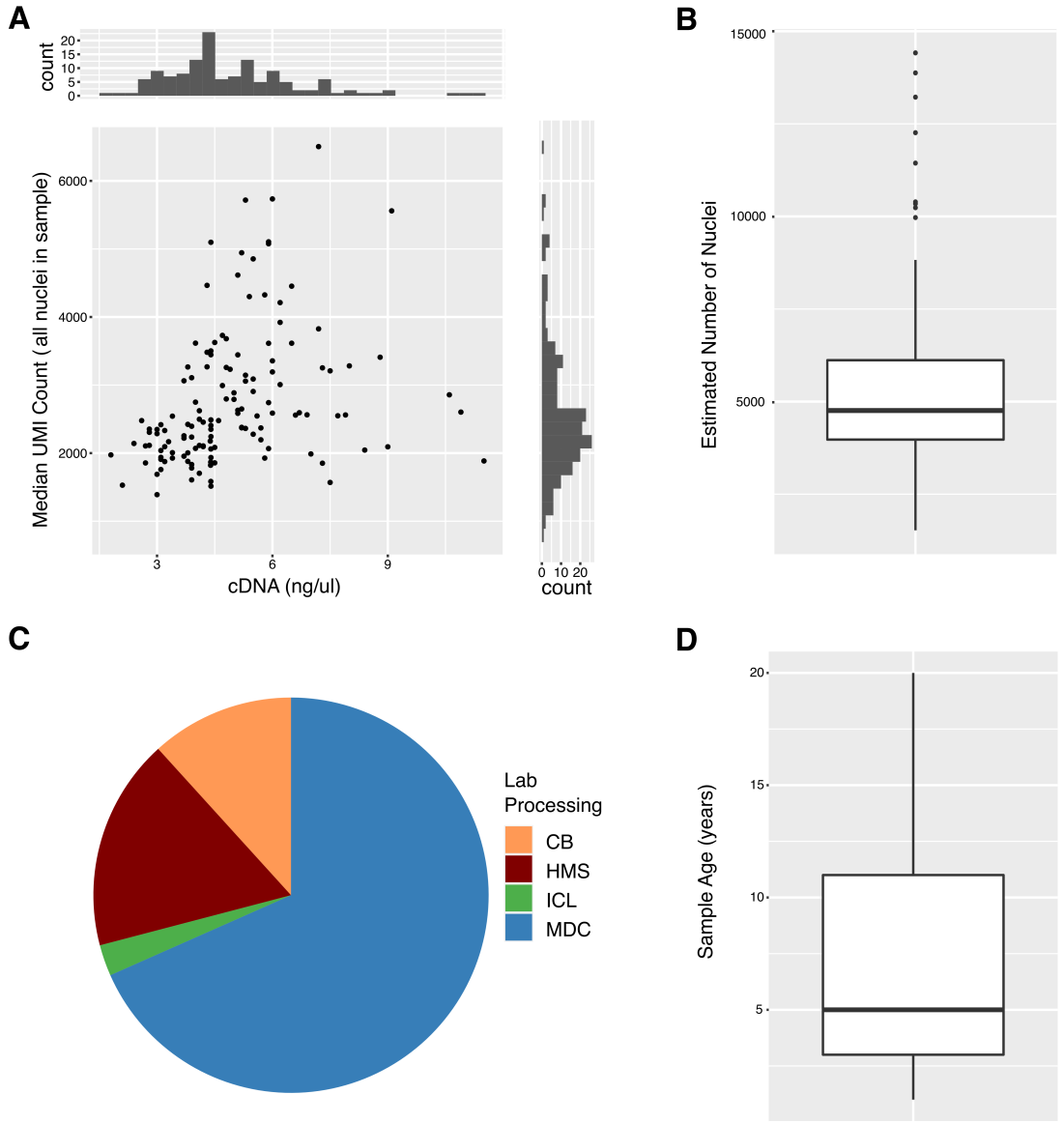

Figure 19: Sample quality information for samples included in the DCM Heart Cell Atlas study. A) Correlation of median UMI counts across all nuclei per sample compared to the cDNA concentration measured per sample. Samples with low cDNA concentration tend to have low numbers of UMIs recovered. B) Estimated number of nuclei per sample. C) Origin of libraries included in this study. D) Distribution of storage time in years for all samples. Sample quality were provided in T2 of the online supplement.

6.3.4 Cell type and state annotation in heart failure

Cell-type and -state marker gene sets established during the healthy human heart project were projected on a healthy control and heart failure combined manifold. Overall we identified 9 major cardiac cell types with 71 distinct states (Figure 20A). All cell types were also characterized in the Healthy Heart Cell Atlas project. Pericytes and smooth muscle cells were jointly annotated as mural cells, and mast cells showed a transcriptional distinct signature from myeloid cells. For some cell types we identified new states, leading to a subdivision of previously annotated cell states, which is highlighted with ".". One example is vFB1, which is now separated into vFB1.0 (canonical ventricular FBs), vFB1.1 and vFB1.2, with a distinct marker gene sets introduced in this chapter. Identification of new cell states is due to the near doubling of nuclei (n=881,081) that are studied in this study compared with the previously described 487,106 nuclei plus cells. Furthermore, processing of new samples (control and disease) using the sensitivity-improved 10x V3 3' chemistry revealed smaller transcriptional differences. Results of the presented work have been partially published in Reichart et al. (2022).

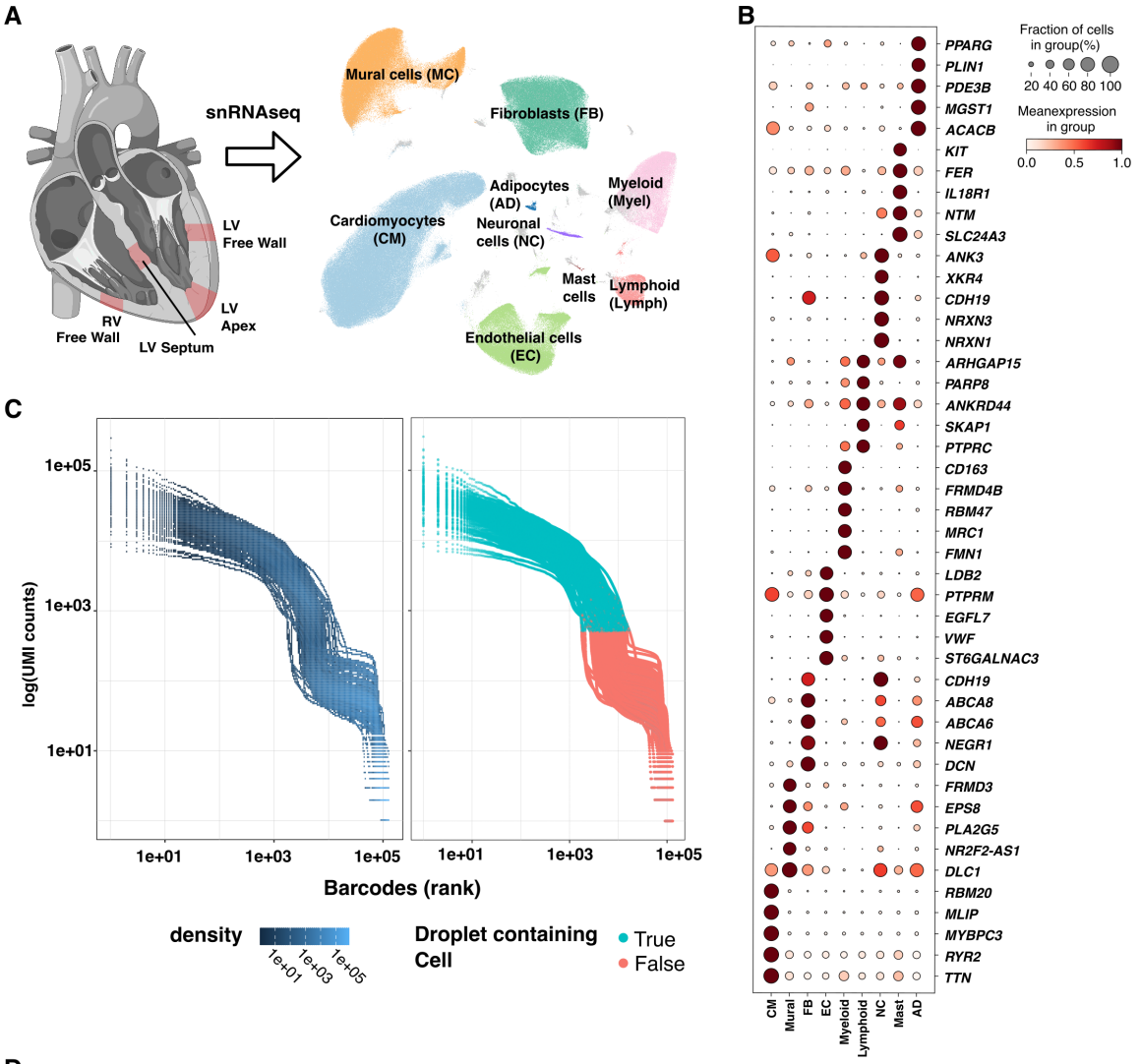
A
snRNAseq
LV
Free Wall
LV
Apex
RV
Free Wall
LV Septum
Mural cells (MC)
Fibroblasts (FB)
Adipocytes
(AD)
Myeloid
(Myel)
Cardiomyocytes
(CM)
Neuronal
cells (NC)
Mast
cells
Lymphoid
(Lymph)
Endothelial cells
(EC)
B
Fraction of cells
in group(%)
20 40 60 80 100
Meanexpression
in group
0.0 0.5 1.0
PPARG
PLIN1
PDE3B
MGST1
ACACB
KIT
FER
IL18R1
NTM
SLC24A3
ANK3
XKR4
CDH19
NRXN3
NRXN1
ARHGAP15
PARP8
ANKRD44
SKAP1
PTPRC
CD163
FRMD4B
RBM47
MRC1
FMN1
LDB2
PTPRM
EGFL7
VWF
ST6GALNAC3
CDH19
ABCA8
ABCA6
NEGR1
DCN
FRMD3
EPS8
PLA2G5
NR2F2-AS1
DLC1
RBM20
MLIP
MYBPC3
RYR2
TTN
CM
Mural
FB
EC
Myeloid
Lymphoid
NC
Mast
AD
C
log(UMI counts)
1e+05
1e+03
1e+01
1e+01
1e+03
1e+05
1e+01
1e+03
1e+05
Barcodes (rank)
density
1e+01
1e+03
1e+05
Droplet containing
Cell
True
False
D
n_genes
n_counts
percent_mito
percent_ribo
solo_score
CM
Mural
FB
EC
Myel
Lymph
NC
Mast
AD

Figure 20: Quality of the generated snRNAseq data. A) 4 cardiac regions were sampled and nuclei were isolated. LV Apex was only available as apical cores received upon VAD implantation. The other cardiac regions were sampled from explanted hearts. The obtained 880.081 nuclei, post quality filtering, identified nine cardiac cell types. B) Marker genes per cardiac cell type. C) Barcode rank plot for the first 100.000 droplets. A two-plateau ("shoulder-and-knee") shape is optimal, in which the first plateau represents droplets containing a nucleus. The lower plateau represents the level of technical noise, ambient RNA, in the dataset. On the right, all droplets containing nuclei as identified by Cellrangers EmptyDrops are shown (turquoise), while empty droplets (red) are not included in any downstream analysis. D) Quality metrics for the nine cardiac cell types. n genes: Identified genes per nucleus. n counts: Identified UMIs per nucleus. percent mito: Percentage of unique reads mapping to the mitochondrial genome. percent ribo: Percentage of unique reads mapping to ribosomal (RPS and RPL) genes. solo score: Softmax-score representing the likelihood of nuclei being doublets. This figure was part of the publication Reichart et al. (2022).

6.3.5 Compositional analysis of cardiac cell types in the failing human heart

Before computing compositional changes of cardiac cell types during heart failure, the robustness of cell type abundance measurements was evaluated by comparing the captured composition of independently sequenced tissue pieces from the same patients' cardiac region. This is of importance as fibrosis is not evenly distributed across the myocardium. Instead, areas with high or low fibrosis content or different amounts of immune cell infiltrates are observed. The impact of patchy fibrosis and a potential sampling bias on cell type abundance measurements is unknown. 50 mg-sized tissue pieces from the same cardiac region were independently processed, sequenced and cell-types were annotated (Figure 21A). A high Pearson correlation of cell types abundances was observed in replicates (r=0.74-0.99), suggesting that 50 mg-sized tissue pieces show small bias by local effects (Figure 21B, C). Replicated samples from the same patient were merged for downstream analysis.

To account for differences in compositional shifts across patient samples, cell-type and -state proportions are tested for significance using prior center log ratio (CLR) transformed proportions. Next, differences in composition between the left ventricular free wall, apical core and intraventricular septum were computed. Cell type and state abundances were comparable across the three regions and are therefore jointly reported as LV (Figure 21D).

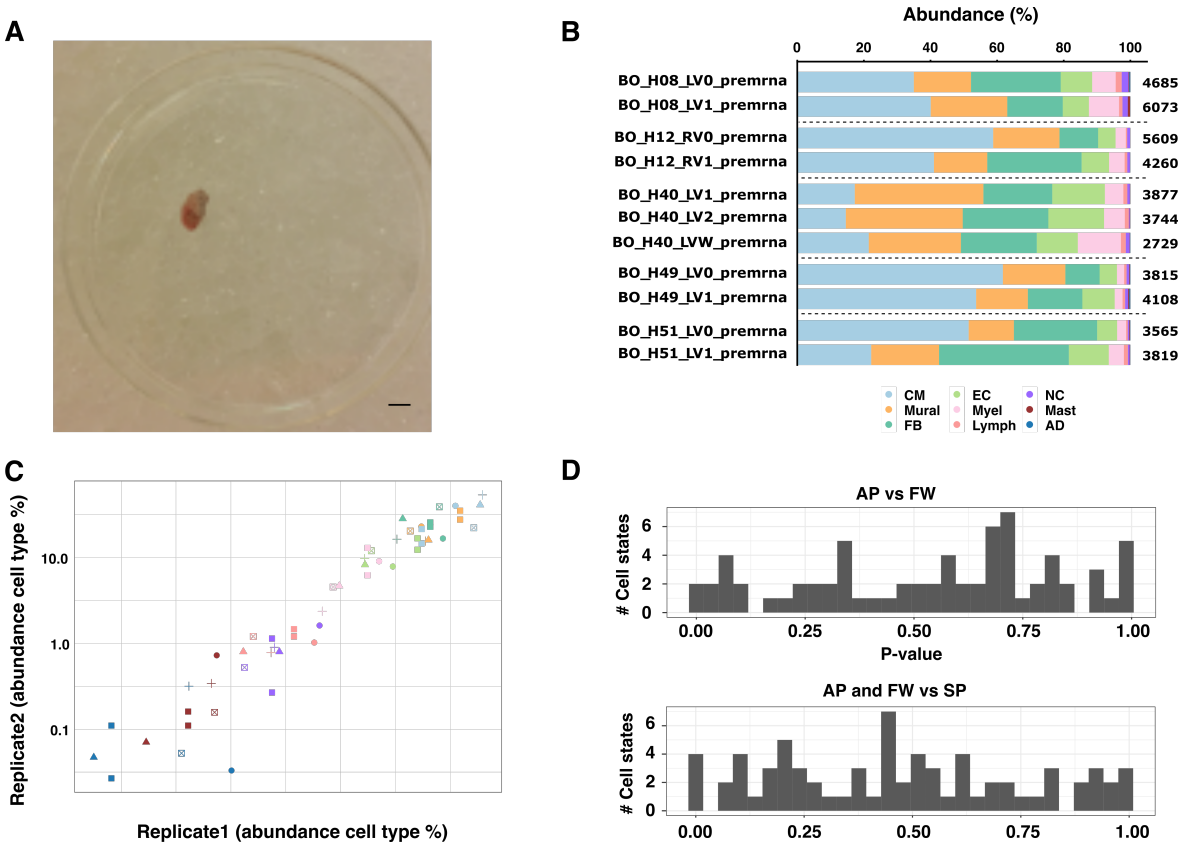

Figure 21: Regionality of cellular composition. A) Exemplary image of a processed cardiac tissue piece. Scale bar: 1cm B) Composition of cardiac cell types per sample. Replicates are separated by dashed lines. The analyzed nuclei number per sample are shown on the right. C) Correlation of the proportion of each cell type (denoted in %) between the two replicates within one sample. Dots are colored by cell type as indicated in B). Shape of dots represents the replicates. For H40 (square), LV1 (Replicate1) was correlated against LV2 and LVW (Replicate2). Both axes are logarithmic. D) A first iteration of subclustering was done and differential abundance analysis was performed. At first, differences in abundance were tested in apex vs. left ventricular free wall (top), followed by AP and FW jointly vs. septum (bottom). Only a very low number of states were found to be significantly different, which is why the three regions were jointly reported as LV. This figure was part of the publication Reichart et al. (2022).

Next effects of PVs on the cellular composition were tested (Figure 22A, B). Boxplots with individual data points are provided in Figure S3. It should be mentioned that RV samples were not available for all patients. Overall, similar trends in compositional changes were observed across genotype subgroups. For example, in LV, CMs were depleted except for the LMNA subgroup. Endothelial, myeloid, and lymphoid cells were more abundant except for the PKP2 (ACM) group. The abundance of fibroblasts was slightly, but not significantly increased. Mural cell abundance was not significantly altered. In RV, CMs were depleted except for the TTN subgroup. EC abundance was significantly increased except for the PKP2 and PVneg subgroups. The RBM20 and PVneg subgroup showed the highest increase in ECs and myeloids compared to CMs in the LV (EC:CM 8, MY:CM 9.8 and 10.3, 14.6 respectively). Similarly in the LV, PKP2 shows the highest increase in the RV (EC:CM 11.4 and MY:CM 13.3).

6.3.6 Genotypes diversify cardiac fibroblast states

The increase of ECM was quantified by measuring levels of hydroxyproline using the Stegemann method (Figure 23A, B). Fibroblasts were determined to be the major cell type secreting collagens (Figure 23C). Fibroblast abundance however was not significantly increased despite the histopathological increase in fibrosis.

Although overall fibroblast abundance was not increased, more collagens are secreted in heart failure patients to various extents compared to healthy controls (Figure 23D). Collagens furthermore showed genotype-specific upregulation, such as COL4A1 and COL4A2, which were up-regulated in LMNA, TTN, and PKP2. In contrast, PVneg hearts upregulated COL4A5 and COL24A1. Increased collagen secretion might be stimulated by various profibrotic ligands, which are detected on fibroblasts (Figure 23E). Examples are TGFβ receptor (Schafer et al., 2017), weakly AGTR1 (Michel et al., 2016), PDGF receptor A (Gallini et al., 2016), CTGF receptors LRP1, 6 and NTRK1 (Daniels et al., 2009). The expression of profibrotic TGFβ2 was universally increased, supporting a mode of fibroblast auto stimulation (Figure 23F). Myeloids and lymphoids additionally significantly upregulated TGFβ1. LMNA, TTN, and PKP2 increased fibrogenic signaling receptor, and EGFR and PKP2 also increased AGTR1 (Figure 23G), which enables EGFR transactivation (Itabashi et al., 2008; Eckenstaler et al., 2021).

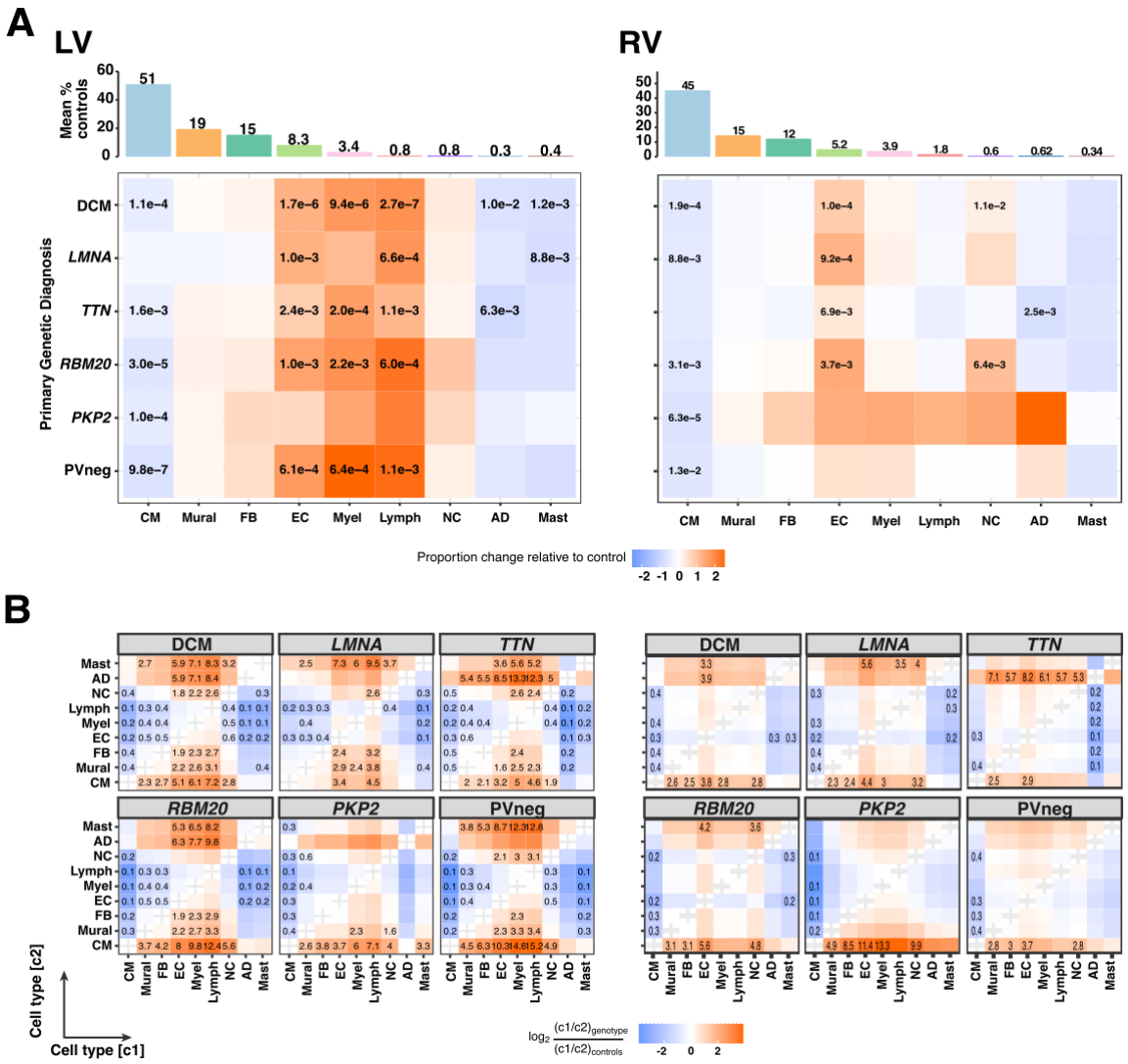

Figure 22: Compositional analysis of control and failing heart samples in LV and RV. A) Upper panel: Mean abundance of identified cardiac cell types in LV and RV of healthy controls. Boxplots with individual data points are provided in Figure S3. Lower panel: Proportional change of cell types in the genotype subgroups and control. The DCM group aggregates all patients with DCM diagnosis. Color scale indicates increase in disease (red) or control (blue). log2-fold changes were computed based on percentages. P values indicated significantly altered proportional changes ($FDR \leq 0.05$) based on CLR-transformed proportions. B) Cell type abundance ratios in the aggregated DCM group and genotype subgroups in LV (left) and RV (right). P values indicated significantly altered proportional changes ($FDR \leq 0.05$) based on CLR-transformed proportions. This figure was part of the publication Reichart et al. (2022). Cell type abundances were provided in T3 and T4 of the online supplement.

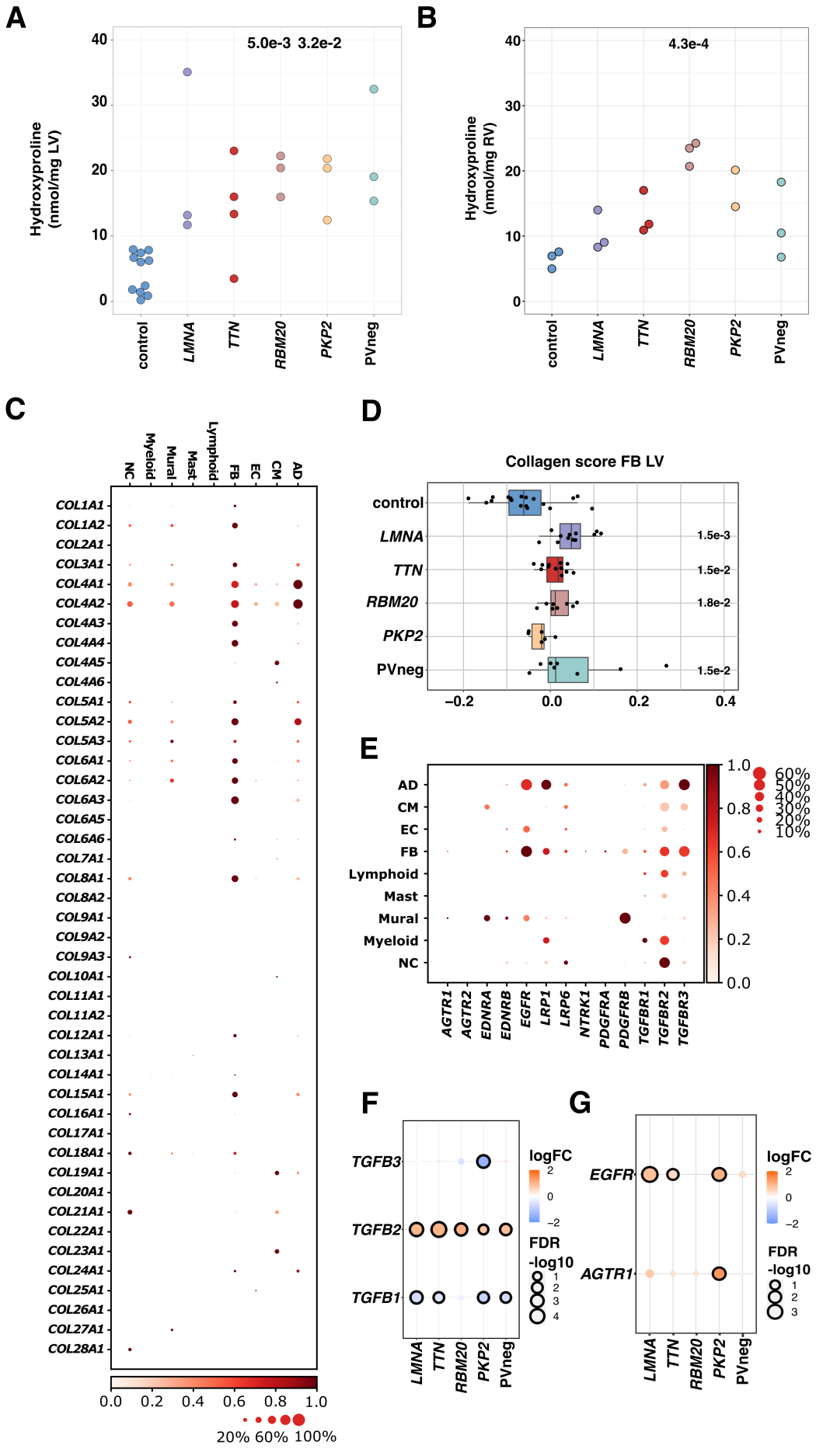
A
Hydroxyproline (nmol/mg LV)
5.0e-3 3.2e-2
control
LMNA
TTN
RBM20
PKP2
PVneg
B
Hydroxyproline (nmol/mg RV)
4.3e-4
C
NC
Myeloid
Mural
Mast
Lymphoid
FB
EC
CM
AD
D
Collagen score FB LV
1.5e-3
1.5e-2
1.8e-2
1.5e-2
E
AD
CM
EC
FB
Lymphoid
Mast
Mural
Myeloid
NC
AGTR1
AGTR2
EDNRA
EDNRB
EGFR
LRP1
LRP6
NTRK1
PDGFRA
PDGFRB
TGFBR1
TGFBR2
TGFBR3
F
TGFB3
TGFB2
TGFB1
logFC
FDR -log10
G
EGFR
AGTR1

Figure 23: Collagen accumulation was observed in heart failure patients. Collagen content of cardiac tissue samples measured using the hydroxyproline assay on A) LV-free wall and B) RV samples. p values of significant hydroxyproline enrichment in genotype subgroups compared to controls are shown above. C) Dotplot shows collagen expression across all cardiac cell types. Collagen expression was enriched in fibroblasts, and to lower extend also observed in mural cells and adipocytes. D) Gene set score enrichment for collagen expression in fibroblasts. E) Dotplot shows expression of pro-fibrotic receptors across cardiac cell types. color, mean expression. Expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states). F) Dotplot shows differential gene expression of TGFβ1-3 across genotypes in LV FBs compared to controls. G) Dotplot shows EGFR and AGTR1 expression across genotypes in LV FBs compared to controls. Size of dot shows fold-change (logFC) and size significance (-log10(FDR)). Significant results are highlighted with framing. Results of differential gene expression were calculated using edgeR. This figure was part of the publication Reichart et al. (2022).

Four previously described ventricular FB states were reidentified during sub-clustering (Figure 24A, B): vFB1.0, vFB2, vFB3 and vFB4. Additionally, two new states, vFB1.1, and vFB1.2, showed distinct transcriptional signatures. This annotation is based on transcriptional similarities, not on assumptions on cellular lineages (Figure 24C, Table T5). vFB1.1 was characterized by increased expression of APOD, APOE, and APOO and has been described in other organs such as the lung as lipogenic fibroblasts (Travaglini et al., 2020). vFB1.2 showed increased expression of multiple cytoskeletal genes such as DAAM1, NAV2, GPC6, suggesting a migratory fibroblast state. GPC6 was furthermore previously reported to be important during chondrocyte differentiation (Melleby et al., 2016). Both states have been validated using single-molecule fluorescent in situ hybridization (Figure 24D, E).

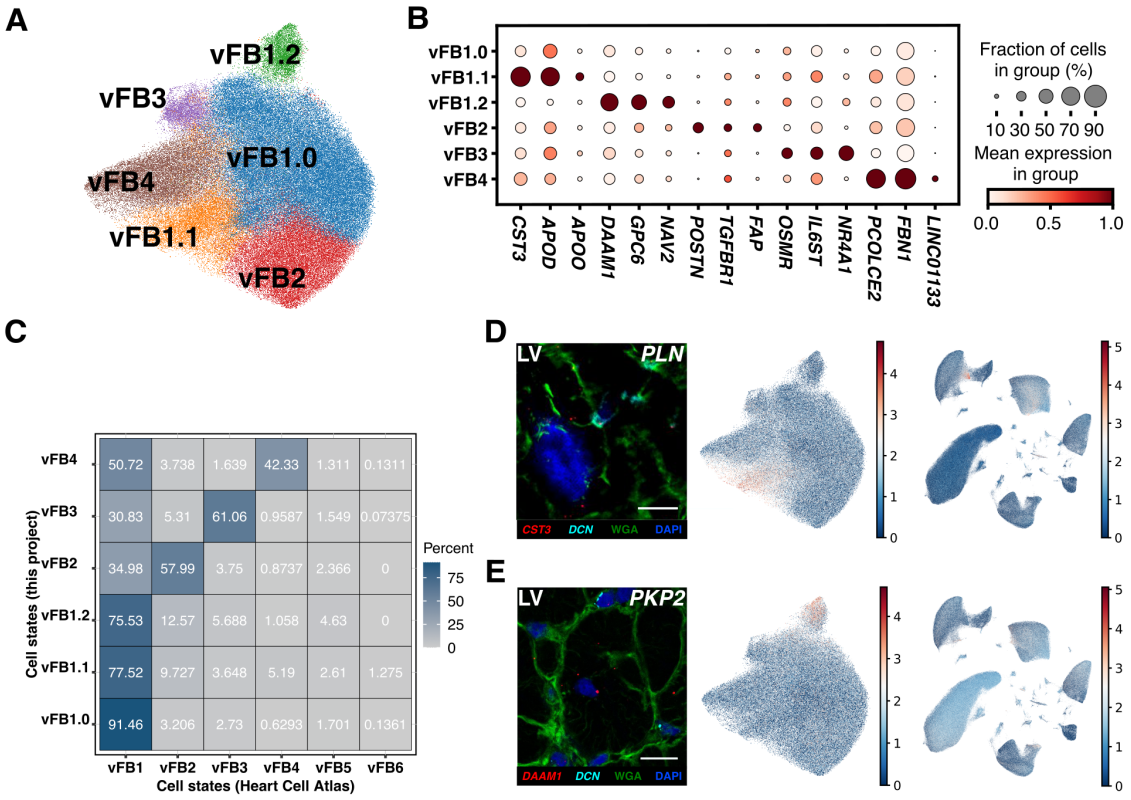

Figure 24: Fibroblast states identified in the healthy human and heart failure patient combined manifold. A) UMAP embedding delineated 6 fibroblast states. B) Dotplot shows selected marker genes of FB states. Dot size represents fraction of expressing cells within a cluster; color, mean expression. Expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states). C) Comparison of annotations per nucleus annotated in the Heart Cell Atlas and the current project. The heatmap shows the fraction of healthy control nuclei in the heart failure dataset (y-axis) and their FB state annotation in the Healthy Human Heart Cell Atlas (x-axis). Single-molecule fluorescent in situ hybridization targeting D) APOD (vFB1.1) and E) DAAM1 (vFB1.2). DCN is used as a FB marker (turqouise), nuclei are DAPI-stained (blue), cell boundaries are delineated using wheat-germ agglutinate (green). Scale bars, 10um. This figure was part of the publication Reichart et al. (2022).

Compositional analysis for fibroblast states was performed (Figure 25 and 26A). FB state abundance, in contrast to FB abundance, was altered. For example, vFB2 was significantly increased in LVs of TTN and PVneg hearts. Other DCM hearts (LMNA and RBM20) showed only modest increase, with PKP2 showing no increase in LV, but a modest increase in RV. Canonical vFB1.0 fraction was decreased across all DCM genotypes in LV. The ratio of

vFB2:vFB1.0 however showed significant dysregulation for LV of TTN, RBM20 and PVneg hearts. The fraction of vFB3 in contrast was modestly decreased in heart failure, with the highest dysregulated vFB2:vFB3 ratio in TTN hearts (LV and RV) only.

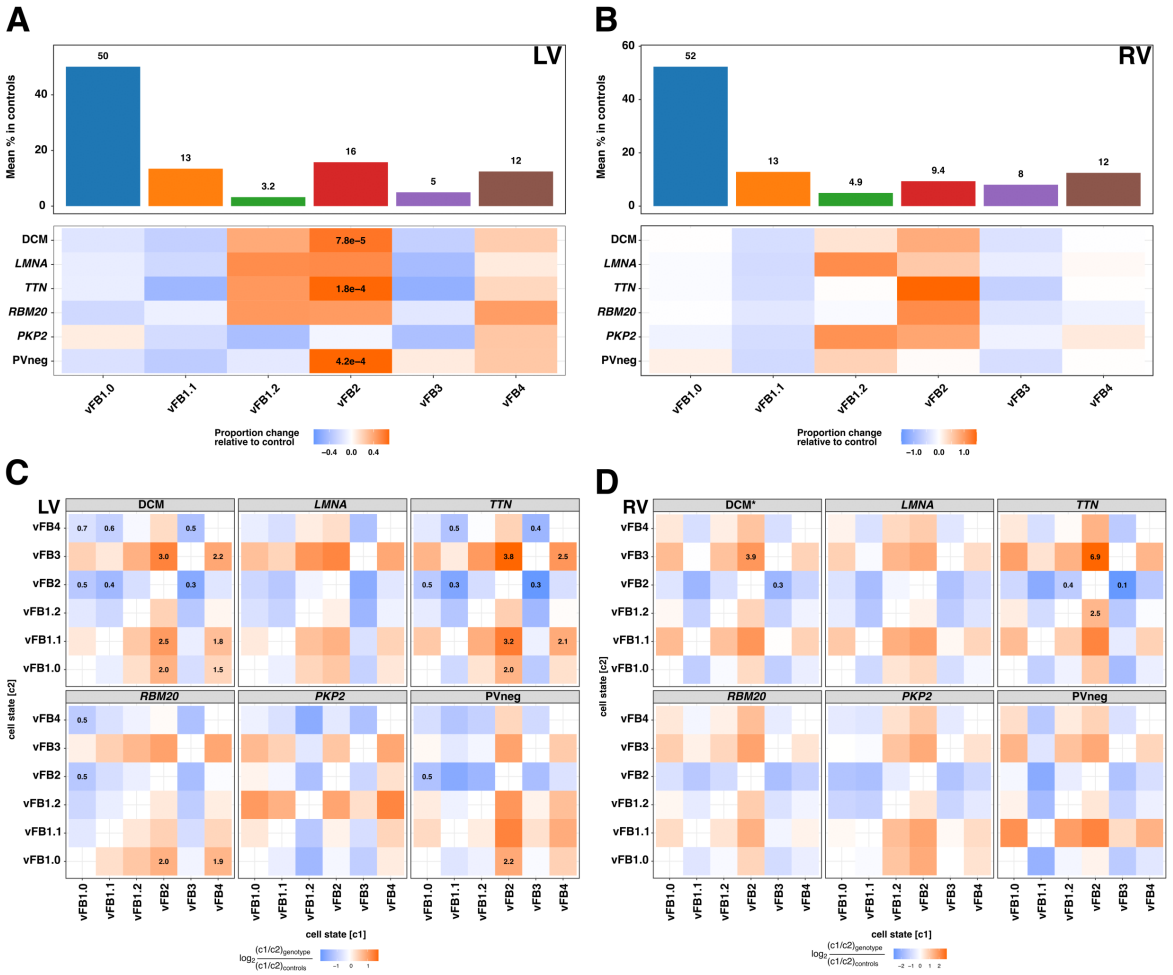

Figure 25: Compositional analysis of fibroblast states. A) Upper panel: Mean proportion of FB states in LV and RV of healthy controls. Boxplots with individual data points are provided in Figure 26A. Lower panel: Proportional change of FB states in the genotype subgroups and control. The DCM group aggregates all patients with DCM diagnosis. Color scale indicates increase in disease (red) or control (blue). log2-fold changes were computed based on percentages. P values indicated significantly altered proportional changes ($FDR \leq 0.05$) based on CLR-transformed proportions. B) Cell type abundance ratios in the aggregated DCM group and genotype subgroups in LV (left) and RV (right). P values indicated significantly altered proportional changes ($FDR \leq 0.05$) based on CLR-transformed proportions. This figure was part of the publication Reichart et al. (2022).

Differential gene expression analysis (DEA) elucidated a high number of differentially regulated genes, especially in canonical vFB1.0, activated vFB2, and ECM-remodelling vFB4 in LV (Figure 26B, C). TTN and PVneg subgroups showed the highest number of uniquely differential expressed genes in LV (Figure 27A, Table T6). Many genes are shared dysregulated in heart failure (Figure 27B, Table T6), existing in parallel to a genotype-specific signature in LV and RV. In LV, DCM subgroups showed an intersecting disease signature, which was absent in RV.

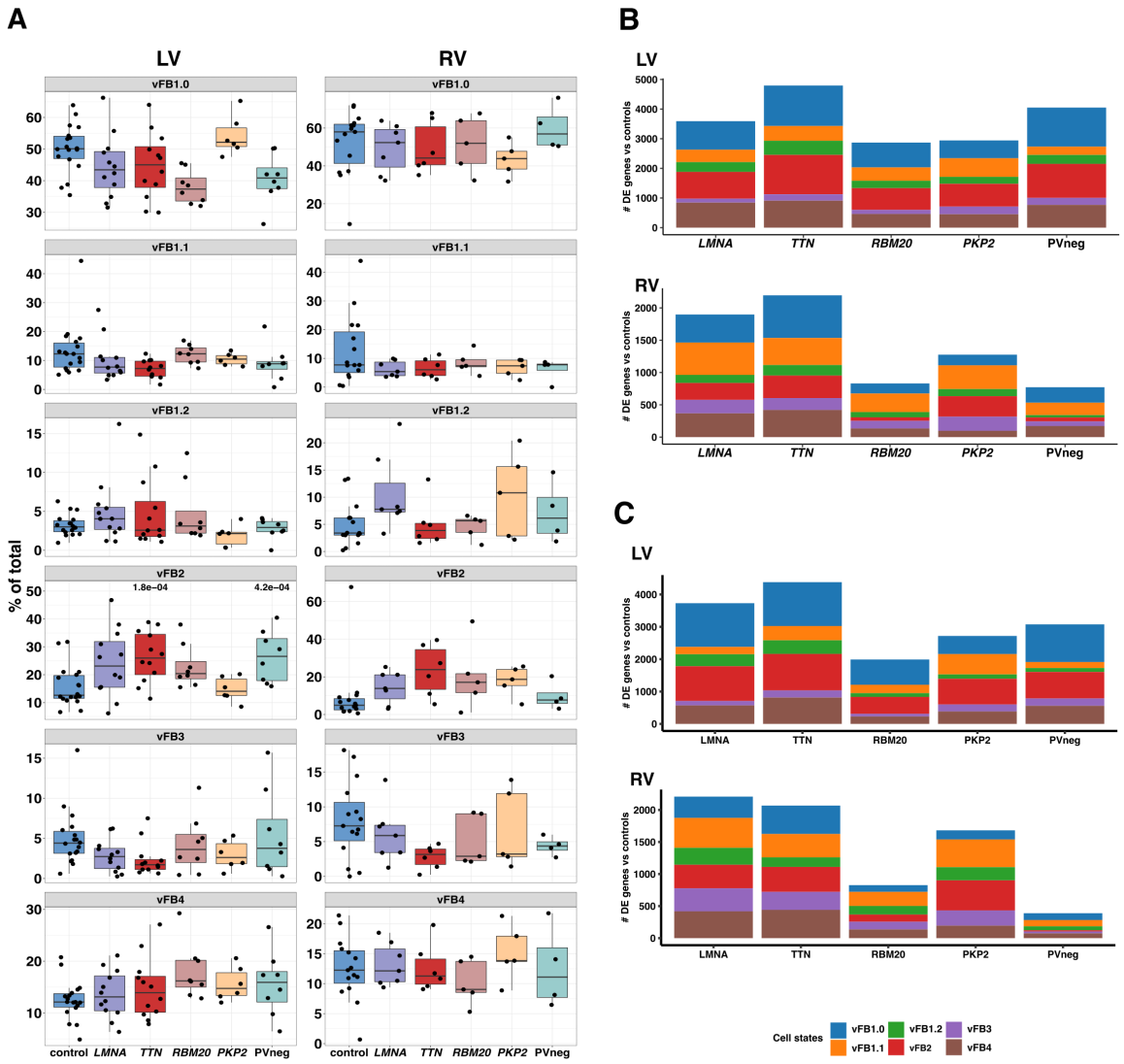

Figure 26: Compositional and differential gene expression analysis of fibroblast states. A) Fibroblast state composition in the aggregated DCM group and genotype subgroups in LV (left) and RV (right). P values indicated significantly altered proportional changes ($FDR \leq 0.05$) based on CLR-transformed proportions. B) Number of differential expressed genes based on the edgeR analysis per genotype subgroup (x-axis) and per cell state (y-axis). Only significantly upregulated genes with $log2FC \geq 0.5$ and $FDR \leq 0.05$ are shown. C) Only significantly downregulated genes with $log2FC \leq -0.5$ and $FDR \leq 0.05$ are shown. This figure was part of the publication Reichart et al. (2022).

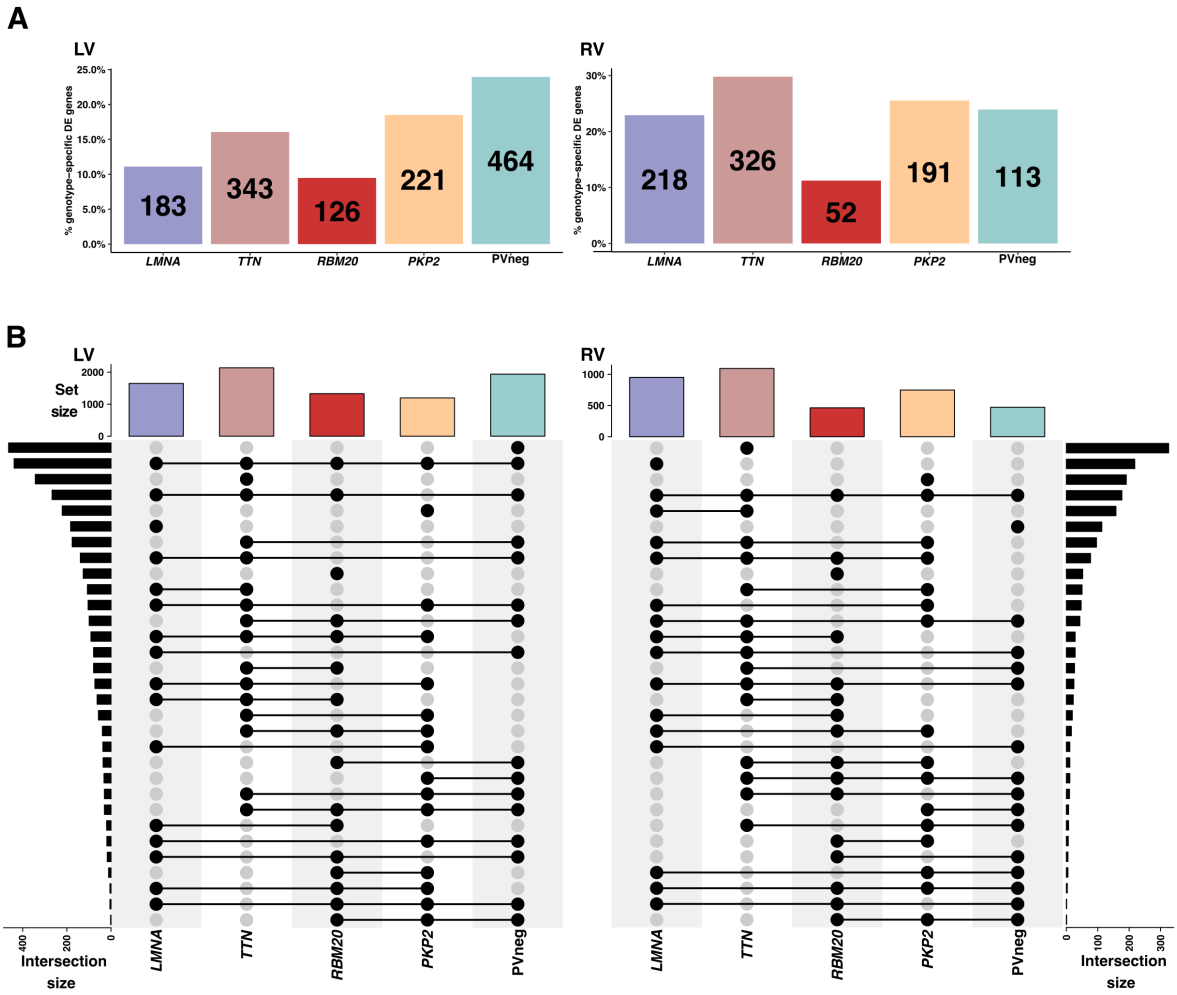

Figure 27: Intersection of upregulated genes across genotype subgroups. A) Barplot showing the fraction of uniquely upregulated genes per genotype subgroup with $log2FC \geq 0.5$ and $FDR \leq 0.05$ across all fibroblast states in LV (left) and RV (right). The absolute number is shown in the bars. B) UpSet plot for all upregulated genes (Figure 26B) depicting the intersection of detected upregulated genes across genotype subgroups. Only unique genes are shown for A) and B), upregulated genes across multiple states are counted as one. This figure was part of the publication Reichart et al. (2022).

Additionally, to study individual dysregulated genes, the enrichment of published gene sets was computed. Below, two examples based on TGFβ and OSM signaling are reported. A list of upregulated genes ($log2FC \geq 0.7$, $FDR \leq 0.05$) from TGFβ cardiac fibroblasts was obtained from a bulk RNA-seq dataset, from which upregulated genes were subsetted and an enrichment score was calculated (Schafer et al., 2017). The score per nucleus was then averaged across all nuclei coming from one patient (Methods). TGFβ-stimulated genes

were enriched in all DCM LVs in vFB2, with the highest enrichment in RBM20 LV (Figure 28A). Surprisingly, PKP2 showed no increase in TGFβ-stimulation despite increased collagen in the LV. IL11 is a lowly expressed cytokine, showing the highest detection levels in RBM20 mutated patients compared to the other patient groups, which was significant using a hypergeometric test (Figure 28B). IL11 was previously shown to be the most upregulated gene upon TGFβ-signalling in vitro fibroblasts and leading to strong profibrotic response, confirmed in this dataset (Schafer et al., 2017) (Figure 28C). This might indicate that IL11 inhibiting-based strategies, which are currently being developed by Boehringer Ingelheim might be most effective in RBM20 mutated patients.

In contrast, OSM-stimulated vFB3 showed decreased expression of OSM signaling genes with the strongest depletion in TTN (Figure 28D). The abundance of CCL2, a cytokine responsible for monocyte and dendritic cell recruitment as shown in other tissues (Yang et al., 2020), was significantly downregulated (Figure 28E, F). Myeloid cells were determined to be the primary source of OSM, which is further discussed in the next section.

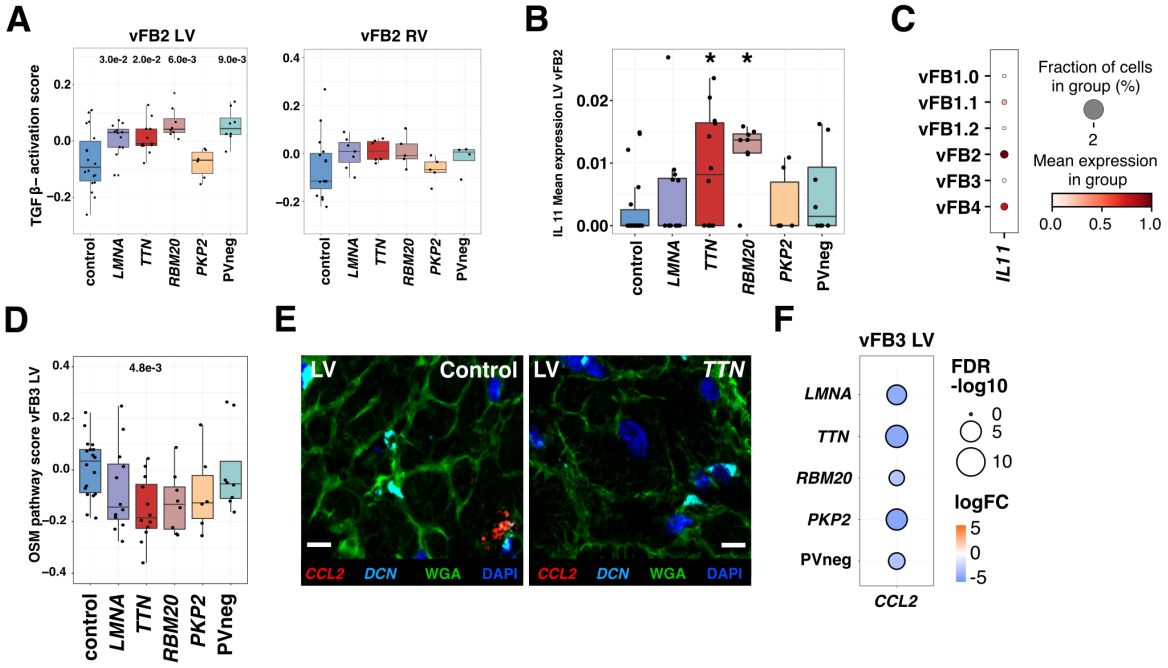

Figure 28: Selected up-and down-regulated genes in vFB2 and vFB3 A) Gene set enrichment of TGFβ-activated genes in vFB2 LV (left) and RV (right). Scores per nucleus were computed per patient. B) Mean expression of IL11 across all vFB2 per patient. *, $p - value \leq 0.05$, hypergeometric test. C) Expression of IL11 across all vFB states. IL11 is enriched in vFB2. D) Gene set enrichment of OSM-pathway genes across vFB3 LV. Scores per nucleus were computed per patient. E) Single-molecule fluorescent in situ hybridization targeting CCL2. DCN is used as a FB marker (turquoise), nuclei are DAPI-stained (blue), and cell boundaries are delineated using wheat-germ agglutinate (green). Scale bars, 10um. F) log-fold change (color) and FDR (dot size) of CCL2 in vFB3 LV as computed by edgeR. The gene shows strong downregulation across all genotypes. This figure was part of the publication Reichart et al. (2022).

6.3.7 Myeloid states of the failing human heart

As part of the immune population of the human heart, we identified a cluster of myeloid cells, which have been identified by the markers F13A1, C1QA, CD68. Myeloid cells expanded during heart failure across all genotypes in LV (Figure 22A). In RV, the increase of myeloids was not significant across genotype subgroups, but was most pronounced in the PKP2 subgroup. The highest increase in relationship to cardiomyocytes in LV was observed in RBM20 and PVneg subgroups, while in RV, PKP2 showed the highest reciprocal relationship (Figure 22B). This increase in myeloid abundance per sample and improved sensitivity in sequencing chemistry (10x V3) elevated marker gene identification, revising the previous myeloid annotation used in the Heart Cell Atlas project.
In total, 14 subclusters were identified, with currently three unclassified states (Figure 29A, Table T7). Myeloid cells were further divided into macrophages, monocytes and classical dendritic cells (cDC). A cluster of proliferating myeloids was identified, with enrichment of cells in G2M and S-phase (Figure 29B).
Non-classical CD16+ monocytes by FCGR3A (CD16), ITGAX (CD11c), and RIPOR2 was the primary monocyte subtype. This cluster was negative for CD14 and showed lower expression of macrophage markers, such as F13A1 or NAV2. VCAN+ monocytes showed an intermediate profile of non-classical monocytes and macrophages, for example higher expression of F13A1 compared to CD16+ monocytes, which is why we assume that VCAN+ monocytes represent a transitional state from tissue-infiltrating monocytes to macrophages. This state showed increased expression of VCAN, LYZ, SORL1, TRERF1, and PLCB1 and weak RIPOR2 expression in contrast to other macrophages. In contrast to monocytes TRERF1, CIITA, and CSF3R were detected.

Three populations of LYVE1 positive populations were identified, with further distinct marker gene expression (Table T8). LYVE1 high and LYVE1 low populations have previously been functionally characterized by Geetika Bajpai starting in 2018 (Bajpai et al., 2018, 2019) and could be subdivided in this study (Figure 29C). In their study, LYVE1 was anti-correlated with CCR2, a marker prominent in recruited macrophage populations. A third tissue resident LYVE1 population, termed LYVE1hi MHCII intermediate showed similarities with the other two LYVE1 populations, such as the high expression of NAV2, SCN9A and DAAM2, but was distinguishable by WASHC2A, EMP1, SLCO3A1. EMP1 was detected in an early microarray experiment of LYVE1+ macrophages with an anti-inflammatory signa-

ture (Pinto et al., 2012).

Dendritic cells were FLT3 positive, a marker also present in PBMC data, such as presented in the online single-cell reference database Azimuth (Hao et al., 2021). Two dendritic cell populations were identified (Table T9). cDC1 were identified by XCR1, BTLA, IRF8, BATF3, and CADM1. cDC2 in contrast showed enriched expression for SIRPA, CLEC10A, ITGAM, SIRPA and ZEB2 (Brown et al., 2019). I was not able to furthermore split cDC2A and cDC2B based on previously reported markers. cDCs showed high expression of HLA genes, CIITA and CD74 (Figure 29C, D), indicating the important role in antigen presentation of cDCs (Murphy and Weaver, 2018).

Two NAMPT, ACSL1 and CEMIP2 enriched populations of recruited macrophages have been identified with further subdivision. Although not explicity annotated, MP OSM were functionally characterized (Abe et al., 2019). MP NFkB showed enrichment of NfKB genes such as NFKB1, NFKBIA, and CCL3 ($MIP-1\alpha$), a chemokine binding to CCR1, CCR3 and CCR5 and functionally described to be involved in wound healing and recruitment of lymphocytes and macrophages (Bhavsar et al., 2015).

The top five marker genes for all myeloid populations are provided in Figure 29F.

TREM2+ macrophages, which were characterized in multiple tissues (Jaitin et al. (2019) termed them LAMs) and species (Rizzo et al., 2020), were described to arise from circulating monocytes, respond to extracellular lipids and are associated with loss of metabolic homeostasis. TREM2 detection however was lower in snRNAseq data compared to scRNAseq.

FOLR2+ macrophages were identified by their FOLR2, C1QA, C1QB, C1QC, and FTL expression, a population characterised as anti-inflammatory and often found in tumors (Puig-Kröger et al., 2009). One macrophage population showed elevated expression of interferon-stimulated genes. As a similar interferon-stimulated gene (ISG) profile was found in mice, this population was named accordingly (Dick et al., 2019). Their transcriptional signature overlaps with previously reported tissue-resident CCR2+ macrophages (Bajpai et al., 2019).

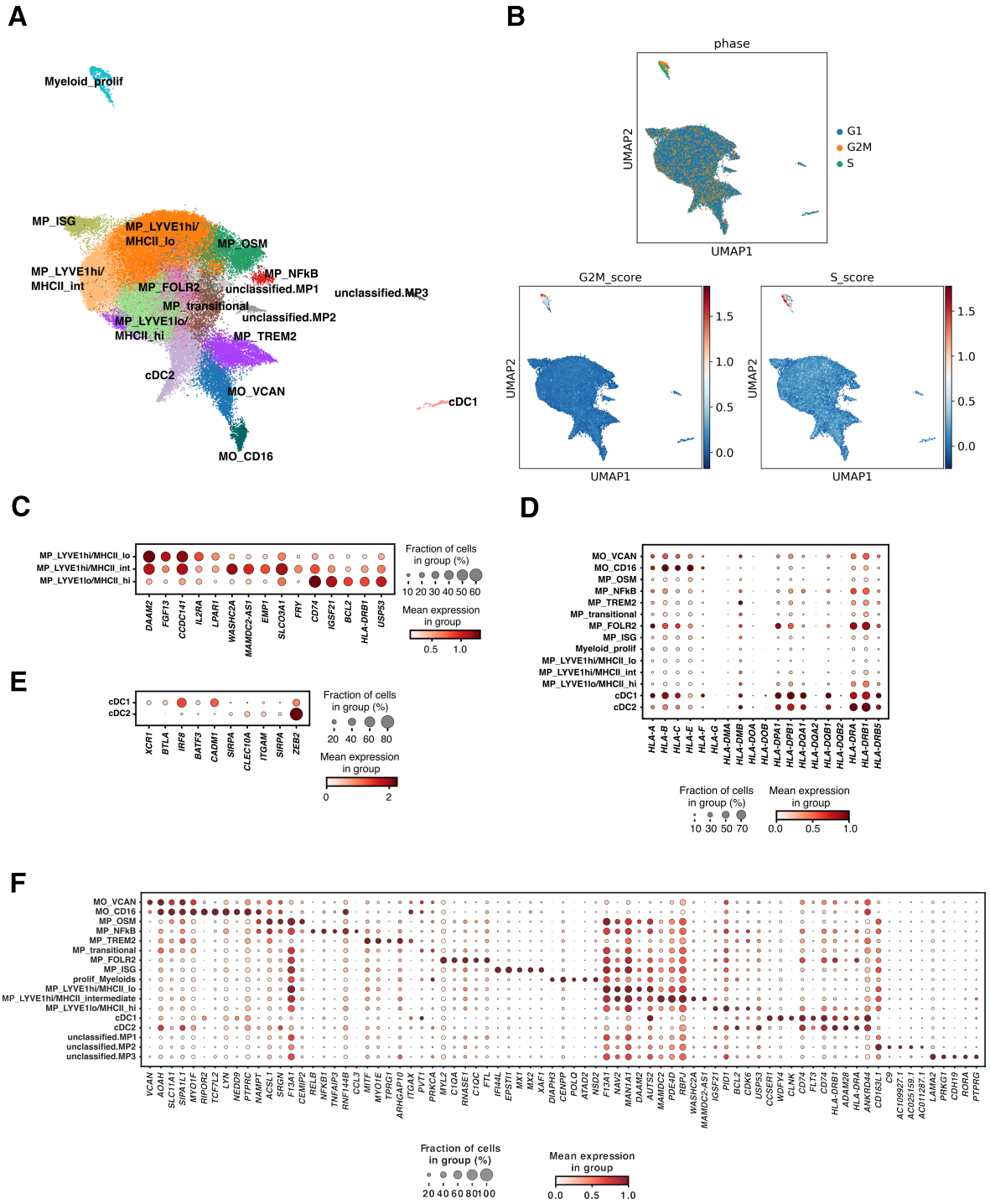
A
Myeloid_prolif
MP_ISG
MP_LYVE1hi/
MHCII_lo
MP_OSM
MP_LYVE1hi/
MHCII_int
MP_NFkB
MP_FOLR2
unclassified.MP1
unclassified.MP3
MP_transitional
MP_LYVE1lo/
MHCII_hi
unclassified.MP2
MP_TREM2
cDC2
MO_VCAN
cDC1
MO_CD16
B
phase
UMAP2
UMAP1
G1
G2M
S
G2M_score
S_score
1.5
1.0
0.5
0.0
C
MP_LYVE1hi/MHCII_lo
MP_LYVE1hi/MHCII_int
MP_LYVE1lo/MHCII_hi
DAAM2
FGF13
CCDC141
IL2RA
LPAR1
WASHC2A
MAMDC2-AS1
EMP1
SLCO3A1
FRY
CD74
IGSF21
BCL2
HLA-DRB1
USP53
Fraction of cells
in group (%)
10 20 30 40 50 60
Mean expression
in group
0.5 1.0
D
MO_VCAN
MO_CD16
MP_OSM
MP_NFkB
MP_TREM2
MP_transitional
MP_FOLR2
MP_ISG
Myeloid_prolif
MP_LYVE1hi/MHCII_lo
MP_LYVE1hi/MHCII_int
MP_LYVE1lo/MHCII_hi
cDC1
cDC2
HLA-A
HLA-B
HLA-C
HLA-E
HLA-F
HLA-G
HLA-DMA
HLA-DMB
HLA-DOA
HLA-DOB
HLA-DPA1
HLA-DPB1
HLA-DQA1
HLA-DQA2
HLA-DQB1
HLA-DQB2
HLA-DRA
HLA-DRB1
HLA-DRB5
Fraction of cells
in group (%)
10 30 50 70
Mean expression
in group
0.0 0.5 1.0
E
cDC1
cDC2
XCR1
BTLA
IRF8
BATF3
CADM1
SIRPA
CLEC10A
ITGAM
SIRPA
ZEB2
Fraction of cells
in group (%)
20 40 60 80
Mean expression
in group
0 2
F
MO_VCAN
MO_CD16
MP_OSM
MP_NFkB
MP_TREM2
MP_transitional
MP_FOLR2
MP_ISG
prolif_Myeloids
MP_LYVE1hi/MHCII_lo
MP_LYVE1hi/MHCII_intermediate
MP_LYVE1lo/MHCII_hi
cDC1
cDC2
unclassified.MP1
unclassified.MP2
unclassified.MP3
VCAN
ACOAH
SLC11A1
SIPA1L1
MYO1E
RIPOR2
TCF7L2
LYN
NEDD9
PTPRC
NAMPT
ACSL1
SRGN
F13A1
CEMIP2
RELB
NFKB1
TNFAIP3
RNF144B
CCL3
MITF
MYO1E
TPRG1
ARHGAP10
ITGAX
PVT1
PRKCA
MYL2
C1QA
RNASET2
C1QC
CTTL1
IFI44L
EPSTI1
MX1
IMX2
XAF1
DIAPH3
CENPP
POLQ
ATAD2
NSD2
F13A1
NAV2
MAN1A1
DAAM1
AUTS2
MAMDC2
PDE4D
RBPJ
WASHC2A
MAMDC2-AS1
IGSF21
PID1
BCL2
CDK6
USP53
CCSER1
WDFY4
CLNK
CD74
FLT3
CD74
HLA-DRB1
ADAM28
HLA-DRA
ANKRD44
CD163L1
C9
AC109927.1
AC025159.1
AC011287.1
LAMA5
PRKG1
RORA
CDH19
PTPRG
Fraction of cells
in group (%)
20 40 60 80 100
Mean expression
in group
0.0 0.5 1.0

Figure 29: Myeloid states identified in failing human hearts. A) uMAP embedding delineated 14 states, including three unclassified states. B) Cell-cycle classification across myeloid states. Highest scores are assigned to a cluster, which was subsequently annotated as proliferating myeloids. G2M- and S-scores were enriched in this state. C) Dotplot shows selected marker genes for tissue-resident LYVE1 states. D) Dotplot shows MHCI and II gene expression for all myeloid states. E) Dotplot shows selected marker gene expression for cDC1 and cDC2. F) Dotplot shows selected marker gene expression for all myeloid states. Dot size represents fraction of expressing cells within a cluster; color, mean expression. For D) and F), expression was scaled from 0 (minimal expression across all states) to 1 (maximum expression across all states). This figure was part of the publication Reichart et al. (2022).

The most striking proportional shift was observed for proliferating myeloid cells, which were nearly not detectable across all genotype subgroups (Figure 30, 31). TREM2 macrophages were increased in LMNA and PVneg LV, and cDC2 were significantly enriched in RBM20. MP ISG was increased in PKP2 LVs. The myeloid and lymphoid population showed the highest number of transcriptionally distinct states compared to all other cell types (Reichart et al., 2022).

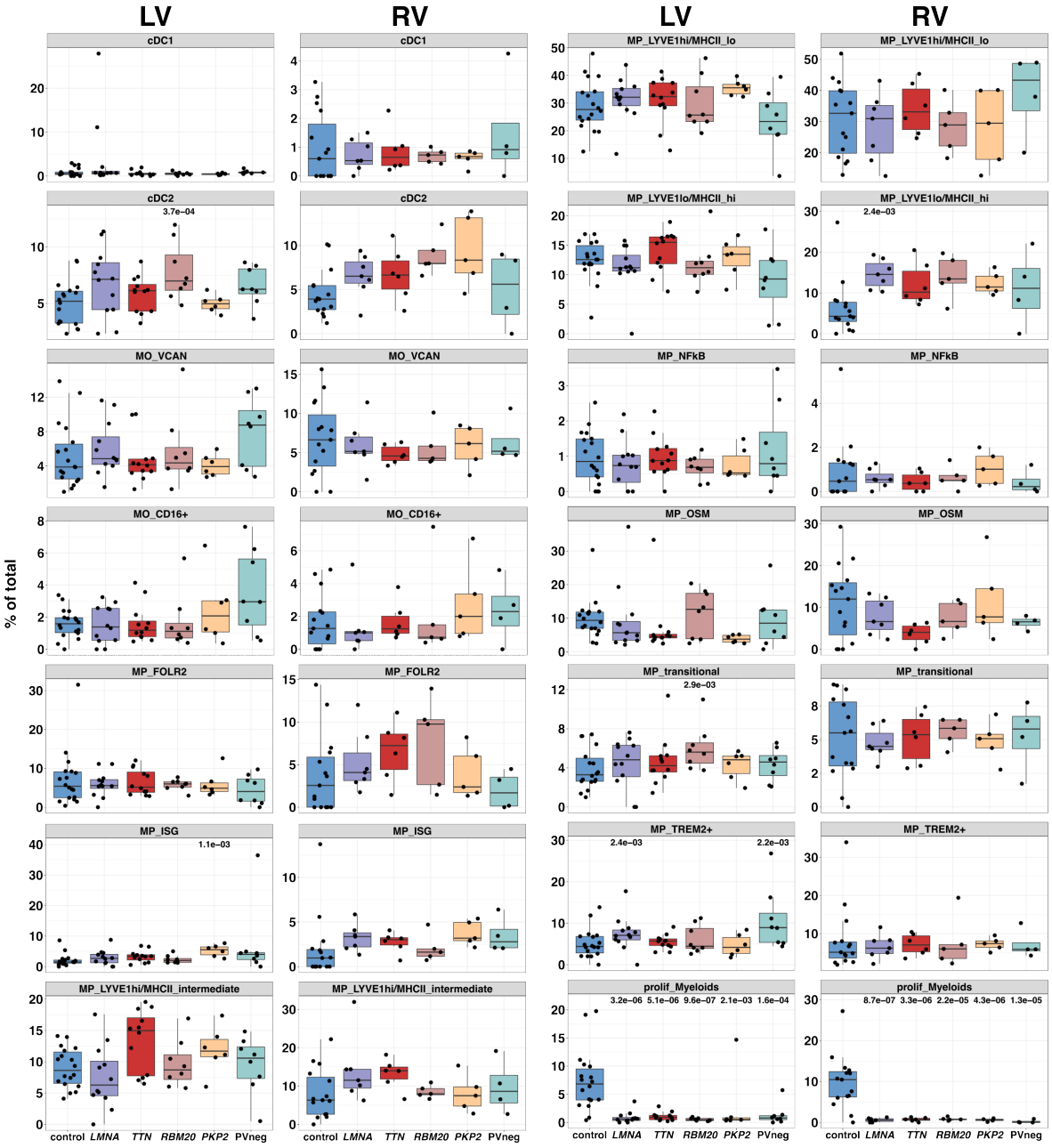

Figure 30: Proportions of myeloid states. Myeloid state composition in the genotype subgroups in LV (left) and RV (right). P values indicated significantly altered proportional changes ($FDR \leq 0.05b$). This figure was part of the publication Reichart et al. (2022).

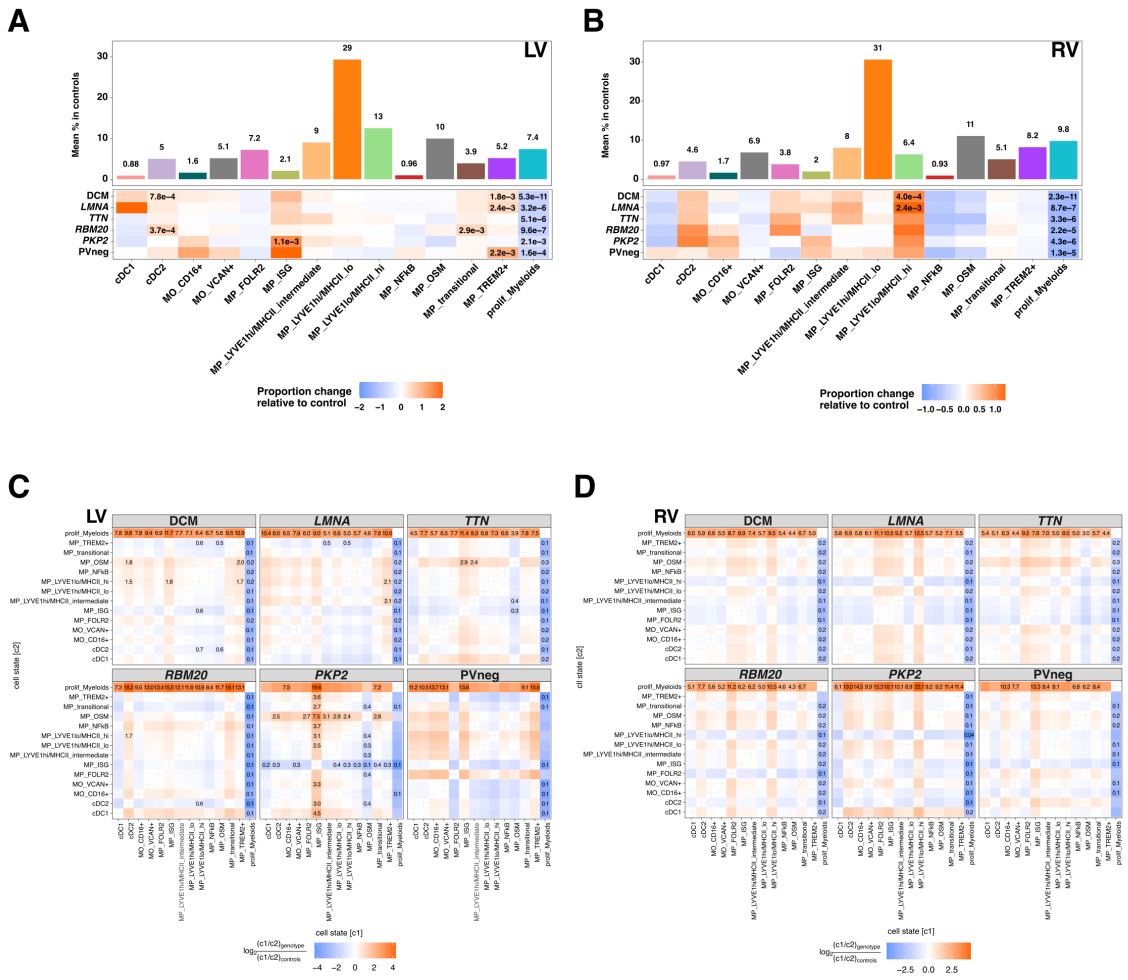

Figure 31: Compositional analysis of myeloid states. A), B) Upper panel: Mean proportion of myeloid states in LV and RV of healthy controls. Boxplots with individual data points are provided in Figure 30. Lower panel: Proportional change of myeloid states in the genotype subgroups and control. The DCM group aggregates all patients with DCM diagnosis. The color scale indicates an increase in disease (red) or control (blue). log2-fold changes were computed based on percentages. P values indicated significantly altered proportional changes ($FDR \leq 0.05$) based on CLR-transformed proportions. C), D) Cell type abundance ratios in the aggregated DCM group and genotype subgroups in LV (left) and RV (right). P values indicated significantly altered proportional changes ($FDR \leq 0.05$) based on CLR-transformed proportions. This figure was part of the publication Reichart et al. (2022).

One focus during the analysis of the myeloid population was MP OSM, the myeloid population stimulating vFB3, which showed a significant dysregulation in Titinopathies. MP OSM

was slightly lower abundant in TTN compared to other DCM subgroups, with one outlier patient having more than 30%. Oncostatin M secretion of this population was downregulated in LMNA and TTN, suggesting a disturbed macrophage-fibroblast interaction especially pronounced in TTN.

Another studied aspect within the myeloid population was antigen presentation. cDC2 was significantly more abundant in RBM20 cases compared to other DCM subgroups, antigen presentation was assessed by calculating MHCII gene enrichment across the major HLA-presenting myeloid states (Figure 29D). MHC II was significantly enriched in RBM20 LV in contrast to laminopathies (Table T10). Although PVnegs also showed enriched MHCII gene expression, two patients were within the third quartile of the control group's MHCII expression distribution (Figure 32).

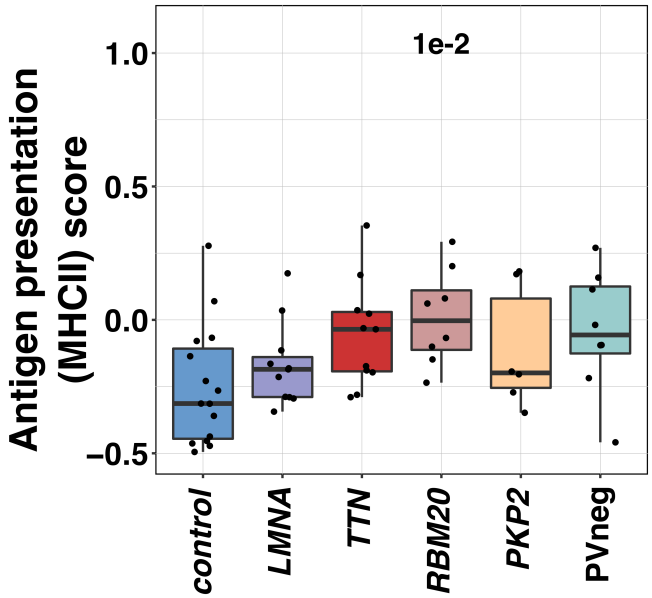

Figure 32: MHC II gene expression in myeloids. Mean enrichment score of antigen presenting MHC II genes in LVs across antigen presenting myeloid states (cDC1, cDC2, MO CD16, MO VCAN, MP FOLR2 and MP LYVE1 lo-MHC II hi) per patient. This figure was part of the publication Reichart et al. (2022).

6.3.8 Recognition of genotype-specific expression signatures using machine learning

We furthermore investigated whether transcriptional signatures are strong enough to construct an unbiased high-accuracy mathematical model assigning correct genotypes to each patient, independent of the previous identification of differentially expressed genes. Machine learnings models allow the identification of discriminating signatures and quantify similarities between predicted genotype subgroups (classes).

I determined multiple prerequisites for good model performance, which are independent of the finally selected model architecture. At first, different cardiac cell types and states need to be modeled differently due to different cellular responses, in line with differences in DEGs. Secondly, cardiac regions need to be modeled differently. Previously described histological features which discriminate LV and RV are reflected in the transcriptional signature. Pooling LV and RV decreased model performance. Thirdly, the training needs to follow a per-patient cross-validation training policy. A training policy that assumes that all nuclei are independent observations and randomly splits the dataset to training and test data risks overfitting to patient-specific transcriptional signatures. This tendency to overfitting was observed for some classification pipelines, such as scanVI. The constructed latent space was independent of the pathogenic variant, but rather overfit to a patient signature, observed after genotype label shuffling per patient and repeating model training (Figure S4). The model should also preferably work on all expressed genes per cell type and not a small subset of pre-selected genes.

At first, machine-learning models which allow the identification of linear relationships between features (=genes) were selected, such as logistic regression (Figure 33). A coefficient is computed for each gene, representing the genes importance for the classification.

The control group was classified from the genotype subgroups with high accuracy. The AUC to differentiate genotype subgroups is between 0.62 and 0.82 (PKP2). From the ROC curves it becomes apparent, that high true positive genotype classifications per nucleus can only be achieved by high false positive rates, complicating the interpretation of feature importance. Model complexity was gradually increased to improve prediction, with graph-attention models showing the best performance. The application of graph attention models on single-cell data were firstly described in 2020 by David van Dijks lab (Ravindra et al., 2020), with which high accuracy of 92% for a multiple sclerosis dataset with two classes was reported. Together with collaborators, a graph attention model was developed with the above-described

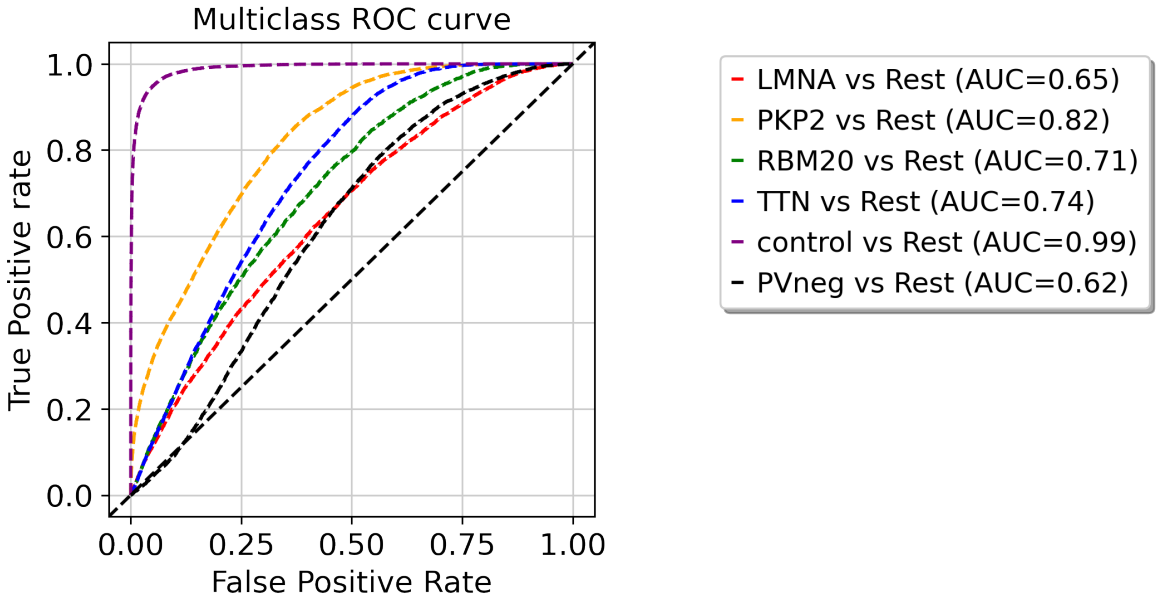

Figure 33: Multiclass genotype subgroup classification using logistic regression. ROC curves showing a function of true and false prediction per nucleus per genotype subgroup (color). The AUC per genotype is shown in brackets in the legend. ROC: Receiver operating characteristic, AUC: Area under the curve.

prerequisites and suitable for multiclass classification.

Cell type-specific neighborhood graphs showed more connectivities within genotypes than across different subgroups, suggesting that nuclei from patients harboring a pathogenic variant within the same gene have a lower euclidean distance between each other compared to nuclei from patients with a PV in a different gene. As for LVAD-IP patients no RV samples were available, and only LV cell types were considered to calculate the aggregated probability (Figure 34 A-C). Due to the large number of nuclei needed for model training, only highly abundant cell types were considered. All cell types were tested for their predictive potential, however, lower abundant populations (e.g. adipocytes, neuronal and lymphoid cells in RV) returned errors and no valuation metric could be computed. Overall, the GAT model shows good performance (accuracy 0.87 and F1 macro 0.91), independently confirming that genotype subgroups carry unique transcriptional signatures.

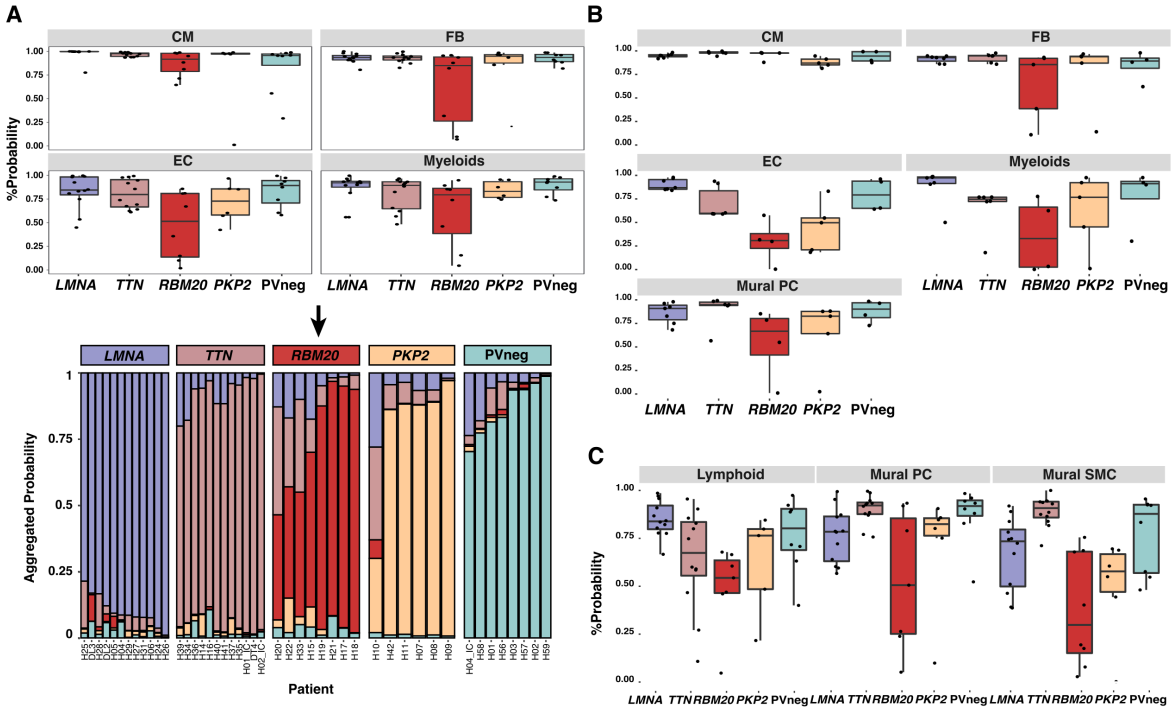

Figure 34: Probability of true genotype probability per cell type per patient returned by GAT. A) Top: Relative number of nuclei which have been assigned with the correct genotype subgroup label per patient. Only LV of the shown cell types was considered. Bottom: Cell type probabilities were then aggregated to calculate a genotype probability per patient. B) Relative number of nuclei which have been assigned with the correct genotype subgroup label per patient's RV. C) Relative number of nuclei that have been assigned with the correct genotype subgroup label per patients LV of lower abundant cell types. This figure was part of the publication Reichart et al. (2022).

6.4 Compatibility of established dataset with future projects

6.4.1 10x 3' v3 to v3.1 differences

10x Genomics announced protocol changes for their 3' Reagent kit in November 2019, leading to the release of the v3.1 chemistry. To allow comparisons with newly collected genotype subgroups, compatibility with the 10x V3 and V3.1 chemistry were compared. Batch effects overlaying disease effects from a newly added genotype subgroup potentially lead to the identification of false positive findings in the differential gene expression analysis. To quantify these batch effects, tissue samples from the same region of already included patients were reprocessed using the new v3.1 chemistry. Libraries have been downsampled to the same

sequencing depth. No significant changes in the number of detected UMIs and genes were observed and no differentially detected genes were identified, suggesting that no batch effects are included when integrating new samples into this dataset (Figure 35). This further supports, that the proposed dataset will be a valuable source for additional future studies related to dilated and arrhyhthmogenic cardiomyopathies.

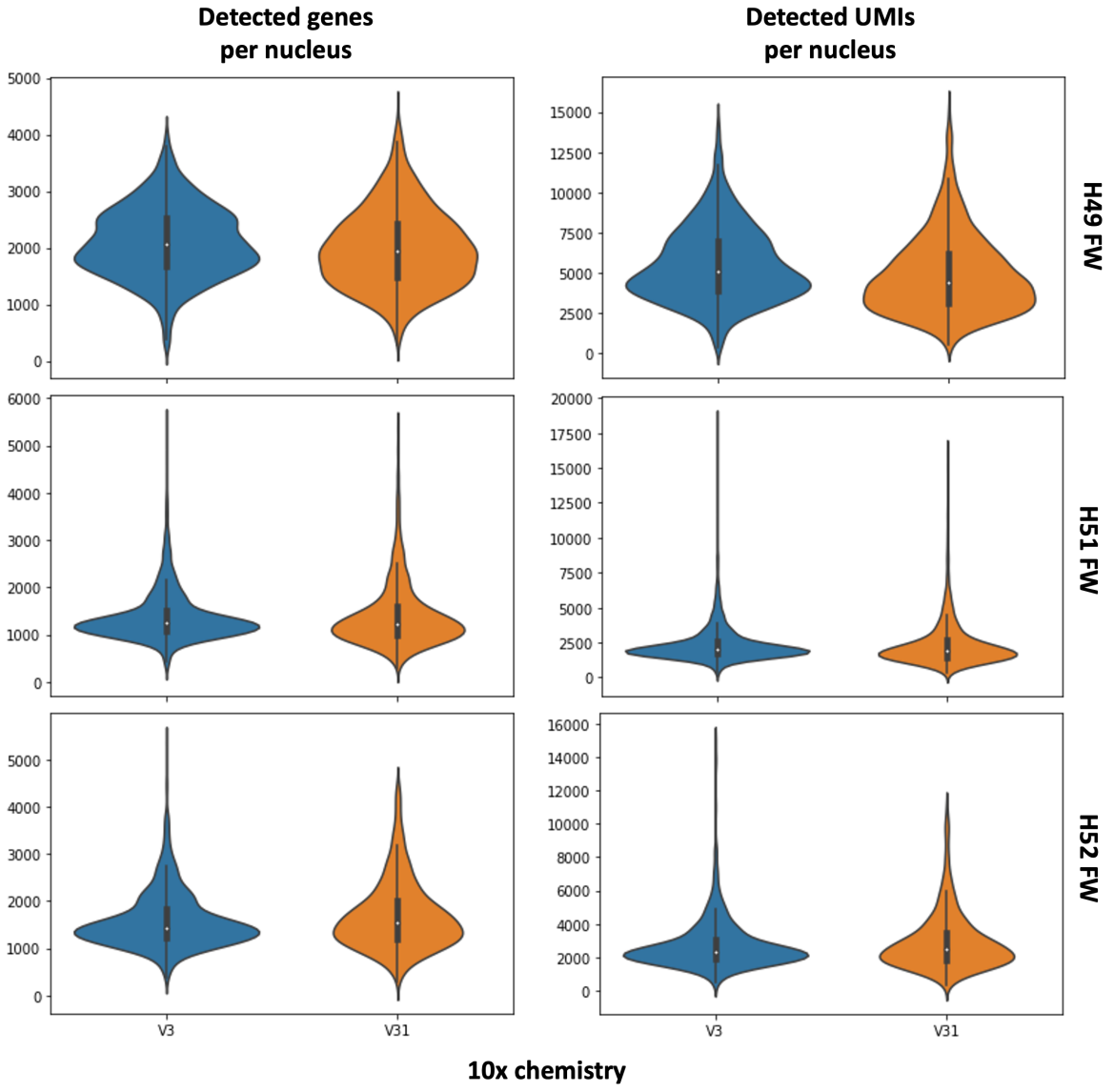

Figure 35: Differences in Gene and UMI detection of the V3 and V3.1 chemistry. No significant changes have been observed. Per patient variability exceeds differences of the 10x chemistry.

7 Discussion

In my doctoral research project, I worked on understanding the cellular composition in the healthy adult human heart and how the architecture of cardiac cell types changes during heart failure, dilated and arrhythmogenic cardiomyopathy. To address this question, my project was divided in three subprojects. At first, a protocol for single-nucleus RNA sequencing was established and optimized in the lab to sequence all cell types in the human heart: from very small, like endothelial cells, to large rod-shaped cardiomyocytes.
In the next phase, in collaboration with the Healthy Heart Cell Atlas Consortium as part of the Human Cell Atlas (HCA), I annotated fibroblast subtypes and characterized regional similarities (for example between left ventricular free wall and left ventricular apex) and heterogeneity, such as atrial vs. ventricular differences, within the 6 studied cardiac regions.
Knowledge of cell type and state annotation plus the sequenced healthy human heart samples were then used to annotate cell types and states in the failing human heart. Here I studied how the cell type and state compositional changes in heart failure, characterized newly identified cell states and focus on transcriptional differences in the fibroblast and myeloid compartment based on a revised myeloid annotation. This large-scale analysis revealed shared and distinct mechanisms between genotype-stratified patient subgroups, supporting the observation that heart failure is not a common final pathway. This atlas provides novel insights into improving personalized treatment strategies for genetic cardiomyopathies.

7.1 Protocol optimization for isolation of intact nuclei

Large amounts of debris are found by microscopy in the tissue homogenate after Dounce homogenization (Figure 8), impacting library quality. The nuclei isolation protocol however needs to be mild enough to prevent nuclei blebbing, a process describing leaky nuclear membrane. Three purification methods were compared: Purification by FACS sorting, filtration, and gradient centrifugation. Methods were compared for applicability for cardiac tissue, throughput and nuclei quality. FACS sorting, although being the method with the lowest throughput, is the most flexible method, as it allows purification of nuclei from small tissue pieces (e.g. apical cores), and showed the highest quality of recovered nuclei, which was evaluated using barcode-rank plots (Figure 7)). FACS gating strategies were optimized to have an unbiased capturing of nuclei from cardiomyocyte and non-myocytes. This was observed for cardiomyocytes, possibly correlating with features of polyploidy. Although multinucle-

ation nor polyploidy determine the state of cardiomyocytes (Yekelchyk et al., 2019) and the results of the differential gene expression analysis remain unchanged, biased sampling alters the results of compositional analysis. Another advantage of FACS sorting is the possibility of enriching and depleting antibody-labeled cell types.

Protocol applicability was furthermore evaluated for failing human hearts. It was previously shown that that some pathogenic variants affect nuclear integrity, such as laminopathies (Paradisi et al., 2005). Levels of ambient RNA were compared between genotype subgroups to evaluate increased levels of ambient RNA (Figure 36).

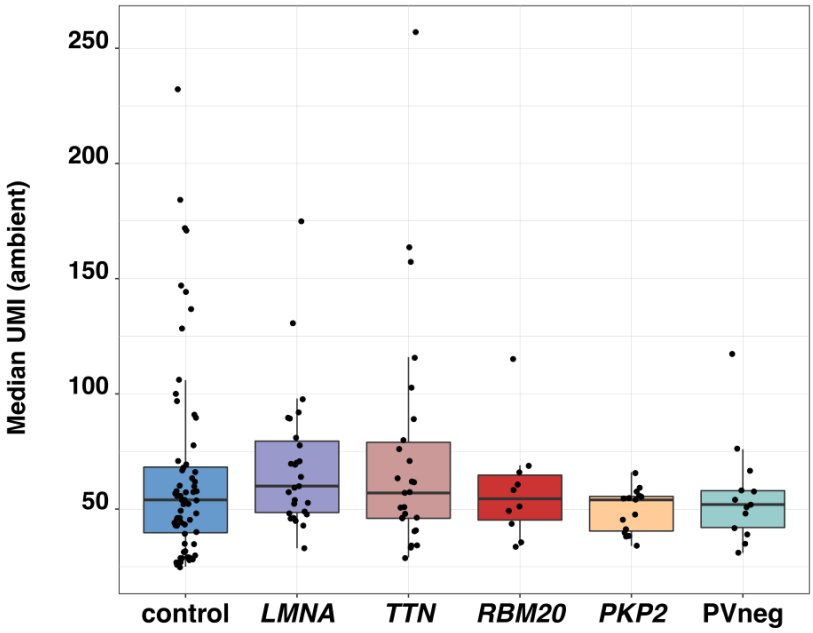

Figure 36: Ambient RNA per genotype subgroup. Median ambient RNA levels were calculated per sample per genotype subgroup. No significant difference between genotype subgroups was identified.

No significantly increased levels of ambient RNA were observed, suggesting that this protocol does not introduce a batch effect linked to the patient's pathogenic variant. cDNA recovery did also not show a significant correlation with storage time, supporting the applicability of the technique to biobanked tissue.

7.2 The fibroblast population of the healthy adult human heart revealed by single-cell sequencing

The second phase of the project consisted of the fibroblast cell state annotation of tissue samples from healthy human hearts, which were collected by the Heart Cell Atlas Consortium

at two different locations (UK and North America). This comprehensive map of snRNAseq data and non-myocyte scRNAseq data serves as a basis for further disease studies. snRNAseq showed better capturing of fibroblasts compared to the applied scRNAseq protocol. Although scRNAseq shows higher recovery of UMIs per gene, due to higher RNA content in the cytosol, both protocols are suitable to call the same cell states in the fibroblast compartment (Figure 14). The size of this atlas of nearly half a million data points is twice as big as the second largest heart cell atlas published so far (Tucker et al., 2020), with more homogenous sample quality (see Figure S2A, B and D of this paper). Given the number of recovered fibroblasts in the Heart Cell Atlas, interrogation of rare cell states is feasible with higher statistical power. The generated fibroblast map is deposited here: www.heartcellatlas.org.

Canonical fibroblasts in the atria and ventricles were identified (FB1, 2), one subtype being stimulated by Oncostatin M (FB3). snRNAseq data show that OSM is presumably secreted by a specific recruited macrophage population, which was subsequently termed MP OSM+. Two fibroblast populations were involved in connective tissue formation: ECM-producing, TGFβ stimulated FB4, and ECM-organizing FB5, defined by the increased expression of proteins involved in ECM turnover and organisation of collagen fibers. This showed that different signalling pathways act on transcriptionally distinct fibroblast states, which jointly form connective tissue. FB3 in contrast showed lowest expression of ECM proteins and was instead defined by increased secretion of cytokines, such as CCL2, suggesting that this state was rather involved in maintaining tissue homeostasis by communicating with other cell types (Daseke II et al., 2020). Oncostatin-stimulated FB3 was most abundant in the atria, especially the right atrium, while TGFβ activated FB4 was more abundant in the ventricles (especially LV). Two states FB6 and 7 represented profiles of chimeric states with endothelial cells and cardiomyocytes. No overlap of CM and fibroblast marker genes in the same cell was identified by RNA in situ hybridization, indicating that those states represent multiplets. This showed that despite rigorous QC filtering and doublet probability calculation cells pass the filtering criteria. This was especially true for large-scale single-cell atlas projects, for which a low number of nuclei passing QC filters per sample can accumulate and form in the combined dataset own clusters. Separate clustering of atrial and ventricular FBs recapitulated the populations described above, such as an OSM-stimulated population in each chamber (aFB4 and vFB3) (Figure 15A-D), and distinct chamber-specific extracellular matrix (ECM) producing and organizing FBs were identified (aFB2 versus vFB2) (aFB1 versus vFB4) (Figure 15E). For example, collagen genes such as COL15A1, COL6A6, and COL4A4

were higher expressed in the atria, while COL8A1, COL6A3, and COL1A1 showed elevated expression in the ventricles (Figure 15E). COL15A1 was shown facilitate the adherence of basement membranes to the connective tissue stroma (Bretaud et al., 2020). A study by Skribic et al. have localized COL8A1 around fibroblast and the cardiac endothelium (Skrbic et al., 2015). They furthermore described inhibitory effect of COL8A1 on RhoA signalling, which is involved in FB migration. Mutations in COL6A3 are associated with muscular dystrophy, resulting in instability of the extracellular matrix (Marakhonov et al., 2018). The function of COL6A6 and COL4A4 in the heart is unknown. I assume that differences in ECM composition are adaptations to differences in pressure and mechanical stress in the atria and ventricles. Similarly, ECM modulators showed chamber-specific expression patterns. Further studies on the protein level are needed to define changes in the extracellular matrix, as ECM turnover is additionally regulated by matrix metalloproteases (MMPs) and tissue inhibitor of metalloproteinases (TIMPs) (Vanhoutte and Heymans, 2010).

The fibroblast atlas was used to study cell state-specific expression of uncharacterized genes and to study regional differences in extracellular matrix production. LINC01013, an uncharacterized lncRNA encoding a small microprotein, was highest expressed in TGFβ stimulated FB4 and co-expressed with other profibrotic markers such as TGFβR1, FAP, POSTN, or COL1A1. This was supported by in vitro stimulation of fibroblasts with TGFβ, done at the Barton lab (Quaife et al., 2022). The co-expression of LINC01013 with other TGFβ activated genes supported the functional role of this translated lncRNA in defining the transcriptional identity of FB4. This study furthermore supported that the function of unknown genes can be studied with scRNAseq and snRNAseq data by computing coexpression or computing enrichment of gene expression modules (Aibar et al., 2017; Kamimoto et al., 2020). In collaboration with the Mayer lab, the regional expression of Versican (VCAN) and ADAMTS5 was studied. Previous lab work has shown that ADAMTS5 controls VCAN turnover (Barallobre-Barreiro et al., 2021). In addition to the regulation on the protein level, VCAN RNA was more abundant in the LA and ADAMTS5 was higher abundant in the LV. Furthermore VCAN and ADAMTS5 were both enriched in fibroblasts, with VCAN being additionally expressed in one myeloid population.

Between 2020 and 2022, multiple groups have analysed their self-generated human ventricular scRNAseq and snRNAseq data and characterized the fibroblast compartment (Tucker et al., 2020; Kuppe et al., 2022; Koenig et al., 2022). The described fibroblast states in this study were also described in at least one of these studies. The best annotation overlap was

with Koenig et al. where all states were independently re-identified. Tucker and Kuppe et al. only reported 4 fibroblast states each. All studies reported the canonical ventricular FB, a TGFβ-stimulated and an ECM remodeling population.

7.3 From the Heart Cell Atlas to studying heart failure

Annotations from the healthy heart atlas were transferred to failing heart snRNAseq data to study transcriptional and compositional changes in dilated and arrhythmogenic cardiomyopathy. Here snRNAseq of LV and RV samples of in total 61 failing human hearts (46 patients within the main genotype subgroups) was applied, yielding more than 880,000 high-quality nuclei. In this dataset, major heart failure-causing mutations are analyzed and transcriptionally compared to each other (Haas et al., 2015). snRNAseq allows understanding differences in compositional shifts and understanding transcriptional changes of rarer populations, for example immune cells or cell states, which are drowning in bulk RNAseq data. Additionally, clinical information on all patient subgroups was statistically evaluated for genotype-specific differences.

Differences in altered left and right ventricular geometry were observed. RBM20 patients showed the highest increase in LVIDs and LVIDd, LMNA patients showed the lowest average increase, with PKP2 patients showing no increase. Instead, PKP2 patients showed increased RVEDD, implying a strong right-ventricular phenotype. This reflects the aggressiveness of a pathogenic variant in RBM20, which is further underlined by the younger age of RBM20 patients. RBM20 patients were mostly between 20 and 30 years in contrast to the other genotype subgroups, where most patients were on average 40 to 50 years old. It is known that heart failure is a significant risk factor for kidney function, induced by increased blood pressure building up in the kidney and reduced oxygen supply. Due to the aggressive progression and early need for transplantation, RBM20 patients have less damage of kidney function (eGFR) compared to the other genotype subgroups. A recent study by Hey et al. identified similar clinical characteristics of RBM20-mutated patients in comparisons to Laminopathies (Hey et al., 2020). In their analysis, they report a mean age of 35 years at diagnosis for RBM20 patients, as for LMNA patients it was above 40. Furthermore, RBM20 patients had more dilated LVs compared to LMNA patients.

Upon cell type annotation of snRNAseq data, compositional analysis was performed per cell type and significances were computed. Here we applied the center log transformation and then calculated the significance to have a robust metric to account for the sum constraint of

100 for relative proportions. Our method performed similarly to the bayesian model included in scCoda (Buettner et al., 2021), or propeller, a linear model-based approach with arcsin square root transformation of proportions (Phipson et al., 2022). CMs overall significantly decreased in heart failure in LV and RV, except for LMNA LV and TTN RV. Overall compositional changes were less pronounced in LMNA than in the TTN group. This might be influenced by the mutation, which not only affects cardiomyocytes, but all cell types in the heart, leading not only to a decline of myocytes (Coste Pradas et al., 2020). In contrast, endothelial and immune cells significantly increased in all DCM subgroups, with highest relative increase to CMs in PVneg and RBM20. The effects in DCM in LV were mirrored for PKP2 patients in RV, reflected by altered RV geometry in ACM (Thiene et al., 2007). The increase of endothelial cells suggests formation of new vessels or microvasculature to support the stressed tissue with oxygen and nutrients, while the increased abundance of immune cells reflects the inflammatory reaction. Fibroblasts were slightly, but not significantly increased in heart failure despite the histopathological finding of fibrosis, confirmed by measuring hydroxyproline levels. This motivated further subcluserting of the fibroblast compartment.
The increase in nuclei number of samples processed with the more sensitive 10x 3' V3 chemistry furthermore increased statistical power to delineate additional cell states. vFB1.1 with increased expression of genes involved in lipid metabolism such as APOD, APOE, APOO were delineated. vFB1.2 showed increased expression of genes associated with cytoskeletal function, suggesting a migratory state. Both populations could not be separated from the canonical vFB1.0 state in the Healthy Heart Atlas project. During sub-clustering, TGFβ activated vFB2 were increased, with highly significant increase in TTN and PVneg hearts. After cell state ratio analysis, canonical vFB1.0 were reduced, which explains the increased proportion of vFB2. Based on marker genes for cell cycling, no significant proliferating population was identified. This indicates that the secretory phenotype of fibroblasts dominated a proliferative phenotype. The increase of vFB2, with approximately stable vFB4 abundance, showed that fibrosis might result from imbalanced TGFβ-driven ECM production and organization. PKP2 hearts showed no proportional increase in vFB2, despite elevated levels of collagen. This observation was either due to the altered half-life of collagens, or other cell types increase their extracellular matrix production. Suspected cell states were mural cells, especially SMC2 or PC3 (Reichart et al., 2022). In contrast to the increase of vFB2, OSM-stimultaed vFB3 decreased. The reciprocal relationship was especially pronounced in the TTN subgroup, indicating disturbance in the macrophage:fibroblast interaction axis.

The comparison of differentially expressed genes in LV and RV identified a common heart failure signature, a DCM-phenotype and genotype specific signature in LV. In RV, no DCM-specific signature was identified, potentially reflected by the mirrored DCM LV-phenotype in PKP2 RV. Differentially expressed genes were calculated by first creating a pseudobulk per cell state per region per patient. So far, t-tests or Wilcoxon rank-sum tests, which estimate the distribution of gene expression per group (e.g. cell type, sample, genotype subgroup) are broadly used and implemented in Seurat and Scanpy for differential gene expression analysis. By this, each patient was given a weight within the distribution depending on the recovered nuclei number per sample. Pseudobulking UMIs and then applying edgeR accounts for differences in nuclei recovery, as each patient was given the same weight in the model. edgeR accounted for differences in library size, which was here reflected by differences in UMI counts per nucleus and overall nuclei recovery per patient. Described fold changes in this publication were hence on the patient level and not based on the distribution of UMI detection across all nuclei from a genotype subgroup. edgeR applied to scRNAseq data was shown to reduce the number of false positive discoveries (Squair et al., 2021).

The myeloid compartment was subclustered to study changes in the macrophage:fibroblast interaction axis, indicated by the decrease of vFB3 in the TTN subgroup. Similarly to the other cell states, annotation of the myeloid compartment benefited from the increased number of nuclei. Myeloids coarsely split into monocytes, macrophages and dendritic cells, which could be further subdivided to in total 14 cell states, and 3 unclassified states. Macrophages, monocytes, and dendritic cells increased in abundance during heart failure. The increase in myeloid cells might arise from the recruitment of circulating monocytes, or proliferation of tissue-resident macrophages at earlier stages of disease progression. At end stage, proliferation of myeloids was significantly downregulated in all genotypes. Furthermore, monocyte recruitment via CCR2/CCL2 axis, which was driven by vFB3, with significant dysregulation in TTN, was not downregulated in RBM20 or PVneg subgroups. Pro-inflammatory TREM2+ macrophages were significantly more abundant in LMNA and PVneg hearts compared to other genotype subgroups. Interferon-stimulated (ISG) macrophages were more abundant in PKP2 LV compared to other genotype subgroups, which might to a preserved left ventricular function during heart failure. Despite observed abundance changes in the myeloid compartment, further studies are needed to define the role of the myeloid compartment during disease progression. It is broadly accepted, that myeloid responses and cytokine peaks differ during disease progression, suggesting additionally differences in the myeloid

composition over disease progression (Liao et al., 2018; Dewald et al., 2005). Despite the increase in macrophages, nearly no macrophage proliferation was observed and recruitment of macrophages via the CCL2 axis was downregulated. Studying earlier disease time points, or acute diseases using endomyocardial biopsies will provide insights into early macrophage responses.

To confirm the results that heart failure is not a common mechanism, machine learning algorithms were applied to classify the most probable mutated gene per patient based on the transcriptional signature signature in the snRNAseq dataset. Interpretable linear machine learning methods such as logistic regression showed low performance in discriminating different genotype subgroups in a multiclass classification setting. The algorithm was trained on all expressed genes per cell type. Regularized linear models such as Lasso or Ridge regression showed no significantly better performance. After comparing additional non-linear algorithms, graph attention networks showed the best performance with high accuracy prediction. This suggested that snRNAseq contains genotype-specific transcriptional signatures, which are recognized by the GAT model. Some patients were classified with lower accuracy by the GAT model. These patients were genotypically distinct from the other patients in the genotype subgroup. For example, patient H10 of the PKP2 group additionally had a mutation in MYBPC3. RBM20 was the subgroup with the highest number of patients not being predicted with more than 75% certainty. RBM20 patients mostly had mutations in the RS domain, either at amino acid position 635 or 638 (Figure 37). Two patients, H22 and H33 had additional mutations in MYBPC3 and TTN respectively, whereas H20s mutation was located in the E-rich domain, showing a functionally distinct phenotype from an RS domain mutated patient (Gaertner et al., 2020). Patient H15 was with an age between 50-59 years older than the average of the RBM20 group. These results suggested that subgroups within genotype subgroups exist, showing the importance to expand this heart failure dataset in the future.

Interpretation of graph networks however remained challenging due to the inclusion of topological features of the snRNAseq data-based kNN graph. Furthermore, the model used node features and the adjacency matrix as inputs. Multiple approaches were proposed to find local interpretations of decisions done by graph attention networks, such as GNNExplainer or GraphLime (Ying et al., 2019; Huang et al., 2022). The applicability of these two methods on our framework however needs to be evaluated.

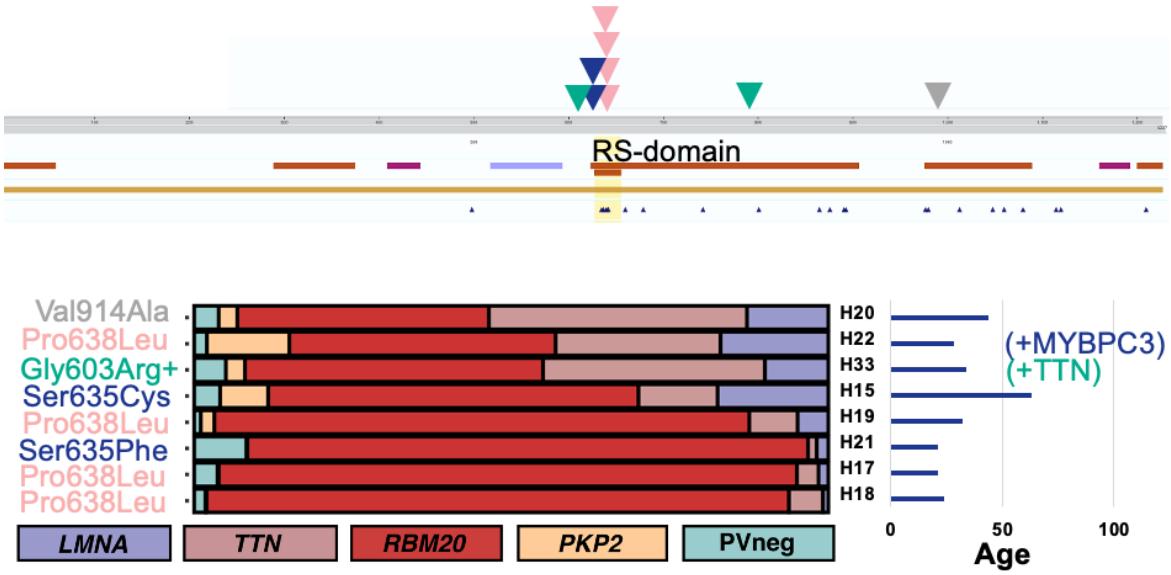

Figure 37: Location of mutations of patients in the RBM20 subgroup. The protein track from Uniprot was plotted on top, with location of the RS domain highlighted. Each arrow on top of the track corresponds to the mutation of one patient. One patient, H33, had a double mutation in RBM20 (turquoise). The stacked barplot below shows the probabilities per genotype per patient, as shown in Figure 34. On the right, the patient age (years) and mutations in other genes are shown.

8 Outlook

snRNAseq is a powerful technology to study changes in cell type composition and transcription. However, there are multiple technical improvements to be considered for future studies. For example, with the current snRNAseq protocols, a counted number of nuclei is loaded on the 10x Chromium controller. The space however which was previously occupied by those cells was unknown. Furthermore, spatial information of individual cell types is lost during tissue dissociation. This readout might be of interest when characterising disease with histological clear pattern, such as borderline myocarditis, perivascular fibrosis, fibrofatty plagues, cardiomyocyte hypertrophy or vessel morphology. TTN genoytpes for example showed a strong dysregulation between the fibroblast:macrophage interaction axis, suggesting genotype-specific changes in specific cellular neighbourhoods. RNAscope allowed imaging of spatial neighbourhoods, but restricts analysis to four probes, limiting interpretability when comparing multiple states. In this study, each cell state was stained with two markers: a pan-cell marker, for example DCN for fibroblasts, and a cell state-specific marker. The emergence of sequencing and imaging based spatial techniques, such as Spatial Transcriptomics, Slide-seq, or Merscope might reveal new insights on how cellular neighbourhoods change in situ and help to identify relevant cell-cell interactions. One challenge for imaging-based techniques is the segmentation of cell types, which was adressed in this project by incorporating WGA in the RNAscope v2 workflow. Due to differences in cell size and localization, segmentation algorithms need to be specifically benchmarked on accuracy. For example, large CMs were easy segmentable due to their large diameter, which was often more than 30μm. Cell types located in the interstitial space such as fibroblasts however appeared to be very compressed with little cytosole colocalizing to the nucleus.

Ambient RNA noise is a challenge in droplet-based snRNAseq technologies. Ambient RNA furthermore complicates the identification of nuclei with low transcriptional complexity, such as neutrophils (Hay et al., 2018). Computational tools for removing ambient RNA were developed, such as SoupX (Young and Behjati, 2020). The disadvantage of this method is the a priori definition of background genes and estimating the amount of ambient RNA per droplet based on this subset. This assumption is violated in the heart, where the genes with highest average noise are cardiomyocyte derived genes, removing biological relevant transcriptional signals from those. As shown in figures 20 and 22, cardiomyocytes are the most abundant cell type with expression of a high number of genes and transcripts per gene.

A promising method was published in the last years, Cellbender V2, which also introduced improved barcode-calling algorithms (Fleming et al., 2019). Alternative algorithms to Emptydrops identifying cell-containing droplets with low UMI counts include also EmptyNN (Yan et al., 2021). Applying those algorithms prior to clustering will improve robustness of marker gene identification. The application of a background-removal will furthermore provide further power on the identification of co-expressed genes and improves cross-tissue comparisons between cell types and states.

This study focused on transcriptional changes in heart failure. Single-cell multiomics technologies will build a more refined understanding on disease mechanisms and cellular plasticity. Measuring transcription and DNA accessibility in the same nucleus allows identification of key transcription factors driving the fate of a cell type or state (Kuppe et al., 2022). First test runs with the 10x multiome approach with human cardiac tissue showed decreased number of UMI recovery, maybe due to a previous permeabilization step needed for the transposase. Furthermore, Hoechst intercalates with DNA, increasing noise of snATACseq data. Further benchmarking experiments are needed to improve the robustness of this protocoll for human cardiac tissue.

Besides technical improvements, the collection of further samples is needed to understand cardiomyopathy subtypes. Especially sex, age and ethnicity driven effects could not further be studied here due to sample availability. Studying protective effects or risk factors associated with those variables might further improve disease prevention. Besides protective effects, the Heart and Diabetes Center in Bad Oeynhausen developed a VAD explantation protocol, leading to successful weaning of the VAD (Gyoten et al., 2021). The reason why weaning is only for a subset of patients possible is incompletely understood. All patients included in this study showed no improvement of cardiac function after VAD implantation. A genotype-stratified analysis of successful weaned patients versus patients who were transplanted might reveal potential protective mechanisms and help understanding cardiac remodelling.

In this study only 4 DCM genotype subgroups and 1 ACM subgroup was analysed, although pathogenic variants in more than 40 genes are currently known to be associated with DCM and ACM (Haas et al., 2015). Incorporating further genotype subgroups might help to identify similar genotype subgroups, for example mutations of proteins located in the same subcellular localization. In particular, it would be interesting to understand wether mutations in proteins of thick or thin filaments cause a similar molecular phenotype. This is why all data collected in this study, even those of rare genotypes, were annotated for cell types and states and were

made publically available for further downstream analysis. Processed and annotated data are deposited on CZIs cellxgene (https://cellxgene.cziscience.com/collections/ e75342a8-0f3b-4ec5-8ee1-245a23e0f7cb) and raw count matrices are available on Zenodo (https://zenodo.org/record/6962685# .YuyuihxBzKE).

References

(2020). Anaconda software distribution.

10x Genomics (2019a). *Single Cell 3' Reagent Kits v2 User Guide*, cg00052 rev f edition.

10x Genomics (2019b). *Single Cell 3' Reagent Kits v3.1 User Guide*, cg000204 rev c edition.

10x Genomics (2020). *Single Cell 3' Reagent Kits v3 User Guide*, cg000183 rev c edition.

Abe, H., Takeda, N., Isagawa, T., Semba, H., Nishimura, S., Morioka, M. S., Nakagama, Y., Sato, T., Soma, K., Koyama, K., et al. (2019). Macrophage hypoxia signaling regulates cardiac fibrosis via oncostatin m. *Nature communications*, 10(1):1–10.

Adams Jr, K. F., Fonarow, G. C., Emerman, C. L., LeJemtel, T. H., Costanzo, M. R., Abraham, W. T., Berkowitz, R. L., Galvao, M., Horton, D. P., Committee, A. S. A., Investigators, et al. (2005). Characteristics and outcomes of patients hospitalized for heart failure in the united states: rationale, design, and preliminary observations from the first 100,000 cases in the acute decompensated heart failure national registry (adhere). *American heart journal*, 149(2):209–216.

Aibar, S., González-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. *Nature methods*, 14(11):1083–1086.

Ammirati, E., Oliva, F., Cannata, A., Contri, R., Colombo, T., Martinelli, L., and Frigerio, M. (2014). Current indications for heart transplantation and left ventricular assist device: a practical point of view. *European journal of internal medicine*, 25(5):422–429.

Amoah, A. and Kallen, C. (2000). Aetiology of heart failure as seen from a national cardiac referral centre in africa. *Cardiology*, 93(1-2):11–18.

Argenziano, M. A., Doss, M. X., Tabler, M., Sachinidis, A., and Antzelevitch, C. (2019). Transcriptional changes associated with advancing stages of heart failure underlie atrial and ventricular arrhythmogenesis. *PloS one*, 14(5):e0216928.

Awtry, E. H. and Philippides, G. J. (2010). Alcoholic and cocaine-associated cardiomyopathies. *Progress in cardiovascular diseases*, 52(4):289–299.

Bajpai, G., Bredemeyer, A., Li, W., Zaitsev, K., Koenig, A. L., Lokshina, I., Mohan, J., Ivey, B., Hsiao, H.-M., Weinheimer, C., et al. (2019). Tissue resident ccr2- and ccr2+

cardiac macrophages differentially orchestrate monocyte recruitment and fate specification following myocardial injury. *Circulation research*, 124(2):263–278.

Bajpai, G., Schneider, C., Wong, N., Bredemeyer, A., Hulsmans, M., Nahrendorf, M., Epelman, S., Kreisel, D., Liu, Y., Itoh, A., et al. (2018). The human heart contains distinct macrophage subsets with divergent origins and functions. *Nature medicine*, 24(8):1234–1245.

Barallobre-Barreiro, J., Radovits, T., Fava, M., Mayr, U., Lin, W.-Y., Ermolaeva, E., Martínez-López, D., Lindberg, E. L., Duregotti, E., Daróczi, L., et al. (2021). Extracellular matrix in heart failure: Role of adamts5 in proteoglycan remodeling. *Circulation*, 144(25):2021–2034.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bensley, J. G., De Matteo, R., Harding, R., and Black, M. J. (2016). Three-dimensional direct measurement of cardiomyocyte volume, nuclearity, and ploidy in thick histological sections. *Scientific reports*, 6(1):1–10.

Bernstein, N. J., Fong, N. L., Lam, I., Roy, M. A., Hendrickson, D. G., and Kelley, D. R. (2020). Solo: doublet identification in single-cell rna-seq via semi-supervised deep learning. *Cell Systems*, 11(1):95–101.

Beyer, C., Schett, G., Gay, S., Distler, O., and Distler, J. H. (2009). Hypoxia. hypoxia in the pathogenesis of systemic sclerosis. *Arthritis research & therapy*, 11(2):1–9.

Bhavsar, I., Miller, C. S., and Al-Sabbagh, M. (2015). Macrophage inflammatory protein-1 alpha (mip-1 alpha)/ccl3: as a biomarker. *General methods in biomarker research and their applications*, page 223.

Bozkurt, B., Coats, A. J., Tsutsui, H., Abdelhamid, M., Adamopoulos, S., Albert, N., Anker, S. D., Atherton, J., Böhm, M., Butler, J., Drazner, M. H., Felker, G. M., Filippatos, G., Fonarow, G. C., Fiuzat, M., Gomez-Mesa, J.-E., Heidenreich, P., Imamura, T., Januzzi, J., Jankowska, E. A., Khazanie, P., Kinugawa, K., Lam, C. S., Matsue, Y., Metra, M., Ohtani, T., Francesco Piepoli, M., Ponikowski, P., Rosano, G. M., Sakata, Y., SeferoviĆ, P., Starling, R. C., Teerlink, J. R., Vardeny, O., Yamamoto, K., Yancy, C., Zhang, J.,

and Zieroth, S. (2021). Universal definition and classification of heart failure: A report of the heart failure society of america, heart failure association of the european society of cardiology, japanese heart failure society and writing committee of the universal definition of heart failure. *Journal of Cardiac Failure*, 27(4):387–413.

Bretaud, S., Guillon, E., Karppinen, S.-M., Pihlajaniemi, T., and Ruggiero, F. (2020). Collagen xv, a multifaceted multiplexin present across tissues and species. *Matrix Biology Plus*, 6:100023.

Brown, C. C., Gudjonson, H., Pritykin, Y., Deep, D., Lavallée, V.-P., Mendoza, A., Fromme, R., Mazutis, L., Ariyan, C., Leslie, C., et al. (2019). Transcriptional basis of mouse and human dendritic cell heterogeneity. *Cell*, 179(4):846–863.

Buettner, M., Ostner, J., Mueller, C. L., Theis, F. J., and Schubert, B. (2021). sccoda is a bayesian model for compositional single-cell data analysis. *Nature communications*, 12(1):1–10.

Burke, M. A., Cook, S. A., Seidman, J. G., and Seidman, C. E. (2016). Clinical and mechanistic insights into the genetics of cardiomyopathy. *Journal of the American College of Cardiology*, 68(25):2871–2886.

Burstein, B., Libby, E., Calderone, A., and Nattel, S. (2008). Differential behaviors of atrial versus ventricular fibroblasts: a potential role for platelet-derived growth factor in atrial-ventricular remodeling differences. *Circulation*, 117(13):1630–1641.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, 36:411–420.

Coats, C. J., Quarta, G., Flett, A. S., Pantazis, A. A., McKenna, W. J., and Moon, J. C. (2009). Arrhythmogenic left ventricular cardiomyopathy. *Circulation*, 120(25):2613–2614.

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). Upsetr: an r package for the visualization of intersecting sets and their properties. *Bioinformatics*.

Corrado, D., Marra, M. P., Zorzi, A., Beffagna, G., Cipriani, A., De Lazzari, M., Migliore, F., Pilichou, K., Rampazzo, A., Rigato, I., et al. (2020). Diagnosis of arrhythmogenic cardiomyopathy: the padua criteria. *International journal of cardiology*, 319:106–114.

Coste Pradas, J., Auguste, G., Matkovich, S. J., Lombardi, R., Chen, S. N., Garnett, T., Chamberlain, K., Riyad, J. M., Weber, T., Singh, S. K., et al. (2020). Identification of genes and pathways regulated by lamin a in heart. *Journal of the American Heart Association*, 9(16):e015690.

Daniels, A., Van Bilsen, M., Goldschmeding, R., Van Der Vusse, G., and Van Nieuwenhoven, F. (2009). Connective tissue growth factor and cardiac fibrosis. *Acta physiologica*, 195(3):321–338.

D'Ascenzi, F., Solari, M., Corrado, D., Zorzi, A., and Mondillo, S. (2018). Diagnostic differentiation between arrhythmogenic cardiomyopathy and athlete's heart by using imaging. *JACC: Cardiovascular Imaging*, 11(9):1327–1339.

Daseke II, M. J., Tenkorang, M. A., Chalise, U., Konfrst, S. R., and Lindsey, M. L. (2020). Cardiac fibroblast activation during myocardial infarction wound healing: fibroblast polarization after mi. *Matrix biology*, 91:109–116.

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. *Developmental cell*, 39(4):480–490.

Derks, W. and Bergmann, O. (2020). Polyploidy in cardiomyocytes: roadblock to heart regeneration? *Circulation research*, 126(4):552–565.

Dewald, O., Zymek, P., Winkelmann, K., Koerting, A., Ren, G., Abou-Khamis, T., Michael, L. H., Rollins, B. J., Entman, M. L., and Frangogiannis, N. G. (2005). Ccl2/monocyte chemoattractant protein-1 regulates inflammatory responses critical to healing myocardial infarcts. *Circulation research*, 96(8):881–889.

Dey, G., Radhakrishnan, A., Syed, N., Thomas, J. K., Nadig, A., Srikumar, K., Mathur, P. P., Pandey, A., Lin, S.-K., Raju, R., et al. (2013). Signaling network of oncostatin m pathway. *Journal of Cell Communication and Signaling*, 7(2):103–108.

Dick, S. A., Macklin, J. A., Nejat, S., Momen, A., Clemente-Casares, X., Althagafi, M. G., Chen, J., Kantores, C., Hosseinzadeh, S., Aronoff, L., et al. (2019). Self-renewing resident cardiac macrophages limit adverse remodeling following myocardial infarction. *Nature immunology*, 20(1):29–39.

Eckenstaler, R., Sandori, J., Gekle, M., and Benndorf, R. A. (2021). Angiotensin ii receptor type 1–an update on structure, expression and pathology. *Biochemical Pharmacology*, 192:114673.

Elliott, P. M., Anastasakis, A., Asimaki, A., Basso, C., Bauce, B., Brooke, M. A., Calkins, H., Corrado, D., Duru, F., Green, K. J., et al. (2019). Definition and treatment of arrhythmogenic cardiomyopathy: an updated expert panel report. *European journal of heart failure*, 21(8):955–964.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The reactome pathway knowledgebase. *Nucleic acids research*, 46(D1):D649–D655.

Fairweather, D., Cooper Jr, L. T., and Blauwet, L. A. (2013). Sex and gender differences in myocarditis and dilated cardiomyopathy. *Current problems in cardiology*, 38(1):7–46.

Felker, G. M., Shaw, L. K., and O'Connor, C. M. (2002). A standardized definition of ischemic cardiomyopathy for use in clinical research. *Journal of the American College of Cardiology*, 39(2):210–218.

Fleming, S. J., Marioni, J. C., and Babadi, M. (2019). Cellbender remove-background: a deep generative model for unsupervised removal of background noise from scrna-seq datasets. *BioRxiv*, page 791699.

Francone, O. L., Gurakar, A., and Fielding, C. (1989). Distribution and functions of lecithin: cholesterol acyltransferase and cholesteryl ester transfer protein in plasma lipoproteins: evidence for a functional unit containing these activities together with apolipoproteins ai and d that catalyzes the esterification and transfer of cell-derived cholesterol. *Journal of Biological Chemistry*, 264(12):7066–7072.

Gaertner, A., Klauke, B., Felski, E., Kassner, A., Brodehl, A., Gerdes, D., Stanasiuk, C., Ebbinghaus, H., Schulz, U., Dubowy, K.-O., et al. (2020). Cardiomyopathy-associated mutations in the rs domain affect nuclear localization of rbm20. *Human Mutation*, 41(11):1931–1943.

Gallini, R., Lindblom, P., Bondjers, C., Betsholtz, C., and Andrae, J. (2016). Pdgf-a and pdgf-b induces cardiac fibrosis in transgenic mice. *Experimental cell research*, 349(2):282–290.

Gladka, M. M. (2021). Single-cell rna sequencing of the adult mammalian heart—state-of-the-art and future perspectives. *Current Heart Failure Reports*, 18(2):64–70.

Gladka, M. M., Molenaar, B., De Ruiter, H., Van Der Elst, S., Tsui, H., Versteeg, D., Lacraz, G. P., Huibers, M. M., 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(2):166–180.

Goldman, L. and AI, S. (2019). Goldman-cecil medicine: Twenty.

Grindberg, R. V., Yee-Greenbaum, J. L., McConnell, M. J., Novotny, M., O'Shaughnessy, A. L., Lambert, G. M., Araúzo-Bravo, M. J., Lee, J., Fishman, M., Robbins, G. E., et al. (2013). Rna-sequencing from single nuclei. *Proceedings of the National Academy of Sciences*, 110(49):19802–19807.

Groenewegen, A., Rutten, F. H., Mosterd, A., and Hoes, A. W. (2020). Epidemiology of heart failure. *European journal of heart failure*, 22(8):1342–1356.

Gyoten, T., Rojas, S. V., Fox, H., Hata, M., Deutsch, M.-A., Schramm, R., Gummert, J. F., and Morshuis, M. (2021). Cardiac recovery following left ventricular assist device therapy: experience of complete device explantation including ventricular patch plasty. *European Journal of Cardio-Thoracic Surgery*, 59(4):855–862.

Haas, J., Frese, K. S., Peil, B., Kloos, W., Keller, A., Nietsch, R., Feng, Z., Müller, S., Kayvanpour, E., Vogel, B., et al. (2015). Atlas of the clinical genetics of human dilated cardiomyopathy. *European heart journal*, 36(18):1123–1135.

Hakim, J. and Manyemba, J. (1998). Cardiac disease distribution among patients referred for echocardiography in harare, zimbabwe. *The Central African journal of medicine*, 44(6):140–144.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., McElrath, J. M., Blish, C. A., Gottardo, R., Smibert, P., and Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.e29.

Harkness, A., Ring, L., Augustine, D. X., Oxborough, D., Robinson, S., and Sharma, V. (2020). Normal reference intervals for cardiac dimensions and function for use in echocar-

diographic practice: a guideline from the british society of echocardiography. *Echo Research and Practice*, 7(1):G1–G18.

Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L., and Salomonis, N. (2018). The human cell atlas bone marrow single-cell interactive web portal. *Experimental hematology*, 68:51–61.

Heinig, M., Adriaens, M. E., Schafer, S., van Deutekom, H. W., Lodder, E. M., Ware, J. S., Schneider, V., Felkin, L. E., Creemers, E. E., Meder, B., et al. (2017). Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome biology*, 18(1):1–21.

Hershberger, R. E. and Jordan, E. (2021). Dilated cardiomyopathy overview. *Genereviews®[internet]*.

Hey, T. M., Rasmussen, T. B., Madsen, T., Aagaard, M. M., Harbo, M., Mølgaard, H., Nielsen, S. K., Haas, J., Meder, B., Møller, J. E., et al. (2020). Clinical and genetic investigations of 109 index patients with dilated cardiomyopathy and 445 of their relatives. *Circulation: Heart Failure*, 13(10):e006701.

Huang, Q., Yamada, M., Tian, Y., Singh, D., and Chang, Y. (2022). Graphlime: Local interpretable model explanations for graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*.

Itabashi, H., Maesawa, C., Oikawa, H., Kotani, K., Sakurai, E., Kato, K., Komatsu, H., Nitta, H., Kawamura, H., Wakabayashi, G., et al. (2008). Angiotensin ii and epidermal growth factor receptor cross-talk mediated by a disintegrin and metalloprotease accelerates tumor cell proliferation of hepatocellular carcinoma cell lines. *Hepatology Research*, 38(6):601–613.

Jaitin, D. A., Adlung, L., Thaiss, C. A., Weiner, A., Li, B., Descamps, H., Lundgren, P., Bleriot, C., Liu, Z., Deczkowska, A., et al. (2019). Lipid-associated macrophages control metabolic homeostasis in a trem2-dependent manner. *Cell*, 178(3):686–698.

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., [...], Zenebe, Z. M., Zhang, K., Zhao, Z., Zhou, M., Zodpey, S., Zucker, I., Vos, T., and Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1789–1858.

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M. V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using cellchat. *Nature communications*, 12(1):1–20.

Kamimoto, K., Hoffmann, C. M., and Morris, S. A. (2020). Celloracle: Dissecting cell identity via network inference and in silico gene perturbation. *BioRxiv*.

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). Kegg: integrating viruses and cellular organisms. *Nucleic acids research*, 49(D1):D545–D551.

KAPABiosystems (2017). *KAPA Library Quantification Kit*, kr0405 - v9.17 edition.

Kassner, A., Oezpeker, C., Gummert, J., Zittermann, A., Gärtner, A., Tiesmeier, J., Fox, H., Morshuis, M., and Milting, H. (2021). Mechanical circulatory support does not reduce advanced myocardial fibrosis in patients with end-stage heart failure. *European Journal of Heart Failure*, 23(2):324–334.

Katz, A. M. and Katz, P. B. (1989). Homogeneity out of heterogeneity. *Circulation*, 79(3):712–717.

Kelder, T., Van Iersel, M. P., Hanspers, K., Kutmon, M., Conklin, B. R., Evelo, C. T., and Pico, A. R. (2012). Wikipathways: building research communities on biological pathways. *Nucleic acids research*, 40(D1):D1301–D1307.

Kindermann, I., Kindermann, M., Kandolf, R., Klingel, K., Bültmann, B., Müller, T., Lindinger, A., and Böhm, M. (2008). Predictors of outcome in patients with suspected myocarditis. *Circulation*, 118(6):639–648.

Koenig, A. L., Shchukina, I., Amrute, J., Andhey, P. S., Zaitsev, K., Lai, L., Bajpai, G., Bredemeyer, A., Smith, G., Jones, C., et al. (2022). Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure. *Nature Cardiovascular Research*, 1(3):263–280.

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2– an r package for gene list functional enrichment analysis and namespace conversion toolset g:profiler. *F1000Research*, 9 (ELIXIR)(709). R package version 0.2.1.

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. *Molecular cell*, 58(4):610–620.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-r., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with harmony. *Nature methods*, 16(12):1289–1296.

Krishnaswami, S. R., Grindberg, R. V., Novotny, M., Venepally, P., Lacar, B., Bhutani, K., Linker, S. B., Pham, S., Erwin, J. A., Miller, J. A., et al. (2016). Using single nuclei for rna-seq to capture the transcriptome of postmortem neurons. *Nature protocols*, 11(3):499–524.

Krüger-Genge, A., Blocki, A., Franke, R.-P., and Jung, F. (2019). Vascular endothelial cell biology: an update. *International journal of molecular sciences*, 20(18):4411.

Kumar, V., Abbas, A. K., and Aster, J. C. (2017). *Robbins basic pathology e-book*. Elsevier Health Sciences.

Kuppe, C., Ramirez Flores, R. O., Li, Z., Hayat, S., Levinson, R. T., Liao, X., Hannani, M. T., Tanevski, J., Wünnemann, F., Nagai, J. S., et al. (2022). Spatial multi-omic map of human myocardial infarction. *Nature*, 608(7924):766–777.

Lake, B. B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., Wildberg, A., Gao, D., Fung, H.-L., Chen, S., et al. (2016). Neuronal subtypes and diversity revealed by single-nucleus rna sequencing of the human brain. *Science*, 352(6293):1586–1590.

Lake, B. B., Chen, S., Hoshi, M., Plongthongkum, N., Salamon, D., Knoten, A., Vijayan, A., Venkatesh, R., Kim, E. H., Gao, D., et al. (2019). A single-nucleus rna-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nature communications*, 10(1):1–15.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2018). Clinvar: improving access to variant interpretations and supporting evidence. *Nucleic acids research*, 46(D1):D1062–D1067.

Liao, X., Shen, Y., Zhang, R., Sugi, K., Vasudevan, N. T., Alaiti, M. A., Sweet, D. R., Zhou, L., Qing, Y., Gerson, S. L., et al. (2018). Distinct roles of resident and nonresident macrophages in nonischemic cardiomyopathy. *Proceedings of the National Academy of Sciences*, 115(20):E4661–E4669.

Lin, C.-Y., Chung, F.-P., Lin, Y.-J., Chang, S.-L., Lo, L.-W., Hu, Y.-F., Tuan, T.-C., Chao, T.-F., Liao, J.-N., Chang, Y.-T., et al. (2017). Gender differences in patients with ar-

rhythmogenic right ventricular dysplasia/cardiomyopathy: clinical manifestations, electrophysiological properties, substrate characteristics, and prognosis of radiofrequency catheter ablation. *International Journal of Cardiology*, 227:930–937.

Lindberg, E. L. and Hübner, N. (2021). *Gene Expression Analysis and Next-Generation Sequencing*, pages 684–688. Springer International Publishing, Cham.

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(7838):466–472.

Liu, Z., Yue, S., Chen, X., Kubin, T., and Braun, T. (2010). Regulation of cardiomyocyte polyploidy and multinucleation by cycling1. *Circulation research*, 106(9):1498–1506.

Lotfollahi, M., Wolf, F. A., and Theis, F. J. (2019). scgen predicts single-cell perturbation responses. *Nature methods*, 16(8):715–721.

Lyden, D., Olszewski, J., Feran, M., Job, L., and Huber, S. (1987). Coxsackievirus b-3-induced myocarditis. effect of sex steroids on viremia and infectivity of cardiocytes. *The American journal of pathology*, 126(3):432.

Maharaj, B. (1991). Causes of congestive heart failure in black patients at king edward viii hospital, durban: A prospective study. *Cardiovascular Journal of Africa*, 2(1):31–32.

Marakhonov, A. V., Tabakov, V. Y., Zernov, N. V., Dadali, E. L., Sharkova, I. V., and Skoblov, M. Y. (2018). Two novel col6a3 mutations disrupt extracellular matrix formation and lead to myopathy from ullrich congenital muscular dystrophy and bethlem myopathy spectrum. *Gene*, 672:165–171.

Marcus, F. I., Fontaine, G. H., Guiraudon, G., Frank, R., Laurenceau, J. L., Malergue, C., and Grosgogeat, Y. (1982). Right ventricular dysplasia: a report of 24 adult cases. *Circulation*, 65(2):384–398.

Maron, B. J., Towbin, J. A., Thiene, G., Antzelevitch, C., Corrado, D., Arnett, D., Moss, A. J., Seidman, C. E., and Young, J. B. (2006). Contemporary definitions and classification of the cardiomyopathies: an american heart association scientific statement from the council on clinical cardiology, heart failure and transplantation committee; quality of care and outcomes research and functional genomics and translational biology interdisciplinary

working groups; and council on epidemiology and prevention. *Circulation*, 113(14):1807–1816.

Matsumura, Y., Takata, J., Kitaoka, H., Kubo, T., Baba, Y., Hoshikawa, E., Hamada, T., Okawa, M., Hitomi, N., Sato, K., et al. (2006). Long-term prognosis of dilated cardiomyopathy revisited an improvement in survival over the past 20 years. *Circulation Journal*, 70(4):376–383.

Maybaum, S., Mancini, D., Xydas, S., Starling, R. C., Aaronson, K., Pagani, F. D., Miller, L. W., Margulies, K., McRee, S., Frazier, O., et al. (2007). Cardiac improvement during mechanical circulatory support: a prospective multicenter study of the lvad working group. *Circulation*, 115(19):2497–2505.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

McKenna, W. J. and Judge, D. P. (2021). Epidemiology of the inherited cardiomyopathies. *Nature Reviews Cardiology*, 18(1):22–36.

Melleby, A. O., Strand, M. E., Romaine, A., Herum, K. M., Skrbic, B., Dahl, C. P., Sjaastad, I., Fiane, A. E., Filmus, J., Christensen, G., et al. (2016). The heparan sulfate proteoglycan glypican-6 is upregulated in the failing heart, and regulates cardiomyocyte growth through erk1/2 signaling. *PloS one*, 11(10):e0165079.

Mewis, C., Riessen, R., and Spyridopoulos, I. (2006). *Kardiologie compact: alles für Station und Facharztprüfung*. Georg Thieme Verlag.

Miao, Z., Moreno, P., Huang, N., Papatheodorou, I., Brazma, A., and Teichmann, S. A. (2020). Putative cell type discovery from single-cell gene expression data. *Nature methods*, 17(6):621–628.

Michel, M. C., Brunner, H. R., Foster, C., and Huo, Y. (2016). Angiotensin ii type 1 receptor antagonists in animal models of vascular, cardiac, metabolic and renal disease. *Pharmacology & therapeutics*, 164:1–81.

Miller, L., Birks, E., Guglin, M., Lamba, H., and Frazier, O. (2019). Use of ventricular assist devices and heart transplantation for advanced heart failure. *Circulation research*, 124(11):1658–1678.

Murphy, K. and Weaver, C. (2018). *Janeway immunologie*. Springer-Verlag.

Nadelmann, E. R., Gorham, J. M., Reichart, D., Delaughter, D. M., Wakimoto, H., Lindberg, E. L., Litviňukova, M., Maatz, H., Curran, J. J., Ischiu Gutierrez, D., et al. (2021). Isolation of nuclei from mammalian cells and tissues for single-nucleus molecular profiling. *Current Protocols*, 1(5):e132.

Paradisi, M., McClintock, D., Boguslavsky, R. L., Pedicelli, C., Worman, H. J., and Djabali, K. (2005). Dermal fibroblasts in hutchinson-gilford progeria syndrome with the lamin a g608g mutation have dysmorphic nuclei and are hypersensitive to heat stress. *BMC cell biology*, 6(1):1–11.

Patel, M. D., Mohan, J., Schneider, C., Bajpai, G., Purevjav, E., Canter, C. E., Towbin, J., Bredemeyer, A., and Lavine, K. J. (2017). Pediatric and adult dilated cardiomyopathy represent distinct pathological entities. *JCI insight*, 2(14).

Phipson, B., Sim, C. B., Porrello, E. R., Hewitt, A. W., Powell, J., and Oshlack, A. (2022). propeller: testing for differences in cell type proportions in single cell data. *Bioinformatics*, 38(20):4720–4726.

Pinto, A. R., Ilinykh, A., Ivey, M. J., Kuwabara, J. T., D'antoni, M. L., Debuque, R., Chandran, A., Wang, L., Arora, K., Rosenthal, N. A., et al. (2016a). Revisiting cardiac cellular composition. *Circulation research*, 118(3):400–409.

Pinto, A. R., Paolicelli, R., Salimova, E., Gospocic, J., Slonimsky, E., Bilbao-Cortes, D., Godwin, J. W., and Rosenthal, N. A. (2012). An abundant tissue macrophage population in the adult murine heart with a distinct alternatively-activated macrophage profile. *PloS one*, 7(5):e36814.

Pinto, Y. M., Elliott, P. M., Arbustini, E., Adler, Y., Anastasakis, A., Böhm, M., Duboc, D., Gimeno, J., De Groote, P., Imazio, M., et al. (2016b). Proposal for a revised definition of dilated cardiomyopathy, hypokinetic non-dilated cardiomyopathy, and its implications for clinical practice: a position statement of the esc working group on myocardial and pericardial diseases. *European heart journal*, 37(23):1850–1858.

Polański, K., Young, M. D., Miao, Z., Meyer, K. B., Teichmann, S. A., and Park, J.-E. (2020). Bbknn: fast batch alignment of single cell transcriptomes. *Bioinformatics*, 36(3):964–965.

Puig-Kröger, A., Sierra-Filardi, E., Domínguez-Soto, A., Samaniego, R., Corcuera, M. T., Gómez-Aguado, F., Ratnam, M., Sánchez-Mateos, P., and Corbí, A. L. (2009). Folate

receptor β is expressed by tumor-associated macrophages and constitutes a marker for m2 anti-inflammatory/regulatory macrophagesfolate receptor β is an m2 macrophage marker. *Cancer research*, 69(24):9395–9403.

Quaife, N., Chothani, S., Schulz, J., Lindberg, E., Vanezis, K., Adami, E., O'Fee, K., Greiner, J., Litviňuková, M., van Heesch, S., et al. (2022). Linc01013 is a determinant of fibroblast activation and encodes a novel fibroblast-activating micropeptide. *Journal of Cardiovascular Translational Research*, pages 1–9.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramirez Flores, R. O., Lanzer, J. D., Holland, C. H., Leuschner, F., Most, P., Schultz, J.-H., Levinson, R. T., and Saez-Rodriguez, J. (2021). Consensus transcriptional landscape of human end-stage heart failure. *Journal of the American Heart Association*, 10(7):e019667.

Ravindra, N., Sehanobish, A., Pappalardo, J. L., Hafler, D. A., and van Dijk, D. (2020). Disease state prediction from single-cell data using graph attention networks. In *Proceedings of the ACM conference on health, inference, and learning*, pages 121–130.

Reichart, D., Lindberg, E. L., Maatz, H., Miranda, A. M., Viveiros, A., Shvetsov, N., Gärtner, A., Nadelmann, E. R., Lee, M., Kanemaru, K., et al. (2022). Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies. *Science*, 377(6606):eabo1984.

Rizzo, G., Vafadarnejad, E., Arampatzi, P., Silvestre, J.-S., Zernecke, A., Saliba, A.-E., and Cochain, C. (2020). Single-cell transcriptomic profiling maps monocyte/macrophage transitions after myocardial infarction in mice. *BioRxiv*.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, 33:495–502.

Sayed, N., Liu, C., Ameen, M., Himmati, F., Zhang, J. Z., Khanamiri, S., Moonen, J.-R., Wnorowski, A., Cheng, L., Rhee, J.-W., et al. (2020). Clinical trial in a dish using

ipscs shows lovastatin improves endothelial dysfunction and cellular cross-talk in lmna cardiomyopathy. *Science translational medicine*, 12(554):eaax9276.

Schafer, S., Viswanathan, S., Widjaja, A. A., Lim, W.-W., Moreno-Moral, A., DeLaughter, D. M., Ng, B., Patone, G., Chow, K., Khin, E., et al. (2017). Il-11 is a crucial determinant of cardiovascular fibrosis. *Nature*, 552(7683):110–115.

Shiina, T., Hosomichi, K., Inoko, H., and Kulski, J. K. (2009). The hla genomic loci map: expression, interaction, diversity and disease. *Journal of human genetics*, 54(1):15–39.

Singh, S. N., Fletcher, R. D., Fisher, S. G., Singh, B. N., Lewis, H. D., Deedwania, P. C., Massie, B. M., Colling, C., and Lazzeri, D. (1995). Amiodarone in patients with congestive heart failure and asymptomatic ventricular arrhythmia. *New England Journal of Medicine*, 333(2):77–82.

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 reports*, 22(3):600–610.

Skrbic, B., Engebretsen, K. V., Strand, M. E., Lunde, I. G., Herum, K. M., Marstein, H. S., Sjaastad, I., Lunde, P. K., Carlson, C. R., Christensen, G., et al. (2015). Lack of collagen viii reduces fibrosis and promotes early mortality and cardiac dilatation in pressure overload in mice. *Cardiovascular research*, 106(1):32–42.

Souders, C. A., Bowers, S. L., and Baudino, T. A. (2009). Cardiac fibroblast: the renaissance cell. *Circulation research*, 105(12):1164–1176.

Squair, J. W., Gautier, M., Kathe, C., Anderson, M. A., James, N. D., Hutson, T. H., Hudelle, R., Qaiser, T., Matson, K. J., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. *Nature communications*, 12(1):1–15.

Stegemann, H. and Stalder, K. (1967). Determination of hydroxyproline. *Clinica chimica acta*, 18(2):267–273.

Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell rna-seq in the past decade. *Nature protocols*, 13(4):599–604.

Technologies, A. (2013). *Agilent High Sensitivity DNA Kit Guide*, g2938-90321 rev b. edition.

Thiene, G., Corrado, D., and Basso, C. (2007). Arrhythmogenic right ventricular cardiomyopathy/dysplasia. *Orphanet journal of rare diseases*, 2(1):1–16.

Thomas, T. P. and Grisanti, L. A. (2020). The dynamic interplay between cardiac inflammation and fibrosis. *Frontiers in Physiology*, 11:529075.

Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., et al. (2017). A subcellular map of the human proteome. *Science*, 356(6340):eaal3321.

Traag, V. A., Waltman, L., and Van Eck, N. J. (2019). From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12.

Tracy, R. E. and Sander, G. E. (2011). Histologically measured cardiomyocyte hypertrophy correlates with body height as strongly as with body mass index. *Cardiology research and practice*, 2011.

Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M., and Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell rna sequencing data. *Genome biology*, 21(1):1–32.

Travaglini, K. J., Nabhan, A. N., Penland, L., Sinha, R., Gillich, A., Sit, R. V., Chang, S., Conley, S. D., Mori, Y., Seita, J., et al. (2020). A molecular cell atlas of the human lung from single-cell rna sequencing. *Nature*, 587(7835):619–625.

Tucker, N. R., Chaffin, M., Fleming, S. J., Hall, A. W., Parsons, V. A., Bedi Jr, K. C., Akkad, A.-D., Herndon, C. N., Arduini, A., Papangeli, I., et al. (2020). Transcriptional and cellular diversity of the human heart. *Circulation*, 142(5):466–482.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. *Science*, 347(6220):1260419.

Uhlén, M., Karlsson, M. J., Hober, A., Svensson, A.-S., Scheffel, J., Kotol, D., Zhong, W., Tebani, A., Strandberg, L., Edfors, F., et al. (2019). The human secretome. *Science signaling*, 12(609):eaaz0274.

Van den Boogaart, K. G. and Tolosana-Delgado, R. (2008). "compositions": a unified r package to analyze compositional data. *Computers & Geosciences*, 34(4):320–338.

van den Brink, S. C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., Robin, C., and Van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature methods*, 14(10):935–936.

van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.-L., et al. (2019). The translational landscape of the human heart. *Cell*, 178(1):242–260.

Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

Vanhoutte, D. and Heymans, S. (2010). Timps and cardiac remodeling:'embracing the mmp-independent-side of the family'. *Journal of molecular and cellular cardiology*, 48(3):445–453.

Wang, L., Yu, P., Zhou, B., Song, J., Li, Z., Zhang, M., Guo, G., Wang, Y., Chen, X., Han, L., et al. (2020). Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. *Nature cell biology*, 22(1):108–119.

Wetzels, J., Kiemeney, L., Swinkels, D., Willems, H., and Den Heijer, M. (2007). Age-and gender-specific reference values of estimated gfr in caucasians: the nijmegen biomedical study. *Kidney international*, 72(5):632–637.

Wiggers, C. J. (1915). *Modern aspects of the circulation in health and disease*. Lea & Febiger.

Wolf, F. A., Angerer, P., and Theis, F. J. (2018). Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):1–5.

Wolock, S. L., Lopez, R., and Klein, A. M. (2019). Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell systems*, 8(4):281–291.

Woods, R. H. (1892). A few applications of a physical theorem to membranes in the human body in a state of tension. *Journal of anatomy and physiology*, 26(Pt 3):362.

Wu, J., Wu, Y.-Q., Ricklin, D., Janssen, B. J., Lambris, J. D., and Gros, P. (2009). Structure of complement fragment c3b–factor h and implications for host protection by complement regulators. *Nature immunology*, 10(7):728–733.

Yan, F., Zhao, Z., and Simon, L. M. (2021). Emptynn: A neural network based on positive and unlabeled learning to remove cell-free droplets and recover lost cells in scrna-seq data. *Patterns*, 2(8):100311.

Yang, H., Zhang, Q., Xu, M., Wang, L., Chen, X., Feng, Y., Li, Y., Zhang, X., Cui, W., and Jia, X. (2020). Ccl2-ccr2 axis recruits tumor associated macrophages to induce immune evasion through pd-1 signaling in esophageal carcinogenesis. *Molecular cancer*, 19(1):1–14.

Yekelchyk, M., Guenther, S., Preussner, J., and Braun, T. (2019). Mono-and multi-nucleated ventricular cardiomyocytes constitute a transcriptionally homogenous cell population. *Basic research in cardiology*, 114(5):1–13.

Ying, Z., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. (2019). Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32.

Young, M. D. and Behjati, S. (2020). Soupx removes ambient rna contamination from droplet-based single-cell rna sequencing data. *Gigascience*, 9(12):giaa151.

Zappia, L. and Theis, F. J. (2021). Over 1000 tools reveal trends in the single-cell rna-seq analysis landscape. *Genome biology*, 22(1):1–18.

Zhao, X., Kwan, J. Y. Y., Yip, K., Liu, P. P., and Liu, F.-F. (2020). Targeting metabolic dysregulation for fibrosis therapy. *Nature reviews Drug discovery*, 19(1):57–75.

9 Supplementary Figures

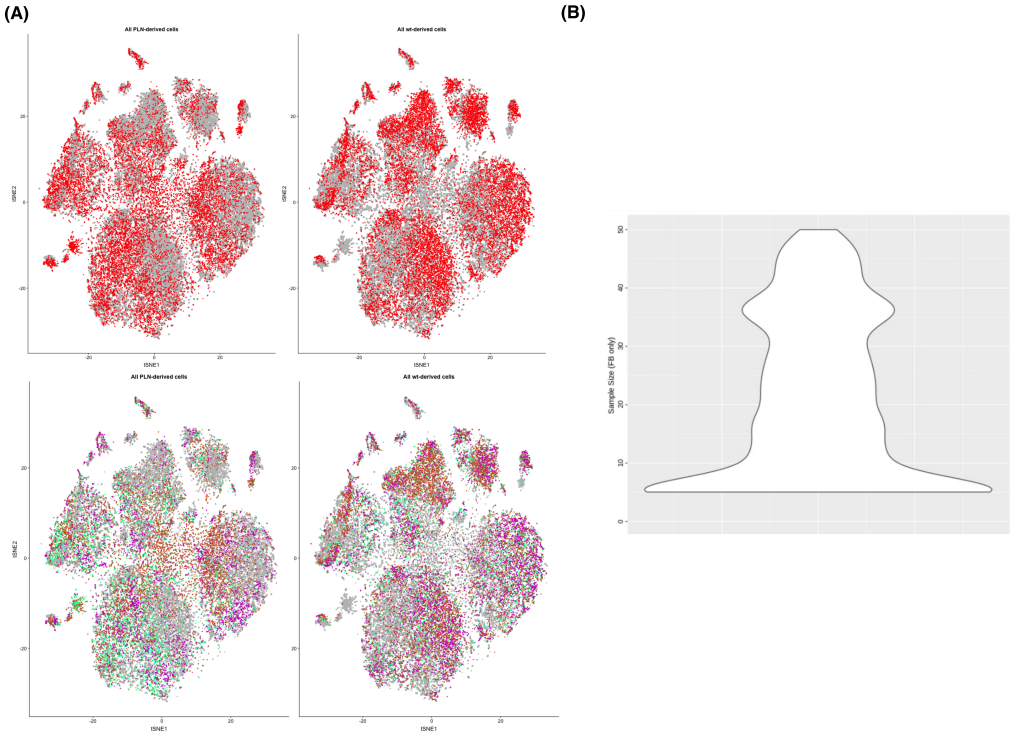

Figure S1: Pilot study using free-walls of 3 wildtype and 3 Phospholamban (PLN)-mutant mice. (A) 2-dimensional tSNE embedding of in total 12.000 nuclei. Each nucleus was colored according condition (top) and sample (bottom). (B) Distribution of sample size required to reach significance for one gene. Analysis was repeated for all genes. A large number of genes can be determined as differentially expressed with a low cohort size (n 10). Larger cohort sizes will yield deeper insights into the molecular mechanisms underlying DCM.

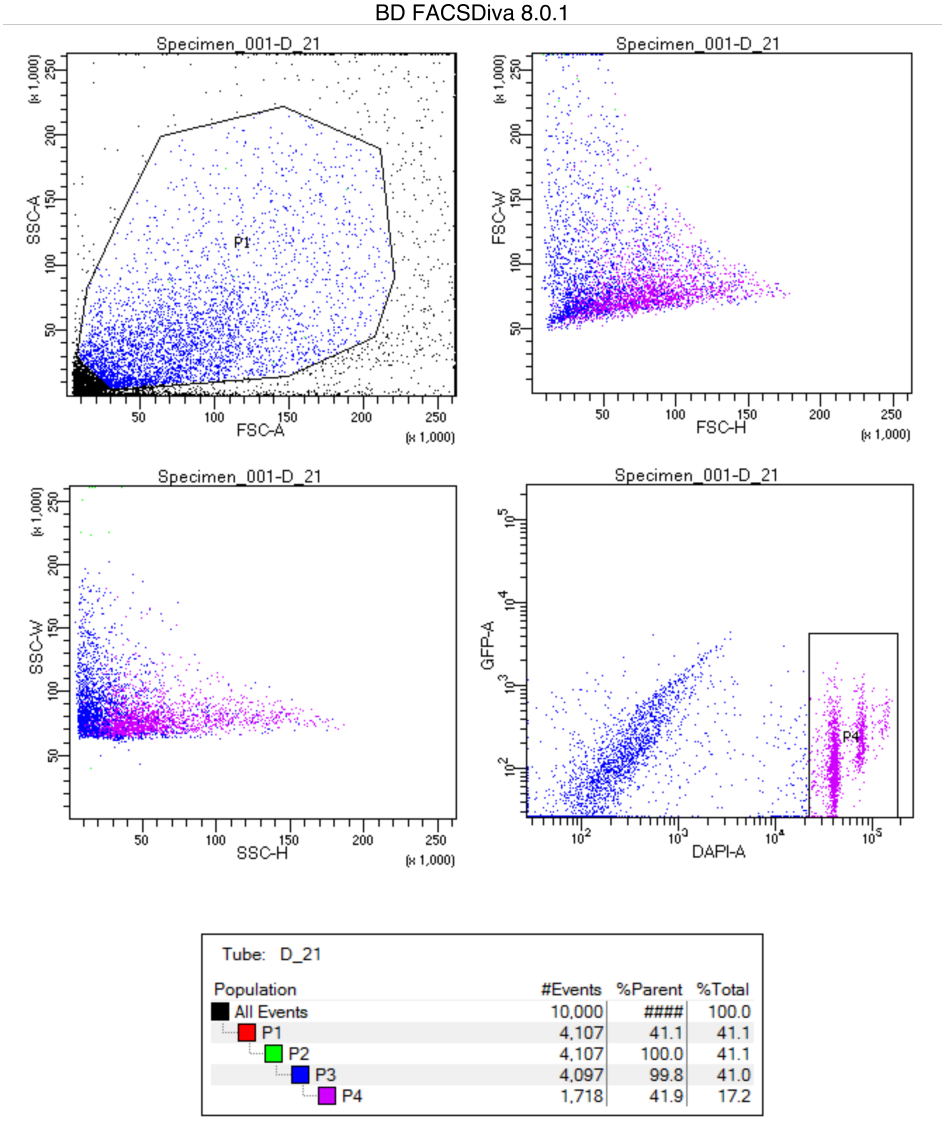

Figure S2: FACS purification of nuclei from human heart tissue homogenate. The P1 gate was used to remove very small particles representing cell debris. P4 was used to sort out NucBlue stained nuclei.

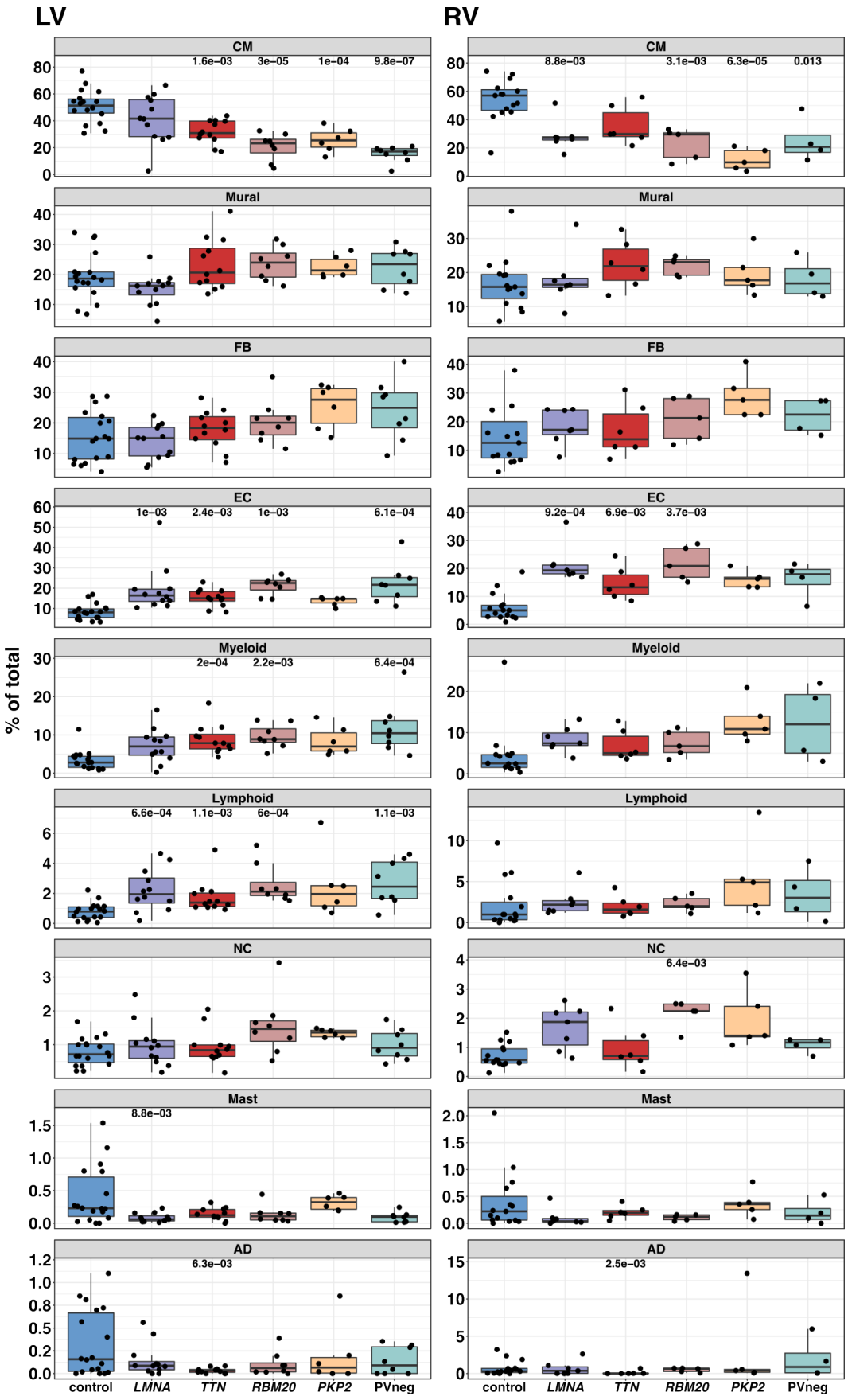

Figure S3: Relative abundance of cardiac cell-types in control and failing heart samples in LV and RV.

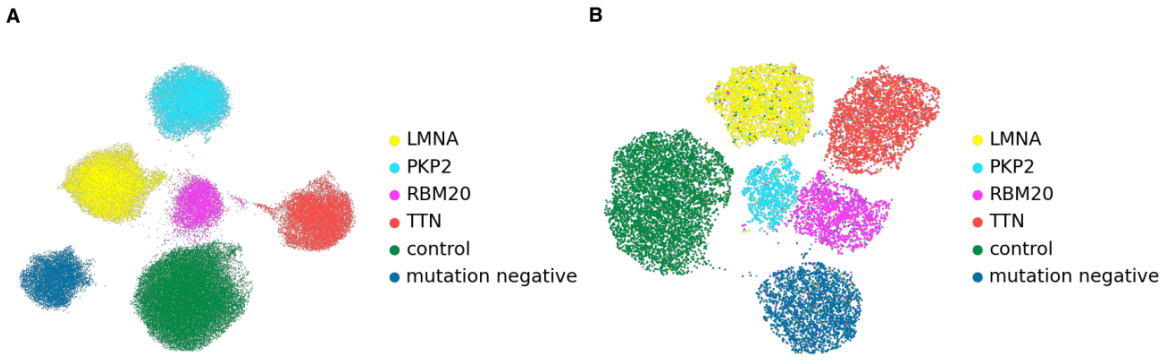

Figure S4: Latent space learned by scanVI. A) Without label shuffling and B) with label shuffling. For label shuffling, genotypes of patients were randomly assigned.

10 Acknowledgement

I thank Prof. Dr. Norbert Hübner for providing me the opportunity to work on this exciting project in his lab. My PhD was a great time and I learned a lot, on a scientific level, but especially how to work in large collaborations.
Furthermore I would like to thank my PhD committee members Prof. Dr. Markus Landthaler and Prof. Dr. Christoph Lippert for the regular discussions and feedback. Many thanks also to everyone who is reading this thesis and giving me feedback on my work and achievements, especially the members of the doctoral committee.
I would furthermore like to thank my great supervisor Dr. Henrike Maatz for the close collaboration on the projects, the input and critical feedback. There was a lot I learned from her, especially how to be target-oriented and how to be short and precise when writing a manuscript. To the members of the DCM Heart Atlas Consortium: Thanks to Prof. Dr. Christine E. Seidman and Prof. Dr. Jonathan Seidmann and Dr. Daniel Reichart, with whom I worked together on understanding the molecular basis of heart failure. Prof. Dr. Hendrik Milting, with whom I had many exciting discussions on heart failure and who has established an impressive biobank for studying human heart failure. Dr. Matthias Heinig, for introducing me to the everyday life of a purely computational biologist and hosting me during my Helmholtz Information & Data Science Academy (HIDA) Exchange to Munich. To my collaborators: Especially Dr. Giannino Patone and Sabine Schmidt, with who together I optimised the snRNAseq protocol and who helped generating this huge dataset. Dr. Andrew Woehler, who introduced me to the exciting world of microscopy. Nikolay Shvetsov for collaborating with me to implement the graph attention model - and also Prof. Dr. Christoph Lippert. A special thank you also to my family: My beloved wife Rumi, who supported me and was daily updated against her will on my worklife and current challenges. My mother, with who I exchanged feelings and thoughts and who gave me tremendous mental support. Finally, I want to thank all open source communities and people with whom I have discussed code online, such as github, stackexchange, and stackoverflow. Many thanks to everyone who supported me on my way.

11 Statement of Contribution by others

Nuclei purification and library preparation of 135 cardiac tissue samples included in the Heart Failure Atlas was done together with Dr. Giannino Patone and Sabine Schmidt (Hübner lab, Max-Delbrück Center for Molecular Medicine, Berlin, Germany). Construction of the graph attention network presented in 6.3.8 was done in collaboration with Nikolay Shvetsov (Hübner lab, Max-Delbrück Center for Molecular Medicine, Berlin, Germany) and Prof. Dr. Christoph Lippert (Hasso Plattner Institute, Potsdam, Germany). Approaches for compositional analysis were discussed, tested, and developed with Dr. Matthias Heinig in the course of the Helmholtz Information and Data Science Academy (HIDA) exchange program. Hydroxyproline measurements was done by Caroline Stanasiuk and Dr. Anna Gärtner at the Heart and Diabetes Center in Bad Oeynhausen. Patient genotyping and variant calling was done by physicians and Prof. Dr. Hendrik Miltings lab at the Heart and Diabetes Center in Bad Oeynhausen. ACMG scoring of patients mutations was done in collaboration with Dr. Anna Gärtner, Prof. Dr. Hendrik Milting, Dr. Daniel Reichart, Prof. Dr. Christine E. Seidman, and Prof. Dr. Gavin Oudit. snRNAseq libraries were sequenced at the Sequencing Core Facility of the Max Delbrück Center for Molecular Medicine, Berlin, Germany. Dr. Daniel Reichart has done RNAscope on healthy human heart samples (RNAscope v1) and validated the CST3+ FB population (RNAscope v2).

12 Permissions

Figures from section 6.2, included in the publication Litviňuková et al. (2020), were generated by me. Figures from section 6.3 were generated by me and were included in the publication Reichart et al. (2022). Reprinted with permission from AAAS.