Article

Integrative Cluster Analysis of Whole Hearts Reveals Proliferative Cardiomyocytes in Adult Mice

Anne-Marie Galow [1],[†] , Markus Wolfien [2],[†] , Paula Müller [3],[4] , Madeleine Bartsch [3],[4] , Ronald M. Brunner [1] , Andreas Hoeflich [1] , Olaf Wolkenhauer [2],[5],[*] , Robert David [3],[4],[*] and Tom Goldammer [1],[6],[*]

[1]   Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN),
18196 Dummerstorf, Germany; galow@fbn-dummerstorf.de (A.-M.G.);
brunner@fbn-dummerstorf.de (R.M.B.); hoeflich@fbn-dummerstorf.de (A.H.)

[2]   Department of Systems Biology and Bioinformatics, University of Rostock, 18051 Rostock, Germany;
markus.wolfien@uni-rostock.de

[3]   Reference and Translation Center for Cardiac Stem Cell therapy (RTC), Department of Cardiac Surgery,
Rostock University Medical Center, 18057 Rostock, Germany; paula.mueller@uni-rostock.de (P.M.);
madeleine.bartsch@med.uni-rostock.de (M.B.)

[4]   Department of Life, Light, and Matter of the Interdisciplinary Faculty at Rostock University,
18059 Rostock, Germany

[5]   Stellenbosch Institute of Advanced Study, Wallenberg Research Centre, Stellenbosch University,
7602 Stellenbosch, South Africa

[6]   Molecular Biology and Fish Genetics, Faculty of Agriculture and Environmental Sciences,
University of Rostock, 18059 Rostock, Germany

[*]   Correspondence: olaf.wolkenhauer@uni-rostock.de (O.W.); robert.david@med.uni-rostock.de (R.D.);
tom.goldammer@uni-rostock.de (T.G.)

[†]   These authors contributed equally to this work.

check for updates

Abstract: The recent development and broad application of sequencing techniques at the single-cell level is generating an unprecedented amount of data. The different techniques have their individual limits, but the datasets also offer unexpected possibilities when utilized collectively. Here, we applied snRNA-seq in whole adult murine hearts from an inbred (C57BL/6NRj) and an outbred (Fzt:DU) mouse strain to directly compare the data with the publicly available scRNA-seq data of the tabula muris project. Explicitly choosing a single-nucleus approach allowed us to pin down the typical heart tissue-specific technical bias, coming up with novel insights on the mammalian heart cell composition. For our integrated dataset, cardiomyocytes, fibroblasts, and endothelial cells constituted the three main cell populations accounting for about 75% of all cells. However, their numbers severely differed between the individual datasets, with cardiomyocyte proportions ranging from about 9% in the tabula muris data to around 23% for our BL6 data, representing the prime example for cell capture technique related bias when using a conventional single-cell approach for these large cells. Most strikingly in our comparison was the discovery of a minor population of cardiomyocytes characterized by proliferation markers that could not be identified by analyzing the datasets individually. It is now widely accepted that the heart has an, albeit very restricted, regenerative potential. However there is still an ongoing debate where new cardiomyocytes arise from. Our findings support the idea that the renewal of the cardiomyocyte pool is driven by cytokinesis of resident cardiomyocytes rather than differentiation of progenitor cells. We thus provide data that can contribute to an understanding of heart cell regeneration, which is a prerequisite for future applications to enhance the process of heart repair.

1. Introduction

The mammalian heart is a complex organ, composed of various cell types such as cardiomyocytes, fibroblasts, endothelial cells, smooth muscle cells, blood cells, and others [1]. Many of them display significant phenotypic, functional, and genetic heterogeneity even among the same basic cell type suggesting that their individual characteristics are influenced by a variety of factors [2]. Distinct subpopulations may be crucial for the complex biological functions of specific structures within the heart.

Genome-wide transcriptome analysis has substantially improved our understanding of the regulatory networks underlying basic cardiac biology and pathophysiological processes [3–6]. However, the resolution of gene expression patterns in the adult heart has long been limited to the level of whole tissue homogenates ("bulk"), thereby inevitably going along with loss of any information on cellular origin and on cell-type-specific changes in gene expression [7].

Recent developments in RNA amplification strategies allowed the reduction of input RNA to minimal amounts, enabling genome-wide sequencing even on the single-cell level. Single-cell RNA sequencing (scRNA-seq) offers the ability to investigate transcriptomic profiles of individual cells, providing information on prevalence, heterogeneity, and gene co-expression [8]. Along with the expansion of scRNA-seq technologies and applications with increasing sensitivity, accuracy, and throughput, new bioinformatics analysis tools (e.g., Seurat, Harmony, MTGO-SC) were developed and tailored to the specific needs [9,10].

The advent of scRNA-seq paved the way for comprehensive studies on heart biology without reliance on the cell surface or genetic lineage tracing markers. Such markers are usually imperfect and may conceal heterogeneity that exists, even within populations of similar cell types [8]. The usage of transcriptome data based on the global gene expression of individual cells renders the classification of cell types particularly quantitative and data-driven, allowing to unravel such heterogeneities, reveal novel cell populations, and identify cellular hierarchies [11–14].

The computational grouping of cells with respect to their subtypes or states is achieved by using unsupervised clustering methods such as Uniform Manifold Approximation and Projection (UMAP). The comprehensive information facilitates the discovery of novel markers defining a tissue or cell type in a hypothesis-free manner as shown in experiments disclosing the cellular composition of complex tissues [14–16]. Single-cell transcriptomics provides a global and unbiased insight in the cellular composition of tissues and interaction of distinct cell types, both at steady-state and in dynamic processes such as development and regeneration [8].

Recently, the Tabula Muris Consortium analyzed several organs from mice using single-cell sequencing to generate a dataset controlled for age, environment, and epigenetic effects to compile a compendium of cell types, referred to as a "tabula muris" [17]. This enabled the direct comparison of cell-type composition between organs and the comparison of shared cell types across them. The compendium comprises single-cell transcriptomic data from 100,605 cells isolated from 20 organs of C57BL/6JN mice including whole hearts [17].

However, the need for viable intact single cells can be problematic for solid organs or tissues that have highly interconnected processes. The unique difficulty of dissociating the adult mammalian heart tissue without damaging constituent cells, particularly cardiomyocytes, poses an immediate technical hurdle [18]. This, in addition to their large and irregular shape, renders most high throughput single-cell techniques vulnerable to severe technical bias [19].

To enable single-cell analysis also of large cells or cells with complex morphology, robust high throughput single-nucleus RNA sequencing (snRNA-seq) methods have been developed

recently [1,2,20–24]. Although snRNA-seq usually detects a lower number of RNA molecules than scRNA-seq, as it excludes transcripts in the cytoplasm, it holds several technical advantages when applied to tissues difficult to dissociate, such as the mammalian heart.

First, snRNA-seq bypasses the preparation of intact single-cell suspensions and thus can be used to study frozen and fixed primary tissues [20,25]. Second, snRNA-seq can minimize the bias of recovering only cells that are easily dissociated or cells of a certain size optimal for platform-specific cell capture. Moreover, it can reduce the impact of aberrant transcriptional changes induced by enzymatic dissociation [1,26]. Although differences in type and proportion of RNA between the cytosol and nucleus do exist [23], it has been shown that nuclear transcriptomes are representative of the whole cell [21,25,26].

However, there is a particularity of the heart tissue that needs to be taken into account when relying on single-nucleus sequencing. In adult mice, ~85–90% of the cardiomyocytes are binucleated [27], so that the number of nuclei does not directly reflect the number of cells. Moreover, this binucleation has implications for cell proliferation. Cardiomyocyte polyploidization was identified as a barrier to heart regeneration [28] and it was demonstrated that binucleated cardiomyocytes that reentered the cell cycle fail to complete cytokinesis [29], rendering the regeneration of heart tissue after injury deficient.

For biomedical studies, inbred mouse lines such as C57BL/6 are commonly used. However, since human populations are genetically diverse, inbred lines may only partially contribute to an understanding of the molecular basis of a given healthy or diseased phenotype. By contrast, results from outbred mice are based on a broader genetic background and thus could be considered as more robust [30,31]. Therefore, we compare the tabula muris data not only with single-nuclei data of the respective inbred strain (C57BL/6NRj) but also with an outbred strain (Fzt:DU), to assess the extent of the technical bias and reveal biological differences in an integrative analysis.

2. Materials and Methods

2.1. Animals

For the tabula muris project, five hearts from 10–15 weeks old C57BL/6JN mice (three female and two male) had been used as described from the Tabula Muris Consortium [17].

For our approach, ten 10-week old male C57BL/6NRj were shipped from Janvier Labs to the central laboratory animal husbandry of the University Medicine Rostock, where they were euthanized by cervical dislocation the next day.

Fzt:DU mice were bred inhouse by the Service Group Lab Animal Facility. Four 12-week old male mice were euthanized for the extraction of the hearts. All mice were handled in accordance with Directive 2010/63/EU on the protection of animals. A notification about the killing of vertebrates for scientific purposes as per § 4 (3) TSchG was sent to the institutional animal welfare officer according to the German animal protection laws.

2.2. Isolation of Single Cells and Single Nuclei

A detailed description of the heart cell isolation procedure for the tabula muris can be found in their supplementary information of the online version under "A detailed discussion of organ cell types" [17].

Single nuclei were isolated as previously described [11]. In brief, whole hearts were harvested whereby the pulmonary artery and vena cava were removed as far as possible. The hearts were then pooled for each strain before using the Nuclei PURE Prep nuclei isolation kit (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's instructions. For this, tissue samples were rinsed with ice-cold PBS and minced thoroughly with a scalpel. Remaining blood was removed by rinsing with PBS before the tissue chunks were further homogenized in chilled lysis solution using a gentle MACS dissociator. Lysate samples were mixed with chilled sucrose cushion solution and layered on 1.8 M sucrose cushion solution for density gradient centrifugation for 45 min at 30.000× *g* and 4 °C. Nuclei pellets were resuspended in chilled PBS containing 1% BSA and 0.2 U/µL RNase inhibitor and cell

debris were removed by using 40 μm Flowmi cell strainers. After another centrifugation for 8 min at 600× *g* and 4 °C, nuclei were resuspended in Nuclei PURE storage buffer, snap-frozen in liquid nitrogen, and stored at −80 °C until processing.

2.3. Single-Cell and Single-Nucleus Sequencing

Single-cell sequencing for the tabula muris project was previously described [17]. In brief, single cells were captured in droplet emulsions using the GemCode Single-Cell Instrument (10x Genomics), and scRNA-seq libraries were constructed as per the 10x Genomics protocol using GemCode Single-Cell 3′ Gel Bead and Library V2 Kit. The samples were diluted in PBS with 2% FBS to a concentration of 1.000 cells per μL. Cells were loaded in each channel with a target output of 5.000 cells per sample. All reactions were performed in the BioRad C1000 Touch Thermal cycler with 96-Deep Well Reaction Module. Amplified cDNA and final libraries were evaluated on a Fragment Analyzer using a High Sensitivity NGS Analysis Kit (Advanced Analytical). Equal volumes of 16 libraries were pooled for sequencing on the NovaSeq 6000 Sequencing System (Illumina).

Sequencing of Fzt:DU and C57BL/6NRj samples were conducted by Genewiz (GENEWIZ Germany GmbH, Leipzig, Germany). Similar to the tabula muris project, single nuclei were captured in droplet emulsions on the 10xGenomics system and sequenced on the NovaSeq 6000 Sequencing System (Illumina, San Diego, CA, USA). In contrast to the tabula muris sequencing, cells were loaded with a target output of 10,000 cells per sample and the snRNA-seq libraries were constructed using Library V3 chemistry.

2.4. Computational Data Analysis

Typical data processing of scRNA-seq involves quality control, normalization, confounding factor identification, dimensionality reduction, and cell-gene level analysis [32]. Preprocessing of the raw data was conducted by using the CellRanger Software (v.3.1.0) provided by 10x Genomics. The snRNA-seq fastq data files were aligned with STAR [33] (v.2.7) to the mm10 genome (Ensembl release 93) index, annotated via GTF file and grouped by barcodes and UMIs resulting in a feature-barcode matrix. Downstream analysis was performed using Seurat [34] (v.3.1.1). After following the standard pipeline of normalization, finding variable features, scaling, and dimensionality reduction by principal-component (PC) analysis, the datasets were merged in one Seurat object to correct for batch effects and allow for an integrative analysis with the upstream processing algorithm Harmony [35] (v.1.0). The Harmony correction procedure for the newly calculated embedding's iteratively uses its original value instead of the corrected value to regress out confounder effects. Based on this approximation, the embedding correction is restricted to a linear model of the original embedding, which results in a more robust normalization [35]. The integrated dataset was then used for UMAP clustering utilizing the formerly generated Harmony embeddings.

To assign the underlying cell types of the generated clusters, we utilized several approaches accounting for the complexity of the dataset. Sets of well-known marker genes, as well as novel cell cluster markers recently identified by other groups working with single-nuclei data [22], were applied as indicated in our provided computational script. In addition, the top 100 transcripts per cluster and the identified cell cluster markers from our dataset served as reference points for the full characterization of all clusters. The experimental protocol, computational script, top 100 transcripts per cluster, and identified cell cluster markers can be retrieved from our publicly available iRhythmics FairdomHub instance (https://fairdomhub.org/studies/739). Raw snRNA-seq data is available in the Single-Cell Expression Atlas via array express (E-MTAB-8751, E-MTAB-8848).

The used raw datasets served as a basis for three independent studies that successively build upon each other:

#1　　Study: Heart cell type composition in entire adult mammalian hearts [11]
#2　　Study: Strain-dependent differences in cell-type composition and velocity [36]
#3　　Study: Discovery of proliferative cardiomyocytes by integrative cluster analysis (this study)

3. Results

Before data integration, the single-cell data from the tabula muris project were deviating, whereas both single-nuclei datasets appeared rather similar when looking at the first two principal components (PC) (Figure 1A). This naturally differing expression is supported by the violin plot where each point is a cell in which the specific position is based on the cell embeddings determined by the PC dimension reduction technique (Figure 1B). Dataset integration with Harmony resulted in compensation of the primarily technique-based batch effects and allowed for a direct comparison of the different datasets (Figure 1C,D). Although the datasets differed not only in the technique used to generate them but also in the detected number of features, it was possible to achieve a well-suited integrated embedding for subsequent downstream analysis.
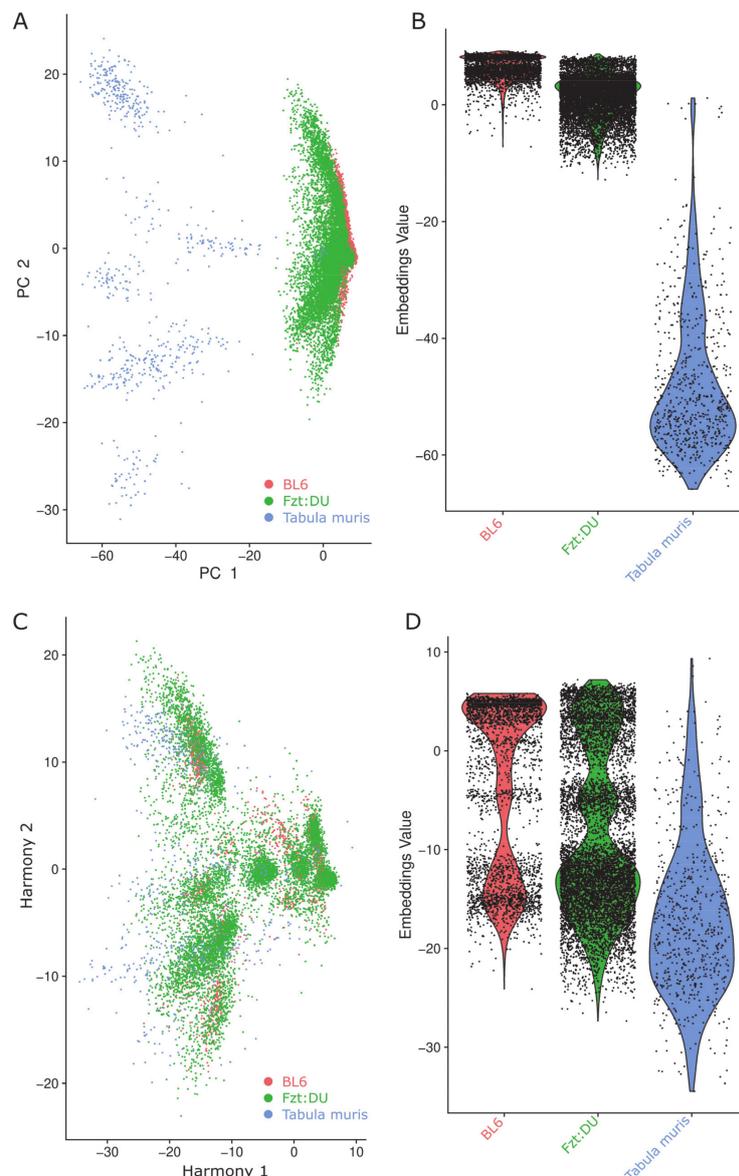

Figure 1. Clustering effects of the data integration with and without utilizing Harmony on the embedding in the first two principal components (PC) on the indicated datasets (BL6, red; Fzt:DU, green; Tabula muris, blue). (**A**) PCA-plot of the three investigated datasets without Harmony processing. (**B**) Violin plot of the investigated datasets without Harmony processing. (**C**) PCA-plot of the three investigated datasets with Harmony processing. (**D**) Violin plot of the three investigated datasets with Harmony processing.

The integrated analysis included a total of 11,672 nuclei and 628 cells representing 20,258 genes in which each cell exhibits an average total expression of 1,199 genes per cell. For the UMAP-clustering, cells were colored by their initial dataset origin to visualize the contribution of different data sets to the various clusters (Figure 2A). Some clusters contained no cells of a certain dataset, other cells appeared to be unique to one dataset, but most clusters were well mixed. The analysis identified 21 distinct clusters, of which 16 were shared by all datasets (see supplementary material provided at the iRhythmics FairdomHub instance under https://fairdomhub.org/studies/739).
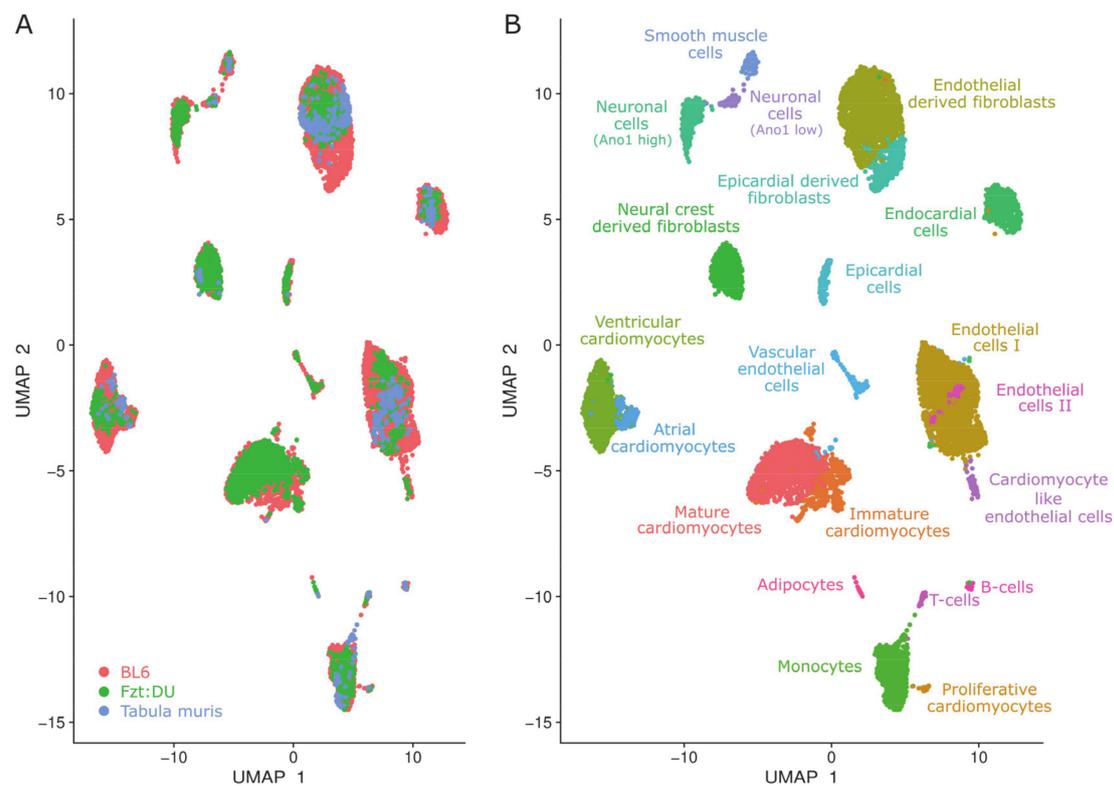
Figure 2. Uniform Manifold Approximation and Projection (UMAP) clustering of the integrated datasets representing the identified cellular clusters that are colored by data origin and cell type. (**A**) snRNA-seq data from whole Fzt:DU (green, 8533 nuclei) and BL6 (red, 3139 nuclei) mouse hearts were integrated with scRNA-seq data from hearts of BL6 mice of the tabula muris project (blue, 628 cells) to attain well-mixed clusters for a common downstream analysis. (**B**) 21 clusters were identified and annotated to specific cell types.

The top five gene markers for all clusters were visualized in a dot plot (Figure 3). An extended visualization of the top ten gene markers per cluster and the top 100 transcripts per cluster are provided online at iRhythmics FairdomHub together with the experimental protocol and the computational script (https://fairdomhub.org/studies/739).
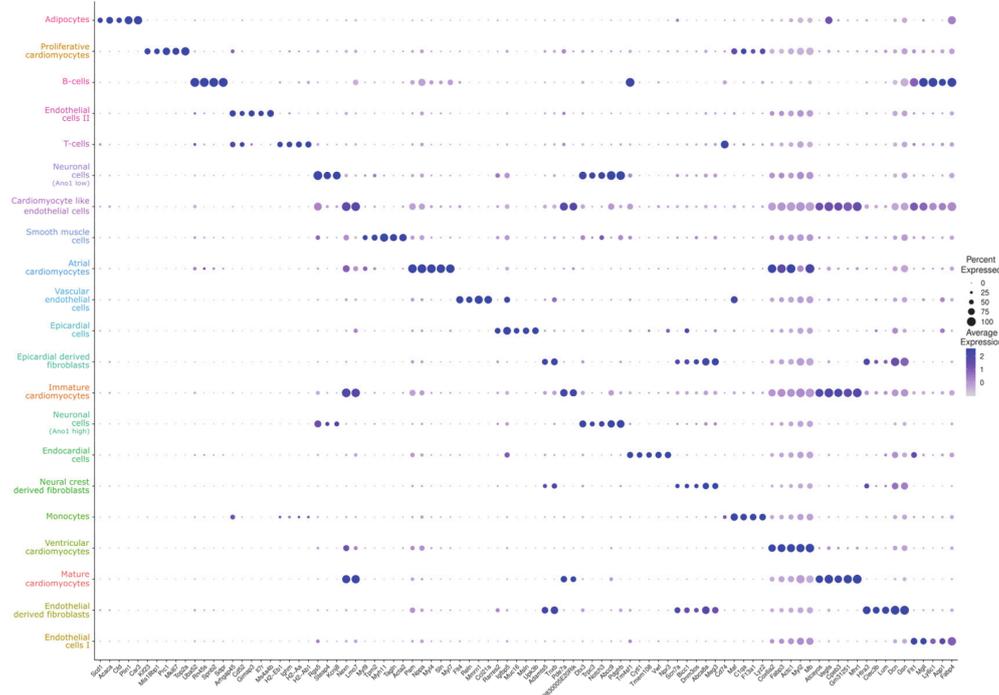
Figure 3. Dot plot of the top five markers for each identified cluster. For each cluster, the relative gene expression in percent is represented by the size of dots, e.g., a value of 100 means that each cell within this cell type expressed this gene. The average expression level for the listed genes is indicated by the color.

In the integrated dataset, the main cell types comprise populations of cardiomyocytes (30.0%), fibroblasts (26.2%), and endothelial cells (20.3%) containing ~3700, ~3200, and ~2500 nuclei/cells respectively (Figure 2B). However, their numbers severely differed between the individual datasets. For example, cardiomyocyte proportions range from only about 9% in the tabula muris data to around 23% for the BL6 data, assuming a binucleation rate of 85–90%.

Cardiomyocyte populations were identified by classical marker genes such as Actn2, Tnnc1, and Actc1. Cardiomyocytes that were characterized by high levels of markers for a mature sarcomere structure such as Myomegalin and cardiac myosin binding protein c as well as markers for contractile function such as Ryr2 but could not be clearly assigned to a ventricular or atrial type, were named "mature cardiomyocytes." Cardiomyocytes characterized by low levels of these genes were considered "immature cardiomyocytes."

When exploring the subpopulations, we failed to detect mature cardiomyocytes in the tabula muris dataset. However, we found some differentiated ventricular and atrial cardiomyocytes with the atrial cells constituting the major population. In contrast, BL6 showed the highest proportion of mature cardiomyocytes, but less could be assigned to clearly ventricular or atrial origin when compared with the Fzt:DU data.

In total, Fzt:DU have fewer cardiomyocytes (15.5%, considering binucleation) than BL6, but remarkably an endothelial cell population was found that shows markers clearly related to cardiomyocyte function (e.g., Gja1, Atp2a2, Ttn, Ryr2, and Myh6) suggesting a trans-differentiation process from an endothelial cell-like phenotype towards a cardiomyocyte-like phenotype in the Fzt:DU hearts [11].

We discovered an additional small population of cells (0.4%; 48 cells/nuclei) located next to the immune cell populations in the UMAP embedding, suggesting related features (Figure 2B). Indeed, cells of this small population share some common markers with monocytes such as Maf, C1qa, and Lyz2 (Figure 3). However, this population also clearly displays cardiomyocyte markers in comparable expression levels to other cardiomyocyte populations (Figure 4A–C).

Surprisingly, this population is characterized by markers related to mitosis and cytokinesis, such as Marker of Proliferation Ki-67 (Mki67) as well as the kinesin-like motor proteins Centrosome-associated protein E (Cenpe) and Kinesin Family Member 23 (Kif23), that both accumulate in the G2 phase of the cell cycle. All other cardiomyocyte populations identified in this study do not express any of these markers (Figure 4D–F).
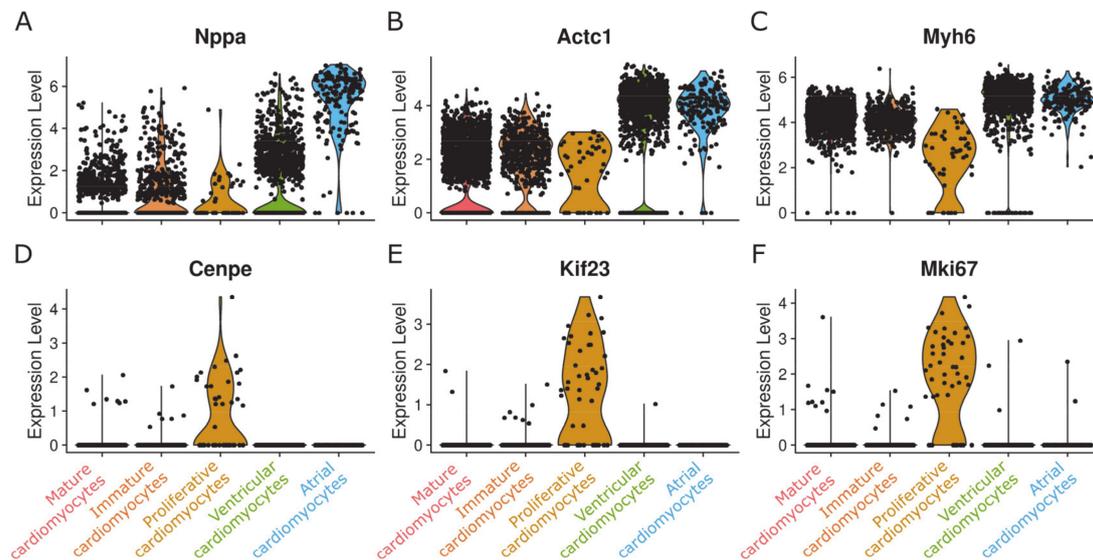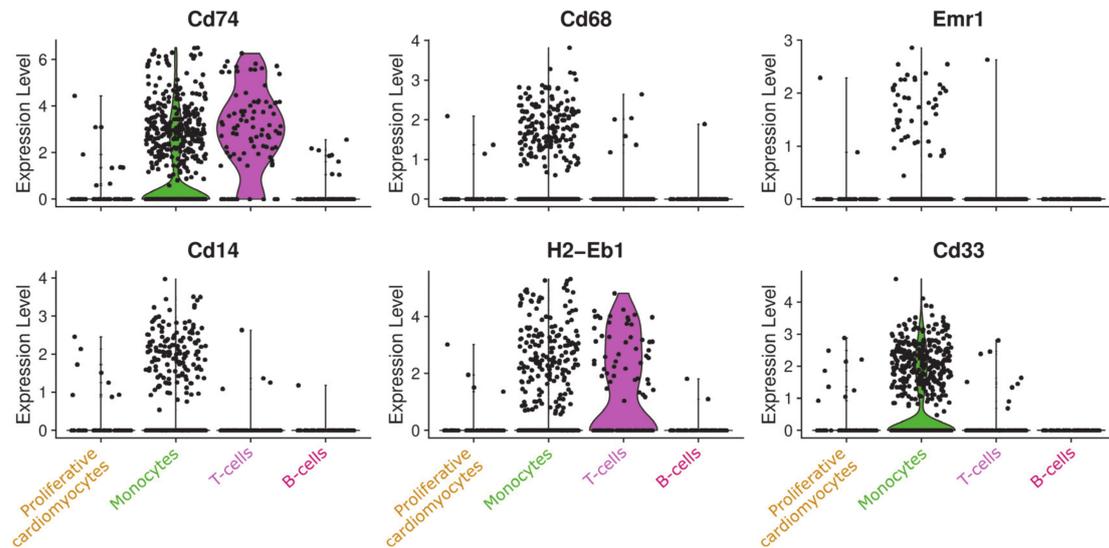

Figure 4. Violin plot of cardiomyocyte cell markers and markers for proliferation. (**A–C**) Three typical cardiomyocyte markers are expressed in comparable levels in all clusters considered as cardiomyocyte populations. (**D–F**) Markers for proliferation and cytokinesis are only expressed in one of the clusters, resulting in the annotation as "proliferative cardiomyocytes".

To confirm that this population does not merely undergo mitosis resulting in binucleation instead of cytokinesis, we checked for Sept7, Anln, and Aurkb, all of which encode proteins involved in the assembly of the cleavage furrow during cytokinesis [37,38]. In fact, all of these genes were expressed by the bona fide "proliferative cardiomyocytes" (Figure 5). In contrast, Anln and Aurkb expression was almost not detectable in the other cardiomyocyte populations (average expression level <0.1), and Sept7 was expressed only in low levels from immature cardiomyocytes and atrial cardiomyocytes (Figure 5).
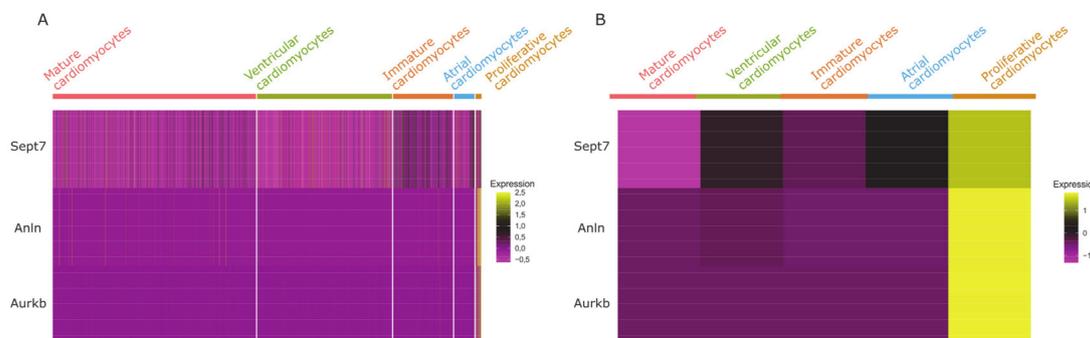
Figure 5. Heatmap of cytokinesis marker expression in cardiomyocyte clusters. (**A**) Expression levels of three factors involved in cleavage furrow assembly during cytokinesis are shown for each individual cell in the cardiomyocyte clusters. (**B**) Expression of cytokinesis markers averaged for each cardiomyocyte cluster shows the result more clearly, demonstrating the notable expression of cytokinesis markers exclusively in the "proliferative cardiomyocyte cluster".

Because of the similarities between this cluster and the monocyte population, noticeable in Figure 3, and the fact that cardiomyocytes are usually considered postmitotic [39], we needed to fully rule out that this cluster is, in fact, a proliferative monoblast population. To this end, we investigated a variety of well-known monocyte and monoblast markers (Cd74, Cd68, Emr1, Cd14, H2-Eb1, and Cd33) and found the cells of the respective cluster to be negative for all of them as visualized as violin plot in Figure 6.


Figure 6. Violin plot of monoblast markers in the proliferating cardiomyocyte population and immune cells. Proliferative cardiomyocytes show no significant expression of the indicated typical monoblast markers.

To explore whether this population might reflect resident cardiac stem cells or another progenitor cell type, we checked for typical stem cell markers. While we could not detect any c-kit or Isl1-positive cells in our dataset, we did find cells expressing Cd34, Pdgfra, or Ly6a/Sca1 (Figure 7). However, these cells are mainly members of the immature cardiomyocyte cluster.

When examining Cd34 more closely, we found the average expression of this relatively unspecific marker to be comparable or even higher in other cell clusters such as the fibroblast or the endothelial cell populations, respectively (supplementary data provided online at iRhythmics FairdomHub, https://fairdomhub.org/studies/739).

For snRNA-seq there is no possibility for a viability test before sequencing due to the isolation protocol. However, in an attempt to assess the potential impact of apoptosis on our results, we checked for the expression of apoptosis-related proteins such as those of the Bcl-2 family and inhibitors of apoptosis proteins (IAPs). We could not observe any remarkably increased expression of pro-apoptotic proteins in our data. When investigating the cardiomyocyte populations, we found no pro-apoptotic p53 or Bak and levels of the anti-apoptotic Bcl-xL and Mcl1 were comparable. We observed a slightly elevated Bax expression (average expression = 1.42) in the proliferative cardiomyocytes in comparison to average expression of around 0.35 in the other cardiomyocyte populations. However, expression levels of the antiapoptotic Bcl-2 were a seven-fold increase in proliferative cardiomyocytes (average expression = 2.0 compared to around 0.29 in other cardiomyocyte populations) and expression of the IAP survivin was even 100-fold increased (average expression = 2.23 compared to around 0.02 in other cardiomyocyte populations).
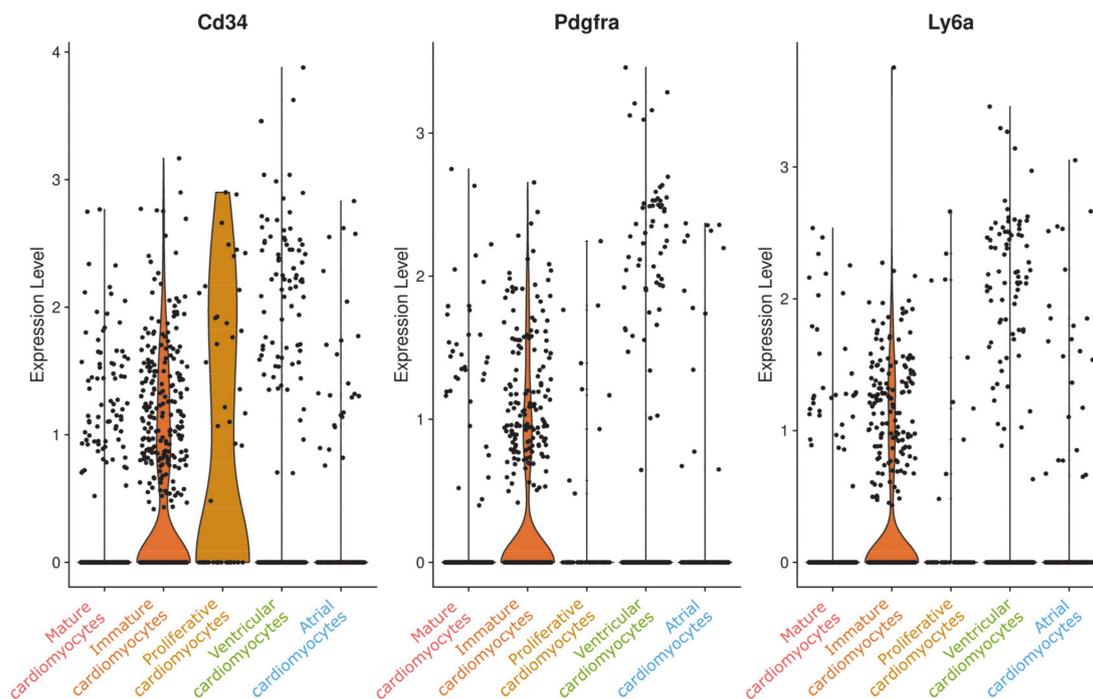
Figure 7. Violin plot of stem cell markers in the cardiomyocyte populations. Typical stem cell markers are expressed in low levels in immature cardiomyocytes, but not in "proliferative cardiomyocytes" except for Cd34, which showed moderate expression in several cell clusters of the integrated dataset.

4. Discussion

This study underlines how significantly the applied technology affects the outcome of cluster analyses and how technical limitations regarding cell capture techniques may lead to an underrepresentation of individual cell types such as cardiomyocytes due to their large cell size and irregular shape [19]. The Tabula Muris Consortium analyzed several organs from mice using single-cell sequencing to compile a compendium of cell types, referred to as a "Tabula Muris," knowing that cardiomyocytes will be underrepresented in their heart data [17]. Here, we directly compared their single-cell RNA sequencing data with our single-nuclei sequencing approach, especially better suited for cardiomyocytes, to likewise assess the extent of technical bias and investigate the effect of such integrative cells to nuclei clustering towards cell type identification. The generated comprehensive information also facilitates the discovery of novel markers defining a tissue or cell type in a hypothesis-free manner, refines cell type subpopulations, and validates existing markers obtained from other tissues [14–16].

It has previously been shown that single-nuclei techniques have several advantages over single-cell sequencing, while still retaining sufficient information for cluster analysis and matching results from whole-cell RNA in neuronal cells [21]. In line with these findings, our data accurately represent the cellular composition of the adult murine heart by relying on single-nucleus RNA sequencing. Our integrative analysis of the two single nuclei and one single-cell sequencing samples required the computational integration of these datasets via Harmony. This was a first essential processing step because the datasets were assayed with two different, yet comparable approaches in which actual biological differences might be interspersed with technical differences [23,40]. The results shown in Figure 1 highlight the necessity of such an approach, because after the Harmony processing the dimension of the embeddings values was greatly improved and aligned for a proper downstream comparison.

Considerable differences became apparent after assigning the clusters to cell types and subpopulations. The most frequent cell types were cardiomyocytes, fibroblasts, and endothelial cells, but their cell numbers differed severely between the individual datasets. For example, cardiomyocyte

proportions ranged from about 9% in the tabula muris data to around 23% for the BL6 data, when taking binucleation into account. Not only do total cell numbers differ, but there is also a clear overrepresentation of atrial cardiomyocytes in the tabula muris data. We attribute this to their smaller size allowing them to pass through the microfluidics easier than e.g., ventricular cardiomyocytes, thereby representing the prime example for cell capture techniques related bias.

On the other hand, our BL6 data showed the highest proportion of mature cardiomyocytes, but less could be assigned to clearly ventricular or atrial origin when compared with the tabula muris data. The reason for this is not completely understood, but a deficiency of snRNA-seq when it comes to the specification of subpopulations cannot be excluded and even seems reasonable taking into account the limitations introduced by exclusively working with nuclear RNA.

Beyond comparison, we demonstrated that the integration of datasets results in an even more informative pool enabling deeper insights also with respect to the underlying individual data sets. Interestingly, we were able to identify a minor cell population (0.4%) in the integrated dataset that could not be detected when analyzing the datasets individually. For example, only two cells of the tabula muris dataset were identified when checking for the sample origins in this cluster, thus not being sufficient to constitute an own cell cluster in the original dataset. However, the pooling of such small populations from several datasets allows for their detection and exploration after all.

Cells of the newly detected cluster are expressing cardiomyocyte markers, but do not cluster together with other cardiomyocyte populations. Although the cells are located next to the immune cell populations in the UMAP embedding and share some common features with monocytes, an affiliation to them could be excluded due to the absence of Cd33, Cd74 (Figure 6), and other typical immune cell markers.

Strikingly, this population is characterized by markers related to mitosis and cytokinesis such as marker of proliferation Ki-67 (Mki67), centrosome-associated protein E (Cenpe) and kinesin family member 23 (Kif23). All other cardiomyocyte populations identified in this study do not express any of those markers, as would be expected from cardiomyocytes, usually considered as being post-mitotic [39].

Mki67 is commonly used to detect and quantify proliferating cells since it stabilizes individual mitotic chromosomes during all phases of mitosis. CENPE is a kinesin-like motor protein that accumulates in the G2 phase of the cell cycle and Kif 23 is a component of the centralspindlin complex required for the myosin contractile ring formation during cytokinesis. CENPE and Kif23 are both markers for late phases of mitosis and cytokinesis, but they are also present at earlier stages [37,38] and therefore, as with Mk67, there are no sufficient markers for complete cell division.

It is known that cardiomyocytes can undergo mitosis, which usually results in binucleation instead of a complete cell division. Because the aforementioned genes are markers for both mitosis and cytokinesis, we checked for additional, more specific genes.

During complete cytokinesis, a contractile ring assembles, which includes a network of actin and myosin filaments as well as septins and anillin. The cleavage furrow is formed under the regulation of several kinases such as Plk1 and Aurora B thereby acting as key regulators of cytokinesis [37,38]. Genes for septin7 and anillin as well as the genes for the mitotic kinases are expressed in considerable amounts by the cells of this cluster, thereby confirming their proliferative status.

When checking for apoptosis-related factors to evaluate the impact of potential cell death during the isolation process, we found the IAP survivin to be severely upregulated in this proliferative cell cluster. Knockout of survivin was shown to lower mitotic rate without increased apoptosis and survivin-deficient cardiomyocytes displayed marked DNA polyploidy indicative of consecutive rounds of DNA replication without cell division [41]. In contrast, overexpression of survivin rescued the cell-cycle defect in cardiomyocytes in spinal muscular atrophy mouse models as seen by the proportions of cells in the G0/G1, S, and G2/M phases of the cell cycle [42]. These data indicate that survivin is not only an antiapoptotic factor but also a critical regulator of cell division, particularly in cardiomyocytes. The increased expression of survivin in the small cardiomyocyte cluster identified in our data further

strengthens the belief that these cardiomyocytes are able to undergo full cytokinesis. Vice versa the transcriptome data underline the postulated role of survivin in controlling cardiomyocyte numbers.

To explore whether this population might, in fact, represent a cardiac stem cell population, we checked for typical stem cell markers such as c-kit, Isl1, Cd34, Pdgfra, or Ly6a/Sca1 (Figure 6). Except for Cd34 there was no relevant expression of any for these genes and inspecting Cd34 more closely, we found that the average expression was comparable or even higher in other cell clusters such as the fibroblasts populations or the endothelial cell populations, respectively. These data reveal no evidence for stem cell-like characteristics in proliferative cardiomyocytes.

It is now widely accepted that the heart has a regenerative potential [43]. However, it is still debated where new cardiomyocytes arise from. Our data suggest that there are cardiomyocytes that undergo cytokinesis, which is why we termed the identified cluster "proliferative cardiomyocytes." In accordance with other studies relying on different techniques, we propose that new cardiomyocytes in adult hearts mainly arise from differentiated cardiomyocytes rather than from resident stem cells [29,44,45]. A graphical illustration of our results can be viewed in Figure 8.
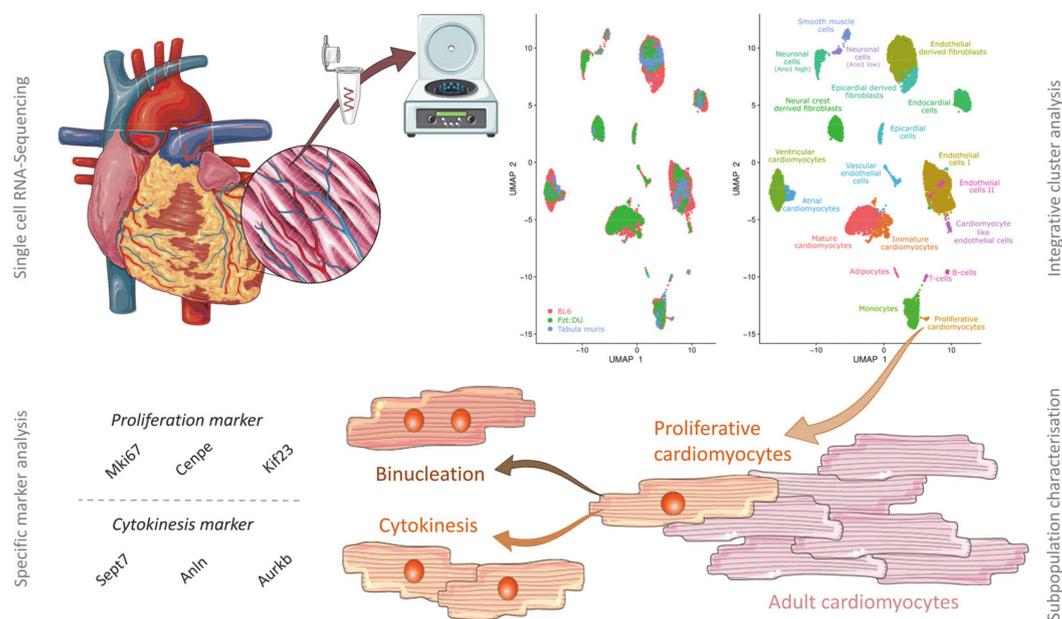
Figure 8. Graphical illustration of the workflow for identification of *bona fide* proliferative cardiomyocytes. Single-cell and single-nucleus sequencing data of whole adult murine hearts from inbred (C57BL/6NRj) and outbred (Fzt:DU) mouse strains were computationally integrated via Harmony. The integrated dataset allowed for novel insights on the mammalian heart cell composition revealing a minor population of cardiomyocytes characterized by proliferation markers and therefore referred to as "proliferative cardiomyocytes." During in-depth gene expression analysis, we found additional markers specific for cytokinesis to be upregulated, suggesting that proliferative cardiomyocytes not only undergo binucleation but also complete cytokinesis. We did not detect relevant expression of any stem cell marker, thus supporting the idea that the renewal of the cardiomyocyte pool is driven by cytokinesis of resident cardiomyocytes rather than differentiation of stem cells.

Furthermore, our study provides information on the frequency of these potentially dividing cells. In comparison with other estimations in humans [44,46] the number of proliferative cardiomyocytes seems to be quite high (0.4% of all nuclei and 1.3% of all cardiomyocyte nuclei). Indeed, not all cells of the "proliferative cardiomyocytes cluster" express all of the proliferation markers. Therefore, we cannot exclude that some cells in this population have re-entered the cell cycle but did not complete karyokinesis, and cytokinesis.

Considering 50% of the cluster (compare dot size in (Figure 3)) as cardiomyocytes truly re-entering the cell cycle and approximately 50% of them completing cytokinesis, as implied by data of Bersell et al. [29], gives a more realistic approximation in our data of around 0.3% of cardiomyocyte nuclei being derived from cells which undergo complete cell division.

Although snRNA-seq bypasses the cell capture technique related bias, it still has one drawback with respect to cardiomyocytes in particular: We cannot distinguish between mononucleated and binucleated cardiomyocytes in our analysis, thus precluding the direct calculation of cardiomyocyte cell numbers and constraining us to base our evaluation on previously determined binucleation rates.

However, for proliferative cardiomyocytes there are contradicting studies about the frequency of dividing binuclear cells. Whereas most studies suggest that proliferating cardiomyocytes were predominantly mononucleated and only mononucleated cardiomyocytes could complete cytokinesis [27,29], there are also studies claiming that mono- and bi-nucleated cardiomyocytes are in fact almost equally able to proliferate [45]. In the absence of fully verified data on the proportion of mono- and bi-nuclear cells capable to proliferate, we cannot draw an exact conclusion on the actual cell count of our identified proliferative cardiomyocyte subpopulation based on the nuclei numbers.

There are highly sensitive plate-based approaches such as Smart-Seq2 [47] and CEL-Seq2 [48] that may allow for microscopic exploration of cells before the sequencing process. This would enable visual control of binucleation on the one hand and cell division processes on the other hand. However, these are low throughput systems and due to the very rare occurrence of proliferative cardiomyocytes, it would be subject to chance whether a proliferative cell could be detected per plate at all, thus rendering plate-based systems rather unsuitable for the identification of such rare cell species at present.

In a systematic comparison of single-cell RNA-sequencing methods, it was found that the high throughput system 10x Chromium had the best, consistent performance [49], which is why we relied on this technique. It allowed us to explore the distribution and features of cell populations in entire mammalian hearts. With the ongoing development of single-cell sequencing technologies, it will in the future certainly be possible to inspect cells also in high throughput systems visually thereby enabling even more comprehensive data.

5. Conclusions

This study illustrates how significantly the applied technology affects the outcome of cluster analyses and demonstrates how the power of cluster analysis can be increased by integrating multiple data sets.

In summary, our data support the hypothesis of other groups postulating that there is a turnover of cardiomyocytes in the postnatal heart. Moreover, our findings suggest that the renewal of the cardiomyocyte pool is driven by cytokinesis of resident cardiomyocytes rather than differentiation of progenitor cells. Our comprehensive sequencing data can contribute to an increased understanding of heart cell biology and regeneration.

References

1. Hu, P.; Liu, J.; Zhao, J.; Wilkins, B.J.; Lupino, K.; Wu, H.; Pei, L. Single-nucleus transcriptomic survey of cell diversity and functional maturation in postnatal mammalian hearts. *Genes Dev.* **2018**, *32*, 1344–1357. [CrossRef] [PubMed]

2. Macosko, E.Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A.R.; Kamitaki, N.; Martersteck, E.M.; et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **2015**, *161*, 1202–1214. [CrossRef] [PubMed]

3. Paige, S.L.; Plonowska, K.; Xu, A.; Wu, S.M. Molecular Regulation of Cardiomyocyte Differentiation. *Circ. Res.* **2015**, *116*, 341. [CrossRef] [PubMed]

4. Potente, M.; Mäkinen, T. Vascular heterogeneity and specialization in development and disease. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 477–494. [CrossRef]

5. Tallquist, M.D.; Molkentin, J.D. Redefining the identity of cardiac fibroblasts. *Nat. Rev. Cardiol.* **2017**, *14*, 484. [CrossRef]

6. Yavari, A.; Bellahcene, M.; Bucchi, A.; Sirenko, S.; Pinter, K.; Herring, N.; Jung, J.J.; Tarasov, K.V.; Sharpe, E.J.; Wolfien, M.; et al. Mammalian γ2 AMPK regulates intrinsic heart rate. *Nat. Commun.* **2017**, *8*, 1–19. [CrossRef]

7. Gladka, M.M.; Molenaar, B.; de Ruiter, H.; van der Elst, S.; Tsui, H.; Versteeg, D.; Lacraz, G.P.A.; Huibers, M.M.H.; van Oudenaarden, A.; van Rooij, E. Single-Cell Sequencing of the Healthy and Diseased Heart Reveals Cytoskeleton-Associated Protein 4 as a New Modulator of Fibroblasts Activation. *Circulation* **2018**, *138*, 166–180. [CrossRef]

8. Massaia, A.; Chaves, P.; Samari, S.; Miragaia, R.J.; Meyer, K.; Teichmann, S.A.; Noseda, M. Single Cell Gene Expression to Understand the Dynamic Architecture of the Heart. *Front. Cardiovasc. Med.* **2018**, *5*, 167. [CrossRef]

9. Picelli, S. Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol.* **2017**, *14*, 637. [CrossRef]

10. Wirka, R.C.; Pjanic, M.; Quertermous, T. Advances in Transcriptomics. *Circ. Res.* **2018**, *122*, 1200–1220. [CrossRef]

11. Wolfien, M.; Galow, M.; Müller, P.; Bartsch, M.; Brunner, R.M.; Goldammer, T.; Wolkenhauer, O.; Hoeflich, A.; David, R. Single-Nucleus Sequencing of an Entire Mammalian Heart: Cell Type Composition and Velocity. *Cells* **2020**, *9*, E318. [CrossRef] [PubMed]

12. Kokkinopoulos, I.; Ishida, H.; Saba, R.; Ruchaya, P.; Cabrera, C.; Struebig, M.; Barnes, M.; Terry, A.; Kaneko, M.; Shintani, Y.; et al. Single-Cell Expression Profiling Reveals a Dynamic State of Cardiac Precursor Cells in the Early Mouse Embryo. *PLoS ONE* **2015**, *10*, e0140831. [CrossRef] [PubMed]

13. Shalek, A.K.; Satija, R.; Shuga, J.; Trombetta, J.J.; Gennert, D.; Lu, D.; Chen, P.; Gertner, R.S.; Gaublomme, J.T.; Yosef, N.; et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **2014**, *510*, 363–369. [CrossRef] [PubMed]

14. Wilson, N.K.; Kent, D.G.; Buettner, F.; Shehata, M.; Macaulay, I.C.; Calero-Nieto, F.J.; Sánchez Castillo, M.; Oedekoven, C.A.; Diamanti, E.; Schulte, R.; et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **2015**, *16*, 712. [CrossRef]

15. Kernfeld, E.M.; Genga, R.M.; Neherin, K.; Magaletta, M.E.; Xu, P.; Maehr, R. A single-cell transcriptomic atlas of thymus organogenesis resolves cell types and developmental maturation. *Immunity* **2018**, *48*, 1258. [CrossRef]

16. Sasagawa, Y.; Nikaido, I.; Hayashi, T.; Danno, H.; Uno, K.D.; Imai, T.; Ueda, H.R. Quartz-Seq: A highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **2013**, *14*, 1–17. [CrossRef]

17. The Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **2018**, *562*, 367–372.

18. Ackers-Johnson, M.; Li, P.Y.; Holmes, A.P.; O'Brien, S.-M.; Pavlovic, D.; Foo, R.S. A Simplified, Langendorff-Free Method for Concomitant Isolation of Viable Cardiac Myocytes and Nonmyocytes From the Adult Mouse Heart. *Circ. Res.* **2016**, *119*, 909–920. [CrossRef]

19. Ackers-Johnson, M.; Tan, W.L.W.; Foo, R.S.-Y. Following hearts, one cell at a time: Recent applications of single-cell RNA sequencing to the understanding of heart disease. *Nat. Commun.* **2018**, *9*, 4434. [CrossRef]

20. Habib, N.; Avraham-Davidi, I.; Basu, A.; Burks, T.; Shekhar, K.; Hofree, M.; Choudhury, S.R.; Aguet, F.; Gelfand, E.; Ardlie, K.; et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **2017**, *14*, 955–958. [CrossRef]

21. Lake, B.B.; Codeluppi, S.; Yung, Y.C.; Gao, D.; Chun, J.; Kharchenko, P.V.; Linnarsson, S.; Zhang, K. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **2017**, *7*, 6031. [CrossRef] [PubMed]

22. Linscheid, N.; Logantha, S.J.R.J.; Poulsen, P.C.; Zhang, S.; Schrölkamp, M.; Egerod, K.L.; Thompson, J.J.; Kitmitto, A.; Galli, G.; Humphries, M.J.; et al. Quantitative proteomics and single-nucleus transcriptomics of the sinus node elucidates the foundation of cardiac pacemaking. *Nat. Commun.* **2019**, *10*, 2889. [CrossRef] [PubMed]

23. Selewa, A.; Dohn, R.; Eckart, H.; Lozano, S.; Xie, B.; Gauchat, E.; Elorbany, R.; Rhodes, K.; Burnett, J.; Gilad, Y.; et al. Systematic Comparison of High-throughput Single-Cell and Single-Nucleus Transcriptomes during Cardiomyocyte Differentiation. *Sci. Rep.* **2020**, *10*, 1535. [CrossRef] [PubMed]

24. Wu, H.; Kirita, Y.; Donnelly, E.L.; Humphreys, B.D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J. Am. Soc. Nephrol.* **2019**, *30*, 23–32. [CrossRef]

25. 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. RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19802–19807. [CrossRef]

26. Zeng, W.; Jiang, S.; Kong, X.; El-Ali, N.; Ball, A.R.; Ma, C.I.-H.; Hashimoto, N.; Yokomori, K.; Mortazavi, A. Single-nucleus RNA-seq of differentiating human myoblasts reveals the extent of fate heterogeneity. *Nucleic Acids Res.* **2016**, *44*, e158. [CrossRef]

27. Soonpaa, M.H.; Kim, K.K.; Pajak, L.; Franklin, M.; Field, L.J. Cardiomyocyte DNA synthesis and binucleation during murine development. *Am. J. Physiol.* **1996**, *271*, H2183–H2189. [CrossRef]

28. González-Rosa, J.M.; Sharpe, M.; Field, D.; Soonpaa, M.H.; Field, L.J.; Burns, C.E.; Burns, C.G. Myocardial Polyploidization Creates a Barrier to Heart Regeneration in Zebrafish. *Dev. Cell* **2018**, *44*, 433.e7–446.e7. [CrossRef]

29. Bersell, K.; Arab, S.; Haring, B.; Kühn, B. Neuregulin1/ErbB4 Signaling Induces Cardiomyocyte Proliferation and Repair of Heart Injury. *Cell* **2009**, *138*, 257–270. [CrossRef]

30. Tuttle, A.H.; Philip, V.M.; Chesler, E.J.; Mogil, J.S. Comparing phenotypic variation between inbred and outbred mice. *Nat. Methods* **2018**, *15*, 994–996. [CrossRef]

31. Wanner, M. The Unexpected Advantages of Outbred Mice in Research. Available online: https://www.jax.org/news-and-insights/2018/december/the-advantages-of-outbred-mice-in-research (accessed on 6 February 2020).

32. Haque, A.; Engel, J.; Teichmann, S.A.; Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **2017**, *9*, 75. [CrossRef] [PubMed]

33. Dobin, A.; Gingeras, T.R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinform.* **2015**, *51*, 11.14.1–11.14.19. [CrossRef] [PubMed]

34. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888.e21–1902.e21. [CrossRef] [PubMed]

35. Korsunsky, I.; Millard, N.; Fan, J.; Slowikowski, K.; Zhang, F.; Wei, K.; Baglaenko, Y.; Brenner, M.; Loh, P.; Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **2019**, *16*, 1289–1296. [CrossRef]

36. Wolfien, M.; Galow, M.; Müller, P.; Bartsch, M.; Brunner, R.M.; Goldammer, T.; Wolkenhauer, O.; Hoeflich, A.; David, R. Single Nuclei Sequencing of entire Mammalian Hearts: Strain-dependent Cell Type Composition and Velocity. *Cardiovasc. Res.* **2020**, cvaa054. [CrossRef]

37. Atilla-Gokcumen, G.E.; Castoreno, A.B.; Sasse, S.; Eggert, U.S. Making the cut: The chemical biology of cytokinesis. *ACS Chem. Biol.* **2010**, *5*, 79–90. [CrossRef]

38. Normand, G.; King, R.W. Understanding Cytokinesis Failure. *Adv. Exp. Med. Biol.* **2010**, *676*, 27–55.

39. Soonpaa, M.H.; Field, L.J. Assessment of cardiomyocyte DNA synthesis in normal and injured adult mouse hearts. *Am. J. Physiol.* **1997**, *272*, H220–H226. [CrossRef]

40. Bakken, T.E.; Hodge, R.D.; Miller, J.A.; Yao, Z.; Nguyen, T.N.; Aevermann, B.; Barkan, E.; Bertagnolli, D.; Casper, T.; Dee, N.; et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS ONE* **2018**, *13*, e0209648. [CrossRef] [PubMed]

41. Levkau, B.; Schäfers, M.; Wohlschlaeger, J.; von Wnuck Lipinski, K.; Keul, P.; Hermann, S.; Kawaguchi, N.; Kirchhof, P.; Fabritz, L.; Stypmann, J.; et al. Survivin Determines Cardiac Function by Controlling Total Cardiomyocyte Number. *Circulation* **2008**, *117*, 1583–1593. [CrossRef]

42. Sheng, L.; Wan, B.; Feng, P.; Sun, J.; Rigo, F.; Bennett, C.F.; Akerman, M.; Krainer, A.R.; Hua, Y. Downregulation of Survivin contributes to cell-cycle arrest during postnatal cardiac development in a severe spinal muscular atrophy mouse model. *Hum. Mol. Genet.* **2018**, *27*, 486–498. [CrossRef] [PubMed]

43. Yester, J.W.; Kühn, B. Mechanisms of Cardiomyocyte Proliferation and Differentiation in Development and Regeneration. *Curr. Cardiol. Rep.* **2017**, *19*, 13. [CrossRef] [PubMed]

44. Lázár, E.; Sadek, H.A.; Bergmann, O. Cardiomyocyte renewal in the human heart: Insights from the fall-out. *Eur. Heart J.* **2017**, *38*, 2333–2342. [CrossRef] [PubMed]

45. Wang, W.E.; Li, L.; Xia, X.; Fu, W.; Liao, Q.; Lan, C.; Yang, D.; Chen, H.; Yue, R.; Zeng, C.; et al. Dedifferentiation, Proliferation, and Redifferentiation of Adult Mammalian Cardiomyocytes After Ischemic Injury. *Circulation* **2017**, *136*, 834–848. [CrossRef]

46. Bergmann, O.; Zdunek, S.; Felker, A.; Salehpour, M.; Alkass, K.; Bernard, S.; Sjostrom, S.L.; Szewczykowska, M.; Jackowska, T.; dos Remedios, C.; et al. Dynamics of Cell Generation and Turnover in the Human Heart. *Cell* **2015**, *161*, 1566–1575. [CrossRef]

47. Picelli, S.; Björklund, Å.K.; Faridani, O.R.; Sagasser, S.; Winberg, G.; Sandberg, R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **2013**, *10*, 1096–1098. [CrossRef]

48. Hashimshony, T.; Senderovich, N.; Avital, G.; Klochendler, A.; de Leeuw, Y.; Anavy, L.; Gennert, D.; Li, S.; Livak, K.J.; Rozenblatt-Rosen, O.; et al. CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **2016**, *17*, 77. [CrossRef]

49. Ding, J.; Adiconis, X.; Simmons, S.K.; Kowalczyk, M.S.; Hession, C.C.; Marjanovic, N.D.; Hughes, T.K.; Wadsworth, M.H.; Burks, T.; Nguyen, L.T.; et al. Systematic comparative analysis of single cell RNA-sequencing methods. *bioRxiv* **2019**, 632216. [CrossRef]