SINGLE CELL RNA-SEQUENCING OF CARDIAC PROGENITOR CELLS ACROSS PATIENT AGE POPULATIONS

A Dissertation
Presented to
The Academic Faculty

by

Arun Jayaraman

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
The Wallace H. Coulter Department of Biomedical Engineering

Georgia Institute of Technology & Emory University
December 2021

SINGLE CELL RNA-SEQUENCING OF CARDIAC PROGENITOR CELLS ACROSS PATIENT AGE POPULATIONS

Approved by:

Dr. Michael Davis, Advisor
School of Biomedical Engineering
Georgia Institute of Technology & Emory University

Dr. Gregory Gibson
School of Biological Sciences
Georgia Institute of Technology

Dr. Peng Qiu
School of Biomedical Engineering
Georgia Institute of Technology & Emory University

Dr. Hee-Cheol Cho
School of Biomedical Engineering
Georgia Institute of Technology & Emory University

Date Approved: [December 10, 2021]

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Michael Davis for providing the project and funding and assisting with analysis of the data. His guidance was immensely helpful for understanding the research and implications of the work along with assisting with writing this thesis.

I would also like to acknowledge the incredible assistance and mentorship of Jessica Hoffman, who conducted much of the wet lab work necessary prior to sequencing and provided much needed computational knowledge when I first started in the group. In addition, I also thank the folks in the Kaushal group at Northwestern University for providing our lab with the single cell sequencing data used in the first study.

Finally, I thank my committee members: Dr. Gregory Gibson, Dr. Peng Qiu, and Dr. Hee Cheol Cho. They provided the insightful questions and commentary necessary to ensure the validity of my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

aCPCs	Adult c-kit+ cardiac progenitor cells
cCPCs	Child c-kit+ cardiac progenitor cells
CHD	Congenital heart disease
CPCs	Cardiac progenitor cells
HF	Heart failure
HLHS	Hypoplastic left heart syndrome
LV	Left ventricle
MI	Myocardial infarction
nCPCs	Neonatal c-kit+ cardiac progenitor cells
RV	Right ventricle
RVEF/LVEF	RV/LV ejection fraction
RVFS/LVFS	RV/LV fractional shortening
TAPSE	Tricuspid annular plane systolic excursion

SUMMARY

Human c-kit+ cardiac progenitor cells (CPCs) have seen success in the treatment of heart failure and myocardial dysfunction.[1,2] However, research has demonstrated the reparative capacity of CPCs appears to be linked to the age of the patient, with younger patients having increased heart function and reduced fibrosis following treatment.[3–5] We hypothesized that these differences may be driven by differing subtypes of CPCs existing in each donor sample. Using the high-resolution capabilities of single cell RNA-sequencing technologies, we hope to elucidate the different subtypes that may be giving rise to the differences in therapeutic outcomes observed during *in vivo* studies. In the first study we analyzed the differences between adult CPCs (aCPCs) and neonatal CPCs (nCPCs). *In vivo* studies indicated injected aCPC had reduced cell retention and cell proliferation due to increased phagocytosis in comparison to nCPCs. We found three distinct subtypes of CPCs following analysis of the data. The two nCPC-enriched clusters correlated strongly with wound healing and cell proliferation, while the third aCPC-enriched cluster indicated some immune response activity. Analysis of selected gene expression in the third cluster indicated reduced expression of *CD47*, an important anti-phagocytic protein, along with reduced expression of several important growth factor and ECM proteins.[6] In the second study we analyzed the differences between pediatric patient populations. Previously published *in vivo* and *in vitro* results indicate reduced fibrosis and immune response and increased chemotaxis when using nCPCs in comparison to child CPCs (cCPCs). Subcluster analysis finds cCPC-enriched clusters upregulated in several fibrosis- and immune response-related genes. Clustering of genes indicates genes correlated in chemotaxis to be

upregulated in nCPC clusters. We identified the surface proteins versican and ITGA2 to be upregulated in fibrosis-related cluster 6 cells. Flow cytometric analysis using antibodies specific to these proteins identified a cell population with high levels of both proteins, consistent with the gene expression profile identified by the cluster 6 cells. We hope that this research will allow for researchers in the future to better optimize for and predict clinical outcomes prior to injection in autologous CPC-based therapies.

CHAPTER 1. INTRODUCTION

1.1 Heart Failure and Adverse Remodeling

1.1.1 Heart Failure in Children

Heart failure (HF), characterized by the heart's inability to maintain systemic blood circulation, is commonly caused in adults by a myocardial infarction (MI) resulting from coronary artery disease.[7,8] In pediatric populations however, the major driver of heart failure is congenital heart disease (CHD), a condition that encompasses a variety of birth defects of the heart.[9,10] Of all patients suffering from CHD, those with diseases resulting in single ventricular physiology are disproportionately more likely to result in patient mortality due to the high complexity of the disease intervention.[11] Hypoplastic left heart syndrome (HLHS) is the most prevalent of the single ventricular CHDs, present in roughly 1 for every 4,000 births.[11,12] Children with HLHS often undergo palliative surgery in 3 stages. The stage I operation redirects the right ventricle (RV) outflow through the aorta while pulmonary blood flow is provided using a shunt. At this stage, the RV is converted into the main ventricle for systemic circulation, subjecting the ventricle to volume and pressure overload. The second stage operation is conducted at roughly 6 months of age where the patient undergoes surgery to disconnect the systemic or ventricular pulmonary blood flow and replaces it with a direct connection to the superior vena cava to reduce the volume load on the RV. In the stage III operation at roughly 3 years of age, blood from the inferior vena cava is redirected to the pulmonary artery to further reduce the volume load on the RV.[11,13]

Surgery is often required for treating HLHS in order to redirect blood flow; however, the overloaded myocardium often undergoes further remodeling during and post-surgery due to increased stress on the RV. This can lead to heart dysfunction and potentially heart failure if the RV is no longer able to provide systemic circulation.[11,13,14]

The exact mechanism by which heart failure arises following surgical palliation in HLHS patients is not fully understood. Research indicates a majority of the deformation occurs due to the stage I Norwood procedure exposing the RV to the greatest volume and pressure overload following the surgery. An echocardiographic study demonstrated the contractility of HLHS patient hearts decreasing following both stage I and stage II procedures by studying how the longitudinal and circumferential strain rates changed over time. The strain rates were shown to be roughly constant across the first 3.5 years in control patients, eliminating the potential confounder of patient age and implicating the surgical procedures as the main source of the contractile variation.[14] In models of pulmonary hypertension, the RV undergoes several maladaptive changes resulting in increased fibrosis and cardiomyocyte hypertrophy.[11,15] Additionally, angiogenic capabilities of HLHS hearts are significantly reduced in both ventricles, resulting in reduced perfusion and oxygen delivery to cardiomyocytes which ultimately leads to myocardial dysfunction.[11,16]

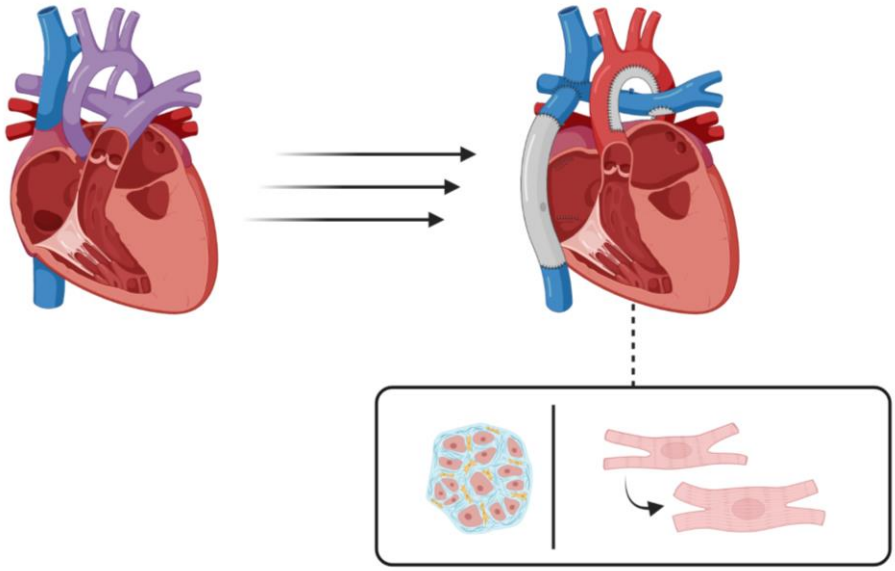

Figure 1. Adverse remodeling associated with HLHS post-operation.
Patients that undergo palliative surgery to treat HLHS often suffer from heart dysfunction post-operation to several maladaptive processes including cardiac fibrosis and cardiomyocyte hypertrophy in the RV. Figure generated using BioRender.

1.1.2 Reverse Remodeling

Reverse remodeling is a process which aims to interfere with the normal remodeling process during myocardial dysfunction to attenuate maladaptive outcomes and improve the heart function of myopathic hearts. The goal of reverse remodeling is to reduce the progress of several processes correlated with negative outcomes including myocardial hypertrophy, ventricular distortion and dilation, and cardiac fibrosis.[17,18]

There are several measures used in clinical applications to track the progress of cardiac remodeling. RV/LV ejection fraction (RVEF/LVEF), a measure of the ventricle's ability to pump blood out of the heart; RV/LV fractional shortening (RVFS/LVFS), the reduction in length of the end-diastolic diameter after the ventricular systole; and tricuspid annular plane systolic excursion (TAPSE), a measure of the longitudinal displacement of

the tricuspid ring, are used as measures of heart contractility.[19–21] The end diastolic volume, a measure of the volume of blood in the ventricle before contraction, is used to assess changes in ventricular dilation.[22] Cardiac fibrosis is often assessed using posterior wall thickness, the thickness measured from the interior of the ventricle to the outside of the heart, or by Masson trichrome staining of the myocardial area.[23] In adult populations, it is common to assess the functional improvements to the LV, while in pediatric populations suffering from single ventricular CHDs, these measures are often made on the RV.

1.2 Therapeutic Approaches to Induce Reverse Remodeling

Several methods for therapeutically inducing the reverse remodeling process have been studied in clinical applications with many drugs developed to target different remodeling processes. For example, angiotensin-converting enzyme inhibitors and angiotensin receptor blockers block activation of the renin-angiotensin-aldosterone system that has been implicated in adverse remodeling during HF and administration of the drugs are strongly correlated with patient survival especially shortly after experiencing MI.[17,18] Similarly, β-Blockers, used to reduce blood pressure by blocking the activity of epinephrine, also demonstrate an ability to limit adverse remodeling with patients having higher LVEFs and lower LV end-systolic volumes when compared to controls.[17,18]

While drugs have shown limited efficacy in limiting adverse remodeling, interest in cell-based therapies for treating heart disease have grown in recent years after a significant body of research has demonstrated the ability of stem cells to regulate the remodeling process. Several different types of cells have been pursued to this end, including mesenchymal stem cells, skeletal myoblasts, and cardiac progenitor cells.[11,24,25]

In this paper we will be focusing on the ability of cardiac progenitor cells to repair the myocardium.

1.2.1 Cardiac Progenitor Cells

Cardiac progenitor cells or cardiac-derived progenitor cells (CPCs) are a heterogenous class of progenitor cells that reside in the cardiac tissue.[26] Several different subtypes of CPCs have been identified for their potential therapeutic benefits and are currently the topic of rigorous study.[27] Of these, c-kit+ CPCs are of particular interest representing one of the two CPC subtypes that have been tested at the level of clinical trials.[28–30] While it was originally believed that c-kit+ CPCs could represent a multipotent CPC subtype that could differentiate into cardiomyocytes and induce cardiac regeneration, subsequent research has demonstrated the therapeutic benefits of all CPCs, including c-kit+ CPCs, is largely due to paracrine signaling to induce cardiac repair.[31] This mechanism of repair is especially beneficial in cardiac applications since the transplantation of these cells is less likely to result in arrhythmogenesis by obstructing the heart syncytium unlike mature cell-based therapies.[31]

1.2.1.1 Clinical Applications of CPCs

A multitude of *in vivo* studies have verified the efficacy of c-kit+ CPCs for inducing a reverse remodeling process using anti-fibrosis and pro-angiogenic signaling. The collective research and literature strongly support the utility of c-kit+ CPCs for improving functional outcomes post-treatment as demonstrated by increased ejection fraction along with reduced fibrotic tissue and myocardial hypertrophy.[1] Recent clinical trials using autologous CPC-based therapies seem to support the use of CPCs for cardiac repair. In

particular, the recent CONCERT-HF clinical trial (NCT02501811) conducted with adults suffering from heart ischemia showed autologous c-kit+ CPC transplantation strongly correlated with reductions in major adverse cardiac events post-treatment.[2]

One major problem regarding the translational potential of many autologous cell-based therapies is high donor variability.[32] In particular, CPCs are known to possess age-dependent reparative capabilities, with older patients having worsened therapeutic outcomes.[3,4] The main focus of this research is to determine if this variability can be captured and analyzed using single cell RNA-sequencing technology and additionally if subpopulations correlated with negative outcomes can be removed from the donor samples to improve therapeutic outcomes. The results of two projects will be presented in chapters 2 and 3 focusing on the characteristic differences of CPCs present in adults and children, respectively. For both studies, neonatal CPCs (nCPCs) were used as the basis of comparison.

CHAPTER 2. COMPARATIVE ANALYSIS OF NEONATE AND ADULT CPC POPLUATIONS USING SINGLE CELL TECHNOLOGY

2.1 Background and Motivation

Published literature and experimental work conducted by researchers in the Kaushal group at Northwestern University indicate a significant difference in the therapeutic outcomes of treatments using neonatal c-kit+ CPCs (nCPCs) and adult c-kit+ CPCs (aCPCs). *In vivo* experimental results by Sharma et al. demonstrate the ability for nCPCs to improve several indicators of heart function including left ventricle ejection fraction, fractional shortening, end diastolic volume, and posterior wall thickness. Proteomic analysis of the two CPC populations found higher expression of the proliferation protein MKI67 in nCPC patient populations. nCPCs were also able to retain expression levels of genes important for maintaining the self-renewal properties over five passaging cycles, while aCPCs showed substantially lower expressions of the genes by the last passage. aCPCs also showed reduced protein expression of c-kit after passaging, a reduced proliferation rate, and reduced telomere length. nCPCs were found to be more resistant to apoptosis in the presence of reactive oxygen species compared to aCPCs and several important paracrine factors relating to proliferation, angiogenesis, and cardioprotection were found to be expressed at higher levels in nCPC conditioned media. nCPC-treated myocardial sections also contained significantly fewer macrophages than aCPC-treated sections.[3]

2.1.1 CPCs Improve Heart Function in a Rat MI Model

Work done by our collaborators in the Kaushal Lab compared several different cell therapies, namely cardiosphere-derived cells (CDCs), adult bone marrow-derived mesenchymal stem cells (BM-MSCs), umbilical cord blood cells (UCBCs), aCPCs, and nCPCs, which have been used in previous studies for inducing reverse remodeling and functional recovery after cardiac dysfunction. The results indicates that both aCPCs and nCPCs (referred to as aMSCs and nMSCs in Figure 2 and Figure 3) are the most efficacious of the cohort for improving several measures of heart function including increases in the LV ejection fraction and LV fractional shortening (Figure 2a-b) while also reducing cardiac fibrosis following repair (Figure 2c-d).

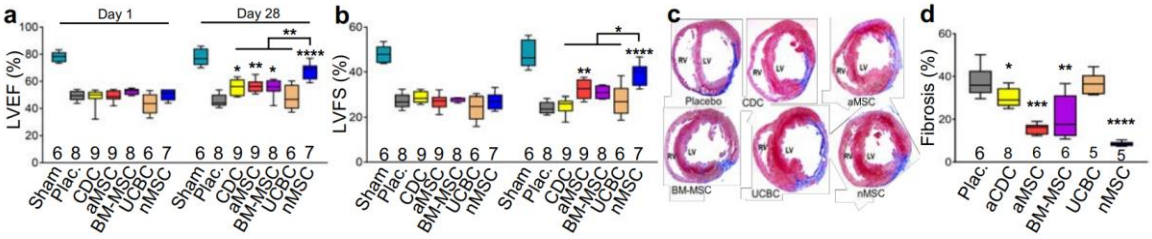

Figure 2. Cardiac functional outcomes following stem cell transplantation in the MI model.
(a) Left ventricular ejection fraction (LVEF) and (b) fractional shortening (LVFS) derived from echocardiography are shown for post-operative Day 1 and Day 28 with different cell therapies (N=6-9). (c) Measurement of fibrosis by Masson trichrome staining and (d) quantitative assessment showed a significant decrease in development of fibrosis in injected nMSC hearts compared with the placebo control and other stem cell types (N=5-8). Data under review.

2.1.2 nCPCs Elude Phagocytosis by Macrophages

A proposed mechanism by which nCPCs may induce increased functional recovery compared to aCPCs is by inhibiting phagocytosis, thereby increasing cell retention, and allowing the cells to continue to release paracrine signals and proliferate. Our collaborators

studied this effect by tagging nCPCs and aCPCs with GFP and measuring the cell abundance after 24h in an *in vitro* co-culture experiment with M1 PKH26 macrophages. aCPCs were found to significantly increase phagocytosis, resulting in reduced cell retention while nCPCs inhibited phagocytosis allowing for increased cell proliferation (Figure 3a-b). *In vivo* experimental results in an immunocompetent rat MI model showed an additional benefit of nCPCs by reducing size of the macrophage population at the site of injection after 48 hours (Figure 3c-d).

Bulk RNA-sequencing analysis indicates mRNA expression CD47, and important anti-phagocytotic protein, was 1.7-fold higher in nCPC populations compared to aCPC populations, consistent with results from the proteomic analysis (Figure 3e-g). Up-regulation of CD47 in nCPCs was verified using immunoblotting (Figure 3h-i).

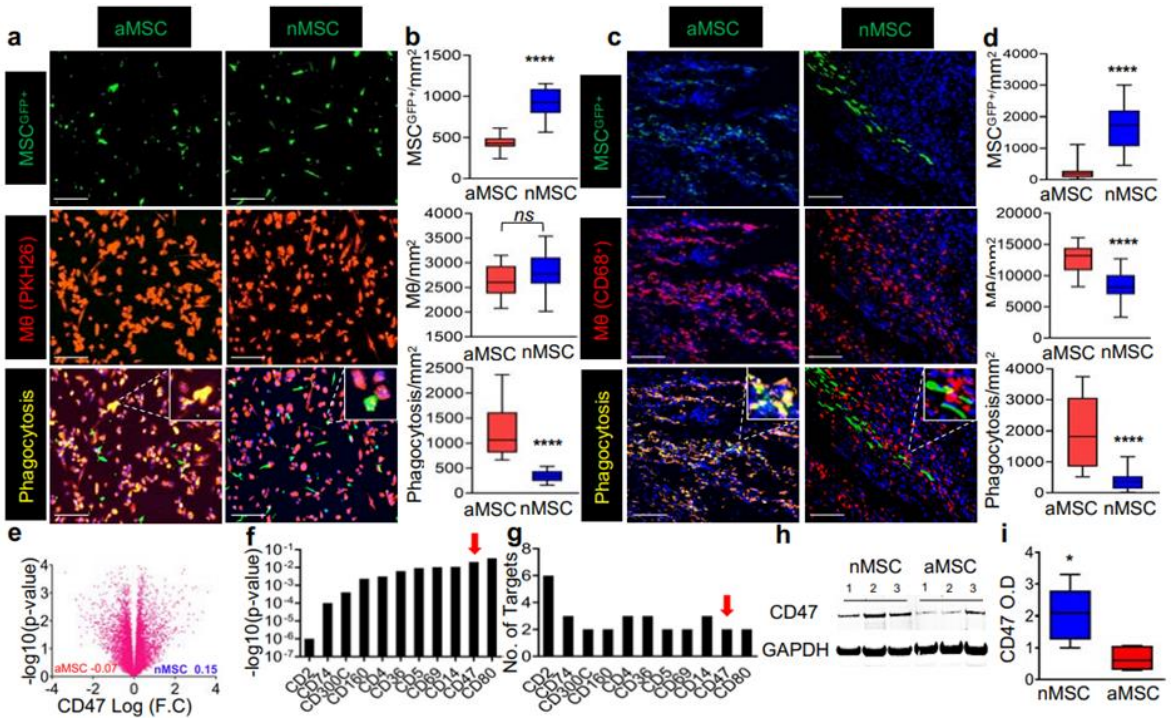

Figure 3. Cardiac nMSCs evade phagocytosis.

(a) Representative images showing Mφ- mediated phagocytosis of nMSCs^{GFP+} or aMSCs^{GFP+} after 24 h co-culture (Scale bars 75 μm). (b) Quantification of nMSCs^{GFP+} or aMSCs^{GFP+} (green) phagocytosis by Mφs (PKH26, red) after 24 hours. c, Representative images showing Mφ- mediated phagocytosis in rat MI model (Scale bars 75 μm), proliferating nMSCs^{GFP+}, and CD68+ macrophages and phagocytosis compared to aMSCs^{GFP+} injection. (d) Quantification of nMSCs^{GFP+} retention and CD68+ phagocytic cells compared with aMSC^{GFP+} injected in MI rats (N=4) at 2 days. (e) Graphical representations of the mRNA sequencing data. The mRNAs are ranked in a volcano plot according to their statistical (P value) significance (-log P; y-axis) and the ratio of their relative abundance ratio (log2-fold change; x-axis) in nMSCs (blue) and aMSCs (red). Histogram representation of genes with significantly increased mRNA (f) and their targets (g). (h) Immunoblots showed expression of CD47 in nMSCs and aMSCs (N=3). GAPDH served as a loading control. (i) Quantification of CD47 in nMSCs and aMSCs (using ImageJ). Data under review.

In order to further resolve differences between nCPC and aCPC populations, we utilized the high-resolution capabilities of single cell RNA-sequencing to identify whether distinct subclusters exist between neonatal and adult CPC populations. In addition, we determined if the subclusters can help explain the differences in therapeutic outcomes as identified in the *in vivo* studies relating to cell proliferation, paracrine signaling, and macrophage recruitment.

2.2 Methods

2.2.1 Experimental Methods

All cell samples were prepared and sequenced by researchers in the Kaushal group at Northwestern University. Tissue biopsies were collected from the right atrial appendage of three neonate and three adult patients during surgery. The cells were sorted for c-kit+ CPCs using magnetic cell sorting and the cells were cultured in flasks. The cells were submitted for sequencing using 10x Chromium technology with Single Cell 3' v2 chemistry.

2.2.2 Computational Methods

A summary of the analysis pipeline is presented in Figure 5.The fastq files were processed and aligned to the GRCh38 human reference genome using CellRanger (10x Genomics, v2.1.0).[33] Raw counts were processed using the Scanpy single cell analysis library in Python.[34] Doublets were subsequently filtered using the Scrublet method as implemented in Python and the dataset was filtered for cells with greater than 1500 total transcripts, more than 780 distinctly expressed genes, and a less than 20% mitochondrial gene fraction (Figure 4).[35] Only genes that were expressed in at least 20 cells were used in subsequent analyses.

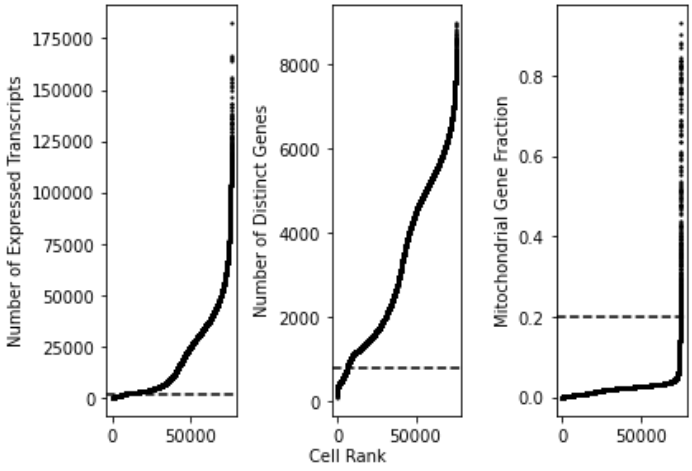

Figure 4. Quality control metrics for filtering low quality cells.
The dataset was filtered using the for cells with more than 1500 total transcripts, more than 780 distinctly expressed genes, and less than 20% mitochondrial gene fraction.

Each batch was normalized using the cell pooling methodology as implanted in the SCRAN R package and utilized a coarse Leiden clustering to identify initial cell pools.[36,37] Jurkat cells were removed for subsequent analyses by identifying cells with normalized CD3E and CD3D expression values greater than 1. The batch balanced k-nearest neighbors

algorithm was used in conjunction with a ridge regression to align individual batches and remove sources of unwanted variance from the dataset.[38] This batch corrected data was clustered using the Leiden algorithm at a resolution of 0.5. The differential gene expression was computed on the clusters with non-batch corrected data using the rank_genes_groups function in Scanpy with the default t-test method. The top 100 upregulated genes in each cluster that were expressed in at least 10% of the rest of the dataset were fed into the Metascape pathway analysis tool. The final dataset contained 61,979 cells, 60.5% of which were aCPCs and 39.5% of which were nCPCs. A summary of the analysis pipeline is shown in Figure 5.

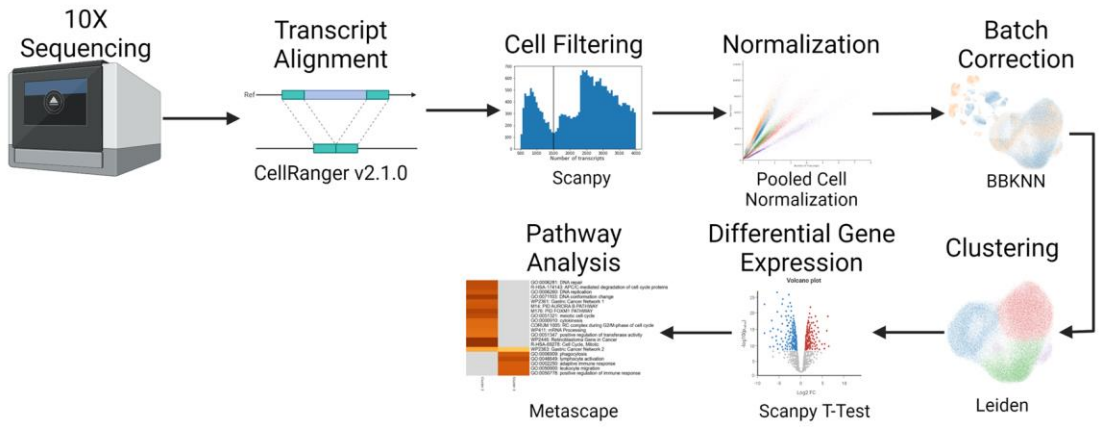

Figure 5. Single cell analysis pipeline.
After sequencing, cells were aligned using CellRanger and processed using Scanpy. Normalization was computed using pooled cell normalization in the SCRAN package and batch correction computed using the batch balanced k-nearest neighbors algorithm. Cluster was computed using the Leiden community finding algorithm at a resolution of 0.5. Differentially expressed genes were computed using a t-test and Metascape was used for pathway analysis.

2.3 Results

2.3.1 Clustering Analysis

nCPCs and aCPCs were clustered into four groups using the Leiden community finding algorithm (Figure 6a-b). Analysis of the cluster compositions indicates nCPCs were especially overrepresented in cluster 2, while aMSCs were overrepresented in cluster 3 (Figure 6c-d). Gene expression analysis indicates the first two clusters have high expression of *CD47*, consistent with previous results, along with higher expression of several key cardioprotective paracrine factors including *PDGFA*, *HGF*, *IGF1*, *FGF2*, *VEGFA*, and *CXCL12* (Figure 6c). Cluster 3 had an especially low expression of *CD47* and *SDF1*/*CXCL12* when compared to the other three clusters. Cluster 1 cells have upregulated fibronectin gene expression, which has previously been implicated as an essential protein in cardiac repair in CPC-based therapies and cluster 2 had high expression of the proliferation-related gene *MKI67* (Figure 6c). Notably, cluster 4 makes up a significantly small proportion of cells (3%) in comparison to the other cell clusters (Cluster 1: 39%, Cluster 2: 29%, Cluster 3: 29%).

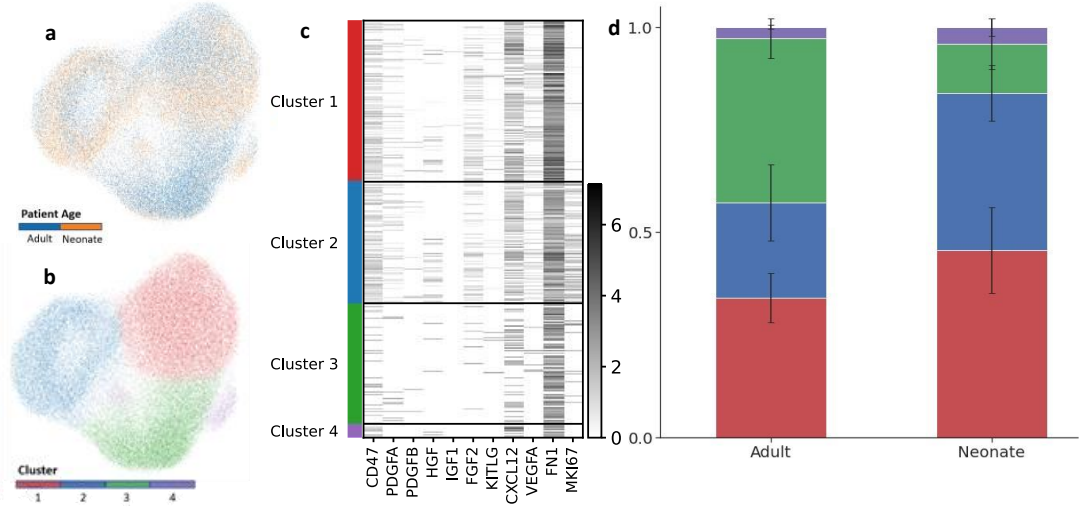

Figure 6. Compositional analysis of CPC single cell data.
(a) UMAP plot colored by age group. (b) Clusters determined using Leiden community detection algorithm. (c) Gene expression of selected genes. (d) Cluster compositions by

age group. Error bars indicate standard deviation of cluster proportions across patient samples.

2.3.2 Differential Gene Expression and Pathway Analysis

Up-regulated genes in the nCPC enriched cluster 2 were associated with cell cycle and proliferation while those in aCPC enriched cluster 3 were correlated with leukocyte migration and cell death (Figure 7b-c). Similar analyses of clusters 1 and 4 indicate the former is upregulated in genes associated with integrin interactions, supramolecular fiber organization, and wounding response while the latter is upregulated in genes involved in ribosomal activity and the *VEGFA-VEGFR2* signaling pathway (Figure 7c). In addition, cluster 1 cells have upregulated fibronectin gene expression, which has previously been implicated as an essential protein in cardiac repair in CPC-based therapies (Figure 7a). These results indicate that the differing functional responses of aCPCs and nCPCs may be attributed to a few subpopulations of cells. Specifically, increased phagocytosis associated with aCPCs may be attributed to the relatively high abundance of the cluster 3 cells in the population, while the increased proliferative and adhesive properties of the nCPCs may be attributed to the higher abundance of cluster 1 and 2 cells.

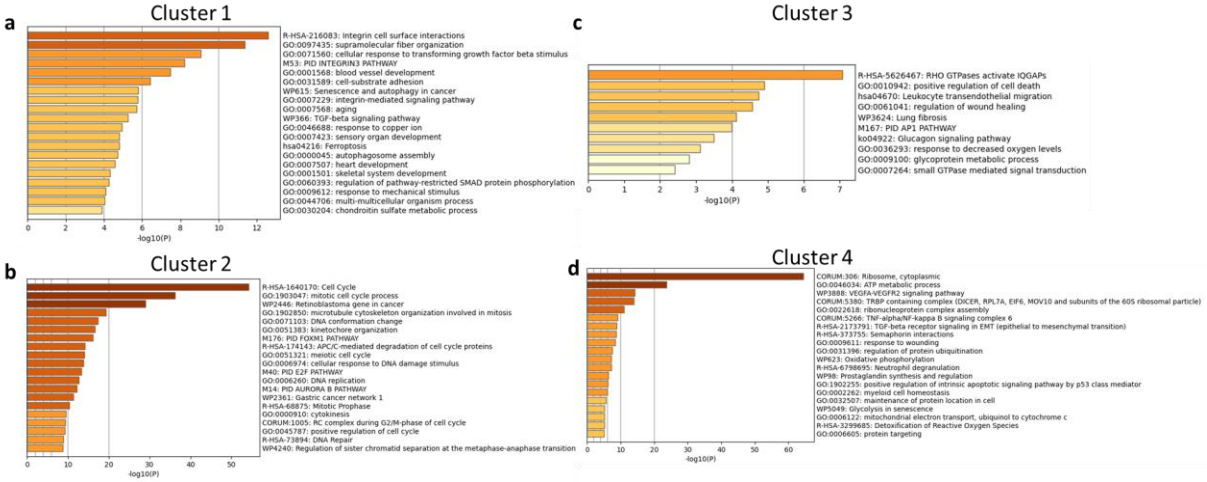

Figure 7. Cluster-specific pathway analysis.
Pathway analysis barplots generated using Metascape. The top 100 differentially expressed genes from each cluster with expressions in greater than 10% of cells in the non-cluster population were used for the analysis.

2.4 Discussion

Experimental results have demonstrated a difference in therapeutic outcomes attributed to nCPCs and aCPCs, posing a considerable challenge for autologous CPC-based therapies in adult patients. We believe much of this difference can be explained by analyzing the relative compositions of CPC subtypes present in a donor sample. In particular, adult patients appear to have high abundance of the pro-phagocytotic cluster 3 cells, while neonatal patients have higher abundances of reparative and proliferative cells characterized as cluster 1 and 2 cells, respectively. In addition, analysis of cardioprotective paracrine signals and proliferative factors indicate that neonatal samples are more capable of evading phagocytosis, promoting neovascularization and repair, and self-renewing. We believe these compositional differences can help to explain the differences in therapeutic outcomes observed in the *in vivo* experimental results.

Deeper analysis of the cluster gene expression indicates the first cluster is highly involved in cardiac repair, with differential expression of collagen genes such as *COL1A2*, *COL4A1,* and *COL8A1* (Figure 8a). Genes such as *FN1* and *ITGAV* have been previously implicated as being especially important for inducing cardiac repair and *NEAT1* and *MALAT1* have been shown to be relevant during cell proliferation and angiogenesis. Cluster 2 related genes correlate strongly with cell proliferation and cell cycle (Figure 8b). *MKI67,* a previously reported gene marker for CPC proliferation, is differentially expressed in this cell cluster.[3] In addition, *H2AFZ,* with its resultant protein H2A.Z.1, and *PCLAF*

have been implicated as important regulators of cell cycle in hepatocellular carcinoma.[39,40] Cluster 3 is moderately correlated with leukocyte migration due to high expression of genes like *CTNNB1* and *CXCR4* and displays high expression of *CTGF*, a marker of fibrosis (Figure 8c).[41]

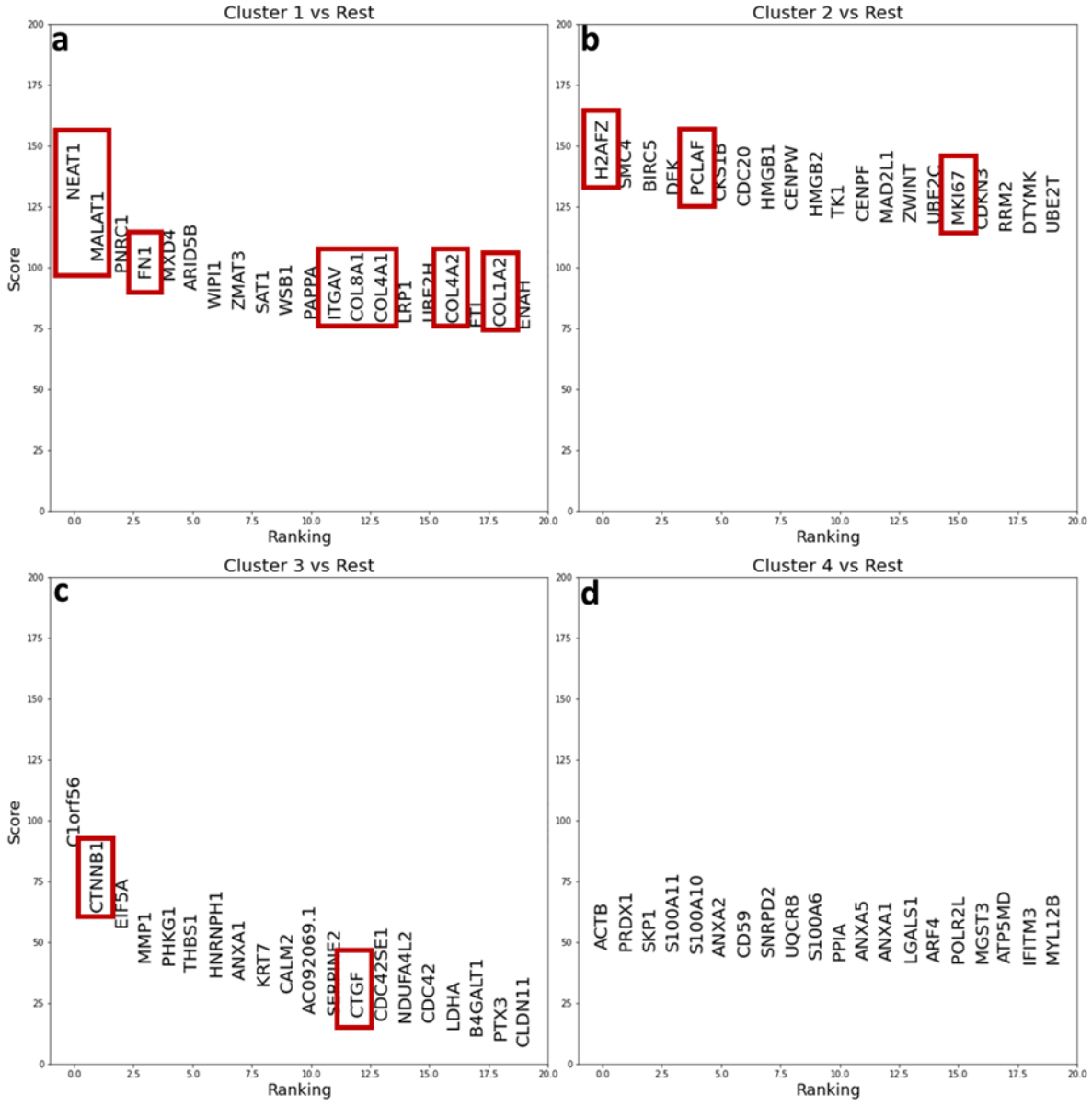

Figure 8. Cluster-specific differential gene expression.
Differential gene expression computed using the Wilcoxon rank sum method for each cluster against the rest of the cells. Genes are ordered by highest z-score.

CHAPTER 3. COMPARATIVE ANALYSIS OF NEONATE AND CHILD CPC POPULATIONS USING SINGLE CELL TECHNOLOGY

3.1 Background and Motivation

Cardiac cell-based therapies have seen success in clinical applications for the treatment of congenital heart disease.[11,42] In particular, human c-kit+ cardiac progenitor cells (CPCs) have been identified as a particularly potent cell type for inducing repair in the damaged myocardium.[11] Recent studies have demonstrated CPCs induce their pro-reparative effects through paracrine signaling mechanisms as opposed to the conventional engraftment, differentiation, integration mechanism attributed to many stem cell therapies.[11,31] Experimental studies indicate the paracrine release of CPCs contributes to anti-fibrotic outcomes and increased neovascularization of the myocardium which drives a reverse remodeling process during wound healing.[4,11] While the exact mechanisms for how CPCs induce this effect continues to be a contentious area of study, a recent clinical trial (CONCERT-HF) conducted with adults suffering from heart ischemia found CPCs improved patient survival post-MI by reducing the reducing the number of major adverse cardiac events.[2]

Existing CPC-based therapies separate CPCs from a tissue sample from the patient's heart.[2] However, as is the case with many autologous cell-based therapies, patient-to-patient variability plays a significant role in the outcome of the patients post-therapy (Figure 9).[43] Previous research by our group has found that the age of the patient can play a major role in determining the reparative capabilities of a patient's CPCs. In particular,

we found that neonatal CPCs (nCPCs) possessed greater anti-fibrotic signaling, reduced immune response, and increased chemotaxis capabilities in comparison to child CPCs (cCPCs).[4] Because of the age dependence of the CPC function, we hypothesized that each patient CPC sample contained many subpopulations of CPCs, and these CPCs were transitioning to reduced reparative states as the patient aged. Based on this, we would expect to see a more heterogeneous cell population among older patient samples, which we hope may help to explain the difference in therapeutic outcomes among these patient samples. Ultimately, by identifying these potentially problematic CPC populations, it may be possible to improve therapeutic outcomes.

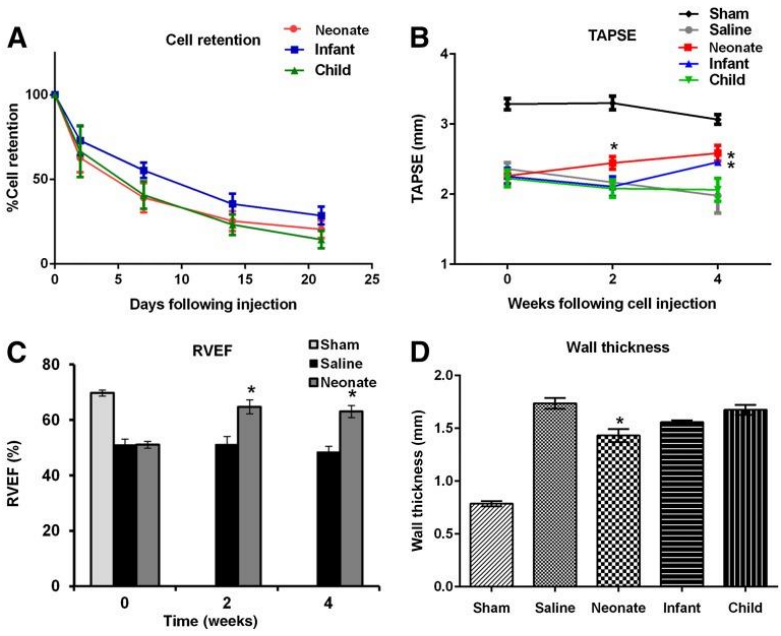

Figure 9. cCPCs result in poor therapeutic outcomes in *in vivo* rat models.
(a) Cell retention after intracoronary injection of labeled CPCs in a pulmonary artery banding rat model showed no significant difference between age groups after 3 weeks. (b) Measure of tricuspid annular plane systolic excursion (TAPSE) over 4 weeks indicates significant improvements when using neonatal or infant CPCs. (c) RV ejection fraction (RVEF) improved significantly 2 weeks after injection of neonatal CPCs. Similar results were obtained when using infant and child CPCs. (d) Wall thickness measured 6 weeks after banding indicate significantly decreased wall thickness when using neonatal CPCs. Adapted from Agarwal et al.[4]

3.2 Methods

3.2.1 Experimental Methods

Cells collected from the right atrial appendage of five neonatal (< 1 month) and five child (3.43 years ± 2.6 years) patients suffering from congenital heart disease were separated for c-kit+ CPCs using magnetic cell sorting. In order to identify the functional differences between neonate and child CPCs, the sorted cells were expanded in culture and submitted for single cell RNA sequencing (10X Genomics) with downstream analysis conducted using Seurat. Patient characteristics for samples used in the study are listed in Table 1. One neonatal patient sample (Patient 985) was removed due to low transcript counts.

Table 1. Metadata and Patient Characteristics.

Sample	*Patient*	*Age Group*	*Passage*	*Heart Disease*	*Sex*
1	903	Neonate	11	Hypoplastic left heart syndrome	F
2	925	Neonate	5	Total anomalous pulmonary venous return	F
3	930	Neonate	5	Total anomalous pulmonary venous return	M
4	926	Child	6	Ventricular septal defect	M
5	1048	Child	12	Atrial septal defect	F
6	896	Child	9	Ventricular septal defect	F
7	938	Child	6	Subaortic Membrane Resection	M
8	1092	Child	15	Atrial septal defect	F
9	985	Neonate	7	Interrupted aortic arch	M
10	2016	Neonate	12	Atrial septal defect	F

Tabulated metadata and patient characteristics for CPC samples used in this study.

Flow cytometry was utilized for characterization of CPC subpopulations based on expression of the versican and integrin alpha 2 surface proteins. Primary anti-versican (Creative Biolabs, CAT # CBMAB-C9301-LY) and anti-integrin alpha 2 (R&D Systems,

CAT # FAB1233P) antibodies conjugated to Alexa Fluor 647 and PE respectively were selected for analysis. Pooled child cells were cultured using a T75 flask in growth serum. Cells were split using 5 mL of TrypLE and detachment was halted by pipetting an equal volume of growth serum. The cells were subsequently suspended in a 2% FBS in PBS flow buffer solution after 5 min of centrifugation at 2000 RPM. Cells were centrifuged and suspended once more in 1 mL of flow buffer. Five separate samples were plated of equal volumes: one unstained, one for each antibody, one for the Zombie Yellow live/dead stain, and one containing both antibodies and the live/dead stain. The antibodies were added to the respective samples using the volumes listed in Table 2. The cells were then incubated for 45 min and centrifuged and resuspended in 200 µL of flow buffer twice. The CytoFLEX flow cytometer was then used for running the experiments. The compensations were computed automatically by the CytoFLEX instrument software.

Table 2. Antibody dilutions for flow cytometry.

Antibody	**Volume**
Anti-versican	5 µL
Anti-integrin alpha 2	1.5 µL
Zombie Yellow	1.5 µL

Antibody dilutions were selected based on manufacturer's instructions.

3.2.2 Computational Methods

Raw reads from single cell sequencing were processed using CellRanger (10x Genomics, v6.0.0).[33] The filtered raw counts data was processed using the Seurat package in R (Satija, v4.0.5).[44] Cells with less than 1000 or greater than 7000 distinctly expressed genes and mitochondrial gene fraction totalling greater than 5% of total transcript counts were filtered from the dataset.

Data from patient samples were integrated by first normalizing counts using the SCTransform method whereby mitochondrial gene fraction and passage number were regressed from the datasets.[45] The datasets were then integrated together using the comprehensive integration methodology implemented in Seurat. Cells were then clustered using the Louvain community finding algorithm and differential expression was computed on non-batch corrected data using the Wilcoxon rank sum method as implemented in Seurat. Trajectory analysis was conducted using Monocle 3 (Tapnell, v3.1) with the "ncenter" parameter in the learn_graph function set to 500. Pseudotimes were computed by setting the root node as the cluster of interest and allowing monocle to compute pseudotime values for the remaining cells.[37,46–49] A summary of the analysis pipeline is shown in Figure 10.

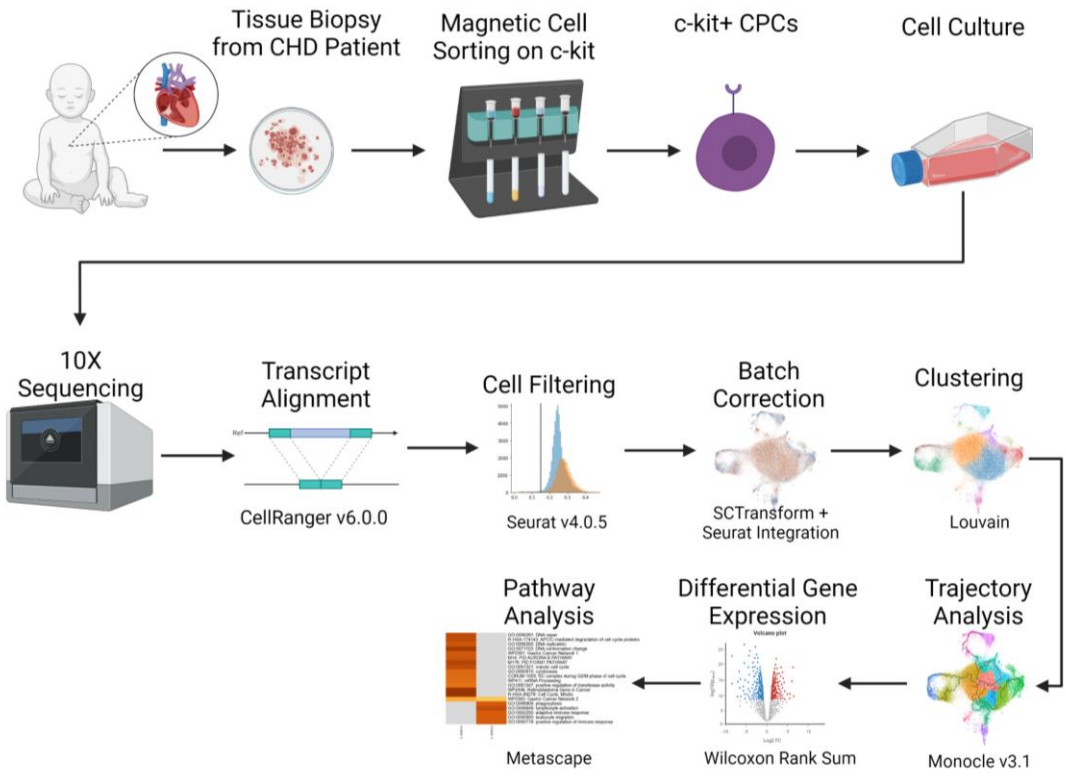

Figure 10. Single cell analysis pipeline.
Top sequence illustrates the steps for separating and culturing CPCs from CHD patients. Bottom sequence summarizes the analysis pipeline and computational tools used in the analysis. Figure generated using BioRender.

Surface proteins were identified using the cell surface protein atlas validated surface proteomes dataset.[50] The surface proteome dataset was filtered for proteins for which there was a high confidence of expression on the cell surface. The dataset was also further filtered for cluster of differentiation proteins for better identification of cell surface proteins. The dataset was analysed for conserved differentially expressed genes across donor cells of the same cluster. The differentially expressed genes were then filtered for only genes present in the filtered surface proteome dataset for determination of highest transcriptionally expressing surface proteins.

3.3 Results

3.3.1 Clustering and Compositional Analysis

An initial clustering using Louvain identified twelve CPC subpopulations (Figure 11a-b). Neonatal samples were largely enriched in the first two clusters while child samples were enriched in clusters 3, 6, 8, and 9. Among the child samples, there was a high level of sample-to-sample variability in comparison to the neonatal samples. Some child samples possessed more neonate-like clustering profiles (Patients 896 and 926) while other samples produced a more dissimilar clustering profile with less cells represented in the first two clusters (Patients 938, 1048, and 1092). Pathway analysis of the clusters indicates that Clusters 0, 1, and 6 are highly involved in the tissue reparative process with genes highly correlated with angiogenesis and fiber organization. The child CPC-enriched cluster 6 had differentially expressed genes highly correlated with supramolecular fiber organization. Clusters 2 and 5 seem to be related to cell proliferation and cell cycle processes.

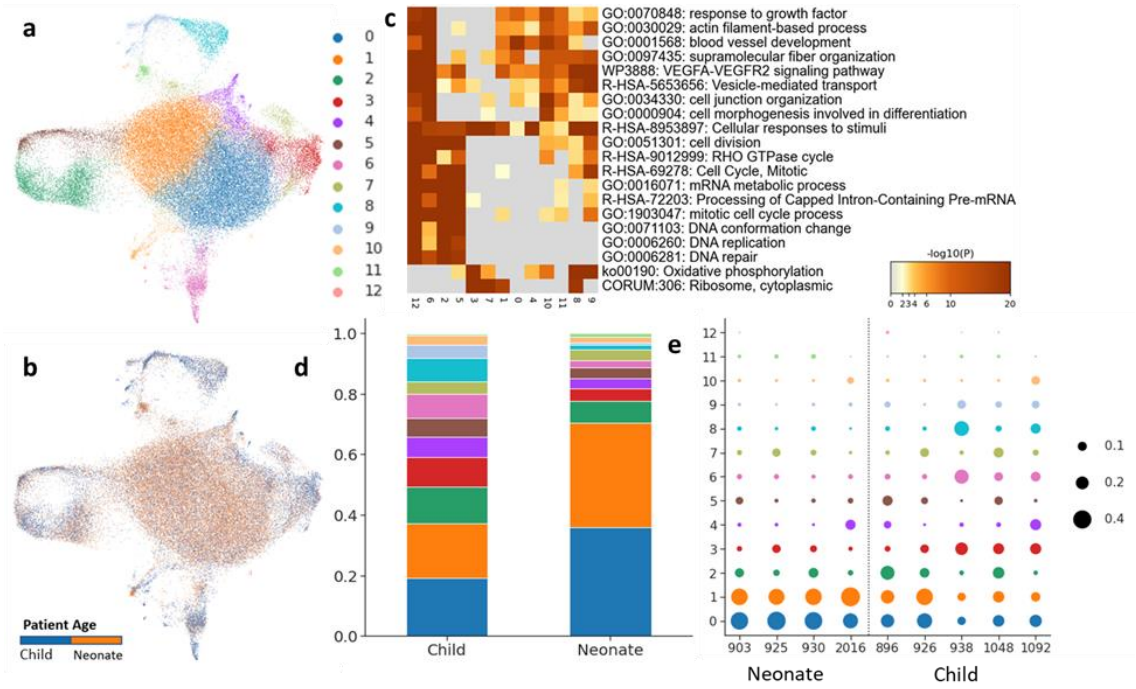

Figure 11. Clustering and cluster compositions of nCPCs and cCPCs.
UMAP projections colored by (a) cell cluster and (b) age group. (c) Metascape pathway analysis heatmap of upregulated genes. Cluster composition as grouped by (d) age group and (e) patient sample.

3.3.2 Trajectory Analysis

Trajectory analysis was conducted using Monocle 3 to understand how gene expression shifts as cells move between CPC subpopulations (Figure 12a). Pseudotimes computed using cluster 2 cells as the root node resulted in highest pseudotimes in cluster 8 cells, indicating the gene profiles of these cells to be the most distinct from the cluster 2 proliferating cells (Figure 12b). Alternatively, pseudotimes computed using cluster 6 as the root node resulted in the largest pseudotimes at cluster 2 (Figure 12b). Co-expression of genes were computed along trajectories and highly co-expressed genes were clustered into 21 gene clusters, henceforth referred to as modules (Figure 12c). Some modules corresponded strongly with certain clusters from the Seurat analysis. For example, cluster

4 cells had high expression of genes in module 8, while cluster 2 and 5 cells had high expression of module 12 genes.

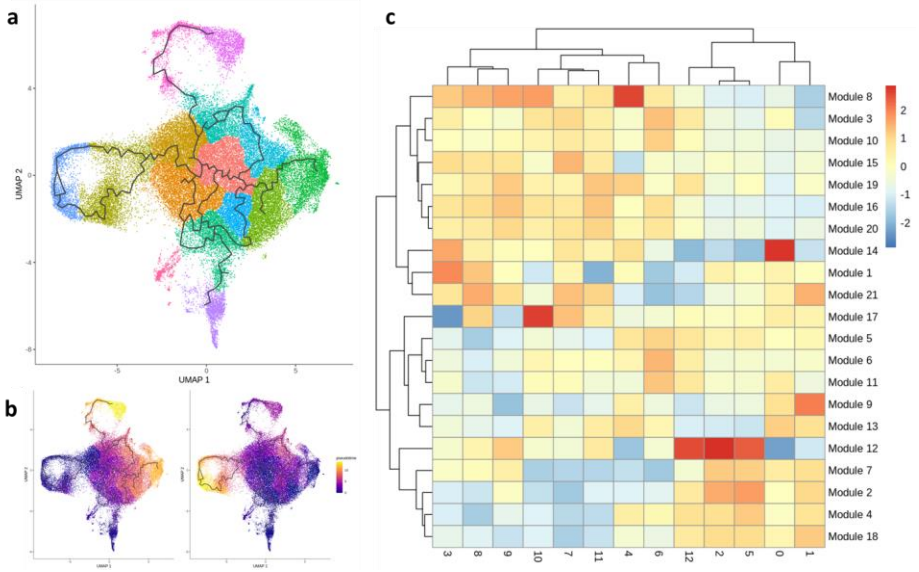

Figure 12. Trajectory analysis and gene clustering.
UMAP projections with trajectories determined by Monocle colored by (a) monocle clusters and (b-c) pseudotime with root nodes set to (b) cluster 2 and (c) cluster 6 from the Seurat analysis. (d) gene module expression heatmap by Seurat cluster. Modules were determined through a Leiden clustering of highly co-expressed genes along trajectories.

Analysing module expression by age group indicates nCPCs are highly upregulated in modules 9, 13, 14, and 21 and cCPCs are upregulated in modules 1, 3, 8, 15, and 16 (Figure 13a). Pathway analysis of genes within each module indicates nCPCs to be enriched in pathways related ribosomal activity via module 21 genes (Figure 13b-c). Both populations are enriched in modules related to chemotaxis (modules 8 and 9), angiogenesis, and ECM organization (modules 3, 9, 13, and 15). In addition, nCPC-related module 13 was enriched in pathways related to small molecule biosynthesis. Module 9 contains *CD34* and *PDGFB*, genes believed be related with paracrine signaling, while module 8 only

contains one such gene, *FGF2*, while also containing high number of immune-related cytokines such as *IL6* and *IL1B*.

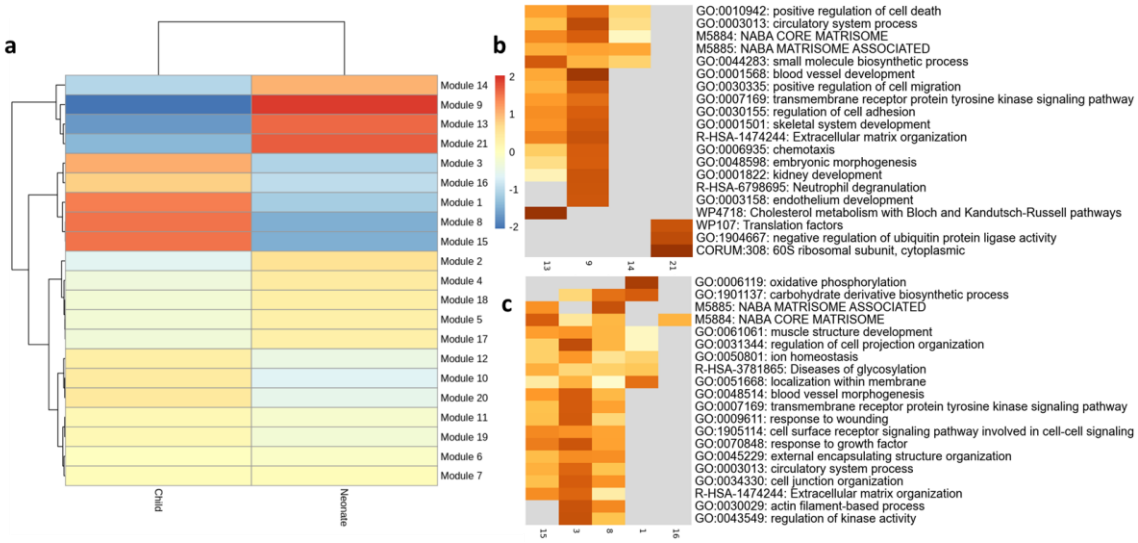

Figure 13. Module expression by age type.
(a) Heatmap of module expression grouped by age group. Metascape pathway analysis heatmap of upregulated modules in (b) neonate and (c) child patients.

3.3.3 Cluster 6 is Upregulated in Several Fibrosis-Associated Factors

A detailed analysis of several fibrosis-related genes indicates Cluster 6 cells may be a potential driver of fibrotic activity. Several different types of collagen genes are highly expressed in this population along with high expression of genes associated with fibrosis *TGFB2*, *CCN1*, *CCN2*, and *FBN1* (Figure 14a-b).[51,52] In addition, the upregulated genes *PDGFRA* and *FAP* are known fibroblast markers, but also believed to correlate with an epithelial-to-mesenchymal transition.[53,54] The cluster is also upregulated in angiogenic markers like *VEGFA* and downregulation of the proliferation-related gene *H2AFZ*.

Pathway analysis indicates this cluster is especially correlated with supramolecular fiber organization and angiogenesis. Pathways associated with TGF-B signaling also

observed in this cluster (not shown). Differentially expressed genes associated with this cluster co-cluster heavily within gene modules 3 and 6.

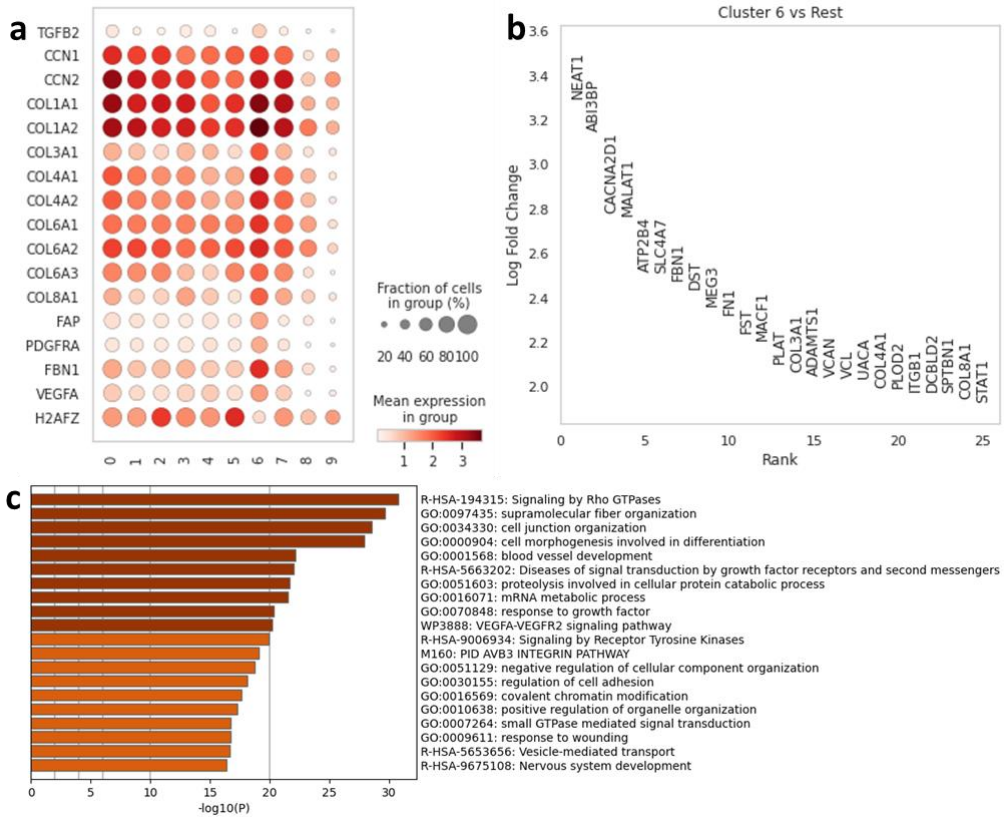

Figure 14. Characteristics of cluster 6 cells.
(a) Dotplot of selected genes relating to fibrosis and angiogenesis. (b) Top 25 differentially expressed genes ordered by log fold change between cluster 6 cells and non-cluster 6 cells. (c) Metascape pathway analysis barplot of upregulated differentially expressed genes.

3.3.4 Cluster 4 is Upregulated in Cytokines

Differential gene expression analysis identified several cytokines were upregulated in cluster 4 cells such as *IL1B, IL6,* and *IL33* (Figure 15a). Pathway analysis indicates strong correlation with immune-related signaling, including genes involved in the IL-10 and IL-17 signaling pathways (Figure 15b). In addition, enriched pathways indicate this cluster exhibits lower proliferative capabilities and apoptotic processes. Many of the

differentially expressed genes from this cluster were captured by the module 8 gene cluster, potentially indicating many of the cytokines expressed by these cells are specific to this cell cluster and are driven by similar biological processes.

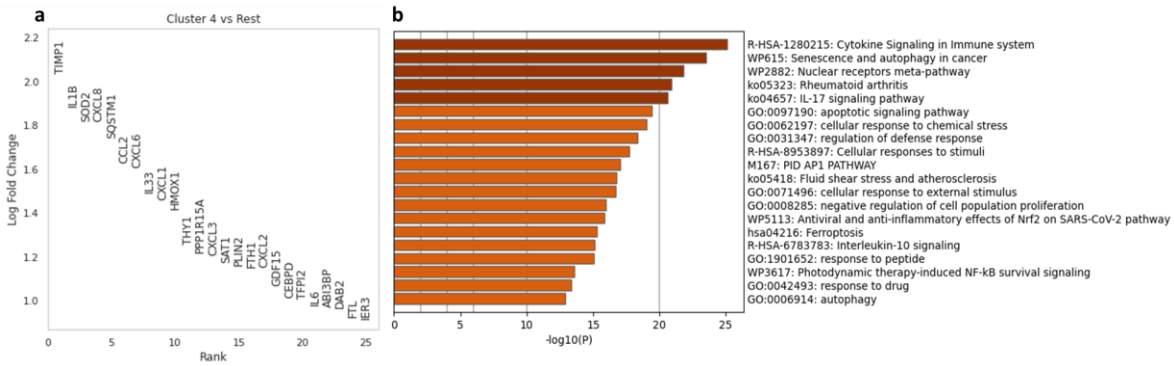

Figure 15. Characteristics of cluster 4 cells.
(a) Top 25 differentially expressed genes ordered by log fold change between cluster 4 cells and non-cluster 4 cells. (b) Metascape pathway analysis barplot of upregulated differentially expressed genes.

3.3.5 Transcriptional Expression of Surface Proteins in Cluster 6 Cells

Differentially expressed surface proteins were selected using the cell surface protein atlas database (Figure 16a).[50] Primary anti-versican and anti-ITGA2 antibodies were selected for characterization of pooled child CPC subpopulations using flow cytometry. A subpopulation of cells with high versican and ITGA2 antibody fluorescence was approximated to make up roughly 11% of the cells in the sample (Figure 16b). Analysis of the histograms of antibody fluorescence verifies the high expression of these proteins on the surface of these cells (Figure 16c-d). Zombie Yellow, a live/dead viability stain, showed higher fluorescence in these cells indicating a higher likelihood of these cells being dead, however we believe the measured fluorescence is still well within the threshold of live cells indicating this is likely our population of interest (Figure 16e).

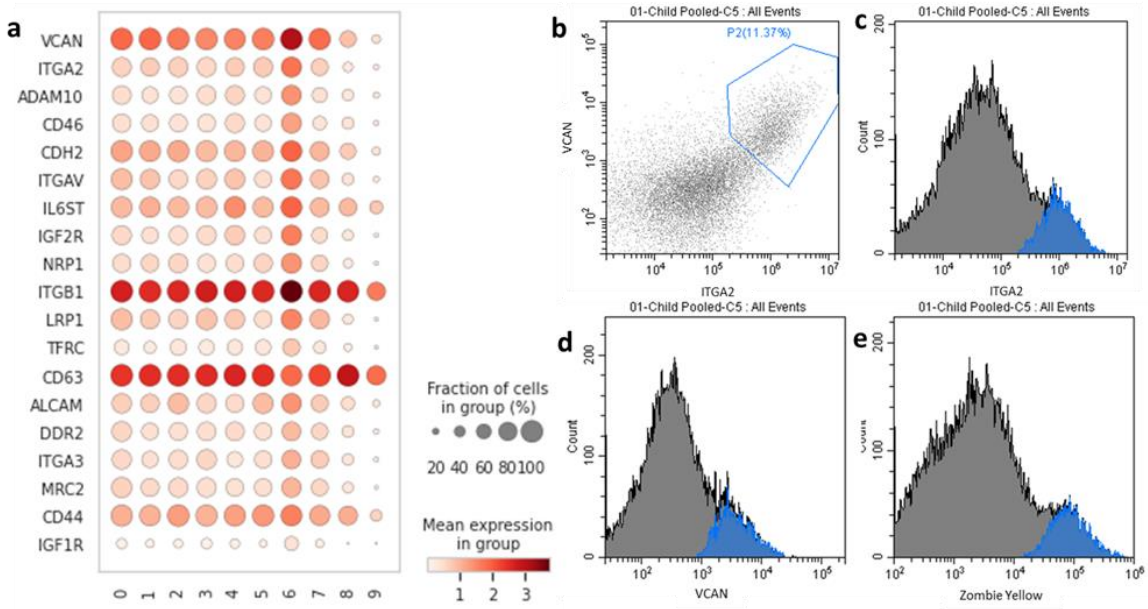

Figure 16. Cluster 6 surface protein characterization.
(a) Transcriptional expression of conserved differentially expressed surface proteins in cluster 6 cells. (b-e) flow cytometry analysis identifies the population of interest. Zombie Yellow was used as a cell viability stain. CytExpert was used to generate plots b-e.

3.4 Discussion

Previous research by our group has uncovered an age-dependence between neonate and child populations for treating HLHS using patient-derived CPCs.[4,5] Specifically, processes associated with anti-fibrotic outcomes and cell chemotaxis appear much stronger in neonatal population in comparison to child patients.[4] Through this research, we hoped to understand how these macroscopic dynamics present at the single cell level, and whether we would be able to discern these populations for selection using cell surface markers. Following clustering, we identified several subpopulations of CPCs, some strongly correlated with CPC proliferation, and others related to collagen synthesis and cytokine signaling (Figure 11c). We identified a high level of sample-to-sample variability among child patients, with some having more neonate-like clustering profiles than others (Figure

11e). In particular, a subset of cCPCs clustered largely in the offshoot branches of the UMAP projection, namely clusters 3, 4, 6, 8, and 9 while nCPCs largely clustered among clusters 0-5 (Figure 11d). Interestingly, there were more proliferating cCPCs than nCPCs in the dataset, as demonstrated by the higher proportion of cCPCs in clusters 2 and 5 (Figure 11d). However, this seems to be largely driven by a few donor samples rather than the overall cCPC population (Figure 11e).

Using trajectory analysis and gene clustering, we also identified several gene modules related to chemotaxis and supramolecular fiber organization (Figure 12). Both age groups contained gene modules related to angiogenesis, supramolecular fiber organization, and cytokine signaling. The modules related to cytokine signaling in nCPCs and cCPCs (modules 8 and 9) differed in the types of cytokines being measured, with nCPC signaling being more strongly associated with chemotaxis while cCPC signaling also contained many inflammatory-related cytokines (Figure 13). This is consistent with previous results indicating cCPCs to be driving an increased immune response and nCPCs driving higher levels of chemotaxis.[4] In addition, nCPCs had upregulation of genes strongly correlated with ribosomal activity, with nCPCs having a higher fraction of ribosomal genes in comparison to cCPCs. Pseudotime analysis indicates cCPC-enriched clusters 6 and 8 are highly distinct from the proliferative cell clusters (Figure 12b). In addition, cluster 6 cells had downregulation of proliferation-related gene *H2AFZ*, even in comparison to other non-proliferating CPCs in the dataset (Figure 14a). This may indicate that the cluster 6 cells represent an especially unique cell subtype from the rest of the CPCs that potentially do not possess the same self-renewing or proliferative properties.

Taking a closer look at specific clusters, we find that cluster 6 cells have high expression of genes related to fibrosis and angiogenesis (Figure 14). In addition, *FAP* and *PDGFRA*, known markers of fibroblast cells, are also upregulated in this cell cluster. These receptors have also been found to express at high levels during the epithelial-to-mesenchymal transition (EMT) which has been previously shown to play a role in the development of fibrosis.[53–55] β-catenin was strongly implicated in the EMT-induced fibrosis in renal tubulointerstitial fibrosis and is also upregulated in these cells.[55] Interestingly, this cluster also has very high levels of fibronectin expression, which has been implicated as a critical protein during cardiac repair. While high expression of fibronectin is present in all CPCs in the dataset, these results indicate that overexpression may potentially lead to adverse outcomes. Increased expression of *VCAN* has also been shown to correlate strongly with fibronectin expression in the literature.[56,57] *VCAN* has been shown to interact with collagen type I and fibronectin in the ECM, which results in reduced cell adhesion. In response, the cells may overexpress fibronectin which may explain the results in our data as both *VCAN* and *FN1* are upregulated in cluster 6.[56]

Another cCPC-enriched subpopulation, cluster 4 had high expression of several inflammation- and immune-related cytokines such as *IL1B* and *CXCL8* (Figure 15). However, unlike cluster 6, cluster 4 also makes up a significant proportion of some neonatal samples as is the case with patient 2016 (Figure 11e). Analysis of enriched biological pathways finds results consistent with bulk RNA-sequencing results from Agarwal et al for cCPC-enriched pathways, including induction of IL-10 and IL-17 signaling pathways.[4] The genes associated with this cell subpopulation all highly co-cluster in gene module 8, which is highly upregulated among cCPC samples (Figure 13a). Module

9, a similar gene module that is upregulated among nCPC samples, contained cytokines more associated with chemotaxis. The difference in these two gene modules may highlight an important difference in cytokine signaling patterns among child and neonate CPC populations.

Based on analysis of conserved genes and surface proteins, we found *VCAN* and *ITGA2* to be strong transcriptional surface protein markers for cluster 6 (Figure 16a). We used flow cytometry to confirm if the transcription upregulation of these genes translated to increase surface protein expression and if this could be used to characterize cell subpopulations. Analysis of the flow cytometry data on pooled child cells, indicated the population of interest to be discernable as cells with high ITGA2 and versican surface protein expression (Figure 16b-d). While the viability stain indicated higher fluorescence in the population of interest, we believe this measure is well within the margin for live cells.

CHAPTER 4. CONCLUSION

4.1 Comparison of Analysis Pipelines

We used separate analysis pipelines for each analysis, with each having its own set of benefits. In the first study, we utilized a pooled cell methodology for normalizing cell counts as implemented in the SCRAN package. This method aims to normalize "pools" of cells defined using a coarse Leiden clustering rather than normalizing on the library size of individual cells.[36] By doing this, the method aims to preserve the differentially expressed genes while also addressing the sparsity inherent to single cell datasets. The recently developed SCTransform normalization method used in the second project opts instead to fit a generalized linear model with a negative binomial error model to pools of genes with similar abundances. The model is then able to estimate residuals that effectively represent normalized expression values.[45] Both the BBKNN and Seurat v3 batch correction methods used in our projects rank highly in benchmarking studies for integrating datasets with cells of the same type.[58–60] BBKNN has been reported to be better suited to highly complex integration tasks and requires less memory and time to run even on larger datasets when compared to Seurat. However, Seurat has better integration with other single cell libraries such as Monocle 3 and the conserved differential expression analysis implemented in Seurat which was necessary for selecting conserved surface proteins across donor samples in the second project.

4.2 Comparisons Between the Two Datasets

Following the analysis of neonate, child, and adult CPC populations, we found several areas of potential overlap. In particular, the most potent area of overlap was among proliferating CPCs, with both datasets showing differential expression of the known CPC proliferation marker *MKI67* in these cells.[3] However, there were also several conserved genes that we believe have been understudied in the context of CPC proliferation. In particular, *H2AFZ*, the genetic precursor to the histone H2A.Z-1, serves as a particularly strong transcriptional identifier of proliferating CPCs in both datasets.[39] Unsurprisingly, analysis of the proliferation-associated module 12 also shows upregulation of many other histone-related genes, such as *HIST1H1A*, *HIST1H2AB*, *HIST1H2BB*, etc. These results indicate epigenetic processes play a significant role in the proliferative capabilities of CPCs and are conserved across patient age groups.

The clusters in both datasets that were most associated with fiber organization and collagen synthesis and angiogenesis had high expression of the long non-coding RNAs *NEAT1* and *MALAT1*. In addition, these cells are also upregulated in fibronectin expression, further validating the pivotal role the protein seems to play during cardiac repair and CPC function. The cCPC-enriched cluster 6 cells had overexpression of many of these genes in comparison to their neonatal counterparts indicating that higher expression of these genes may lead to potentially adverse outcomes. Interestingly, a subset of older patient CPCs in both datasets (aCPC-enriched cluster 3 and cCPC-enriched cluster 6) had upregulation of *CTNNB1*, the genetic precursor to β-catenin. Unlike in the aCPC samples, we did not find an overwhelming difference in *CD47* expression between the two age groups, agreeing with previous results indicating nCPCs and cCPCs have similar

retention times *in vivo* in contrast to aCPCs. Interestingly, in both datasets, *CD47* had the highest level of expression among proliferating CPCs.

4.3 Future Areas of Research

In the study of cCPCs, we have shown it is possible to characterize CPC subclusters using surface protein expression (Figure 16). Based on these results, it may be possible separate out cell subclusters of interest for further characterization and validation using *in vitro* assays to assess the ability of subclusters to promote remodeling processes of interest, such as fibrosis, angiogenesis, chemotaxis, and inflammation. In addition, these assays can be used to optimize the cluster compositions for the outcomes of interest before *in vivo* experimentation. Based on this, it may be possible to engineer CPC compositions to improve therapeutic outcomes.

Our group previously worked to characterize CPC samples as either "good" or "poor" performing cells based on predictive modelling on several outcomes such as TAPSE and RVEF.[5] It may be possible assess samples analysed using single cell technology in a similar fashion by correlating the *in vitro* outcomes of individual subclusters to the cluster compositions of each sample. In addition, such a model may potentially be expanded to bulk RNA-sequenced data through the use of recently developed deconvolutional methods such as BisqueRNA and MuSiC.[61,62] Such a model may allow for more rapid assessment of the viability of CPC samples for the treatment of HF and the necessity for potential compositional optimization on donor samples.

Previous research has indicated c-kit protein expression can decrease rapidly for CPC samples collected from adult patient after several passaging cycles.[3] Such reductions

in c-kit expression are also correlated with reductions in proliferative capacity and increases in markers of senescence.[3] Simultaneous characterization of single cell protein and gene expression is now possible using the Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq) technology. This would allow for genetic profiles to be attached to protein expressions of interest. Some proteins implicated in CPC function that could be characterized using this technology are c-kit, PECAM-1, ITGAV, ITGA2, PDGFRα, and LAMP3.[3,5,63]

The combination of the RNA sequencing data presented in this study along with protein expression and *in vitro* or *in vivo* characterization can be used to build a predictive model for clinical trial outcomes and indicate where potential areas of optimization may exist by tuning cluster-specific compositions. Ideally, quickly matching protein expression or bulk RNA-sequencing data to subcluster proportions may allow practitioners to quickly assess the viability of the patient sample or optimize cluster compositions before injection to improve clinical outcomes and reduce the undesirable variability inherent to autologous CPC-based therapies.

REFERENCES

1. Bolli, R., Tang, X.-L., Guo, Y. & Li, Q. After the storm: an objective appraisal of the efficacy of c-kit+ cardiac progenitor cells in preclinical models of heart disease. *Can. J. Physiol. Pharmacol.* **99**, 129–139 (2021).

2. Bolli, R. *et al.* A Phase II study of autologous mesenchymal stromal cells and c-kit positive cardiac cells, alone or in combination, in patients with ischaemic heart failure: the CCTRN CONCERT-HF trial. *Eur. J. Heart Fail.* **23**, 661–674 (2021).

3. Sharma, S. *et al.* A Deep Proteome Analysis Identifies the Complete Secretome as the Functional Unit of Human Cardiac Progenitor Cells. *Circ. Res.* **120**, 816–834 (2017).

4. Agarwal, U. *et al.* Age-Dependent Effect of Pediatric Cardiac Progenitor Cells After Juvenile Heart Failure. *Stem Cells Transl. Med.* **5**, 883–892 (2016).

5. Shoja-Taheri, F. *et al.* Using Statistical Modeling to Understand and Predict Pediatric Stem Cell Function. *Circ. Genomic Precis. Med.* **12**, e002403 (2019).

6. Kojima, Y. *et al.* CD47-blocking antibodies restore phagocytosis and prevent atherosclerosis. *Nature* **536**, 86–90 (2016).

7. CDC. Coronary Artery Disease | cdc.gov. *Centers for Disease Control and Prevention* https://www.cdc.gov/heartdisease/coronary_ad.htm (2021).

8. Heart Failure | NHLBI, NIH. https://www.nhlbi.nih.gov/health-topics/heart-failure.

9. Fahed, A. C., Roberts, A. E., Mital, S. & Lakdawala, N. K. Heart Failure in Congenital Heart Disease: A confluence of acquired and congenital. *Heart Fail. Clin.* **10**, 10.1016/j.hfc.2013.09.017 (2014).

10. CDC. What are Congenital Heart Defects? | CDC. *Centers for Disease Control and Prevention* https://www.cdc.gov/ncbddd/heartdefects/facts.html (2019).

11. Bittle, G. J. *et al.* Stem Cell Therapy for Hypoplastic Left Heart Syndrome. *Circ. Res.* **123**, 288–300 (2018).

12. CDC. Congenital Heart Defects - Facts about Hypoplastic Left Heart Syndrome. *Centers for Disease Control and Prevention* https://www.cdc.gov/ncbddd/heartdefects/hlhs.html (2019).

13. Philadelphia, T. C. H. of. Staged Reconstruction Heart Surgery. https://www.chop.edu/treatments/staged-reconstruction-heart-surgery (2014).

14. Tham, E. B. *et al.* Insights into the Evolution of Myocardial Dysfunction in the Functionally Single Right Ventricle between Staged Palliations Using Speckle-Tracking Echocardiography. *J. Am. Soc. Echocardiogr.* **27**, 314–322 (2014).

15. Zungu-Edmondson, M., Shults, N. V., Wong, C.-M. & Suzuki, Y. J. Modulators of right ventricular apoptosis and contractility in a rat model of pulmonary hypertension. *Cardiovasc. Res.* **110**, 30–39 (2016).

16. Salih, C., Sheppard, M. N. & Ho, S. Y. Morphometry of coronary capillaries in hypoplastic left heart syndrome. *Ann. Thorac. Surg.* **77**, 903–907 (2004).

17. Bhatt, A. S., Ambrosy, A. P. & Velazquez, E. J. Adverse Remodeling and Reverse Remodeling After Myocardial Infarction. *Curr. Cardiol. Rep.* **19**, 71 (2017).

18. Hellawell, J. L. & Margulies, K. B. Myocardial Reverse Remodeling. *Cardiovasc. Ther.* **30**, 172–181 (2012).

19. Ejection Fraction. *Cleveland Clinic* https://my.clevelandclinic.org/health/articles/16950-ejection-fraction.

20. Fractional shortening for estimation of ejection fraction. *ECG & ECHO* https://ecgwaves.com/topic/fractional-shortening-for-estimation-of-ejection-fraction/.

21. Sun, X. *et al.* Tricuspid annular plane systolic excursion (TAPSE) can predict the outcome of isolated tricuspid valve surgery in patients with previous cardiac surgery? *J. Thorac. Dis.* **8**, 369–374 (2016).

22. End-diastolic volume: What is it, and how do doctors use it? https://www.medicalnewstoday.com/articles/325498 (2019).

23. Obasare, E. *et al.* Natural history of severe aortic stenosis: Diastolic wall strain as a novel prognostic marker. *Echocardiography* **34**, 484–490 (2017).

24. Zhou, Y. *et al.* Direct injection of autologous mesenchymal stromal cells improves myocardial function. *Biochem. Biophys. Res. Commun.* **390**, 902–907 (2009).

25. Murry, C. E., Field, L. J. & Menasché, P. Cell-Based Cardiac Repair. *Circulation* **112**, 3174–3183 (2005).

26. Amini, H., Rezaie, J., Vosoughi, A., Rahbarghazi, R. & Nouri, M. Cardiac progenitor cells application in cardiovascular disease. *J. Cardiovasc. Thorac. Res.* **9**, 127–132 (2017).

27. Bianconi, V. *et al.* Endothelial and cardiac progenitor cells for cardiovascular repair: A controversial paradigm in cell therapy. *Pharmacol. Ther.* **181**, 156–168 (2018).

28. Bolli, R. *Myocardial Regeneration Using Cardiac Stem Cells Harvested From Right Atrial Appendages in Patients With Ischemic Cardiomyopathy*. https://clinicaltrials.gov/ct2/show/NCT00474461 (2017).

29. Davis, B. R. *A Phase II, Randomized, Placebo-Controlled Study of the Safety, Feasibility, & Efficacy of Autologous Mesenchymal Stem Cells & C-kit+ Cardiac Stem*

Cells, Alone or in Combination, Administered Transendocardially in Subjects With Ischemic HF. https://clinicaltrials.gov/ct2/show/NCT02501811 (2021).

30. Ge, Z., Lal, S., Le, T. Y. L., dos Remedios, C. & Chong, J. J. H. Cardiac stem cells: translation to human studies. *Biophys. Rev.* **7**, 127–139 (2014).

31. Marbán, E. A mechanistic roadmap for the clinical application of cardiac cell therapies. *Nat. Biomed. Eng.* **2**, 353–361 (2018).

32. Lipsitz, Y. Y. *et al.* A roadmap for cost-of-goods planning to guide economic production of cell therapy products. *Cytotherapy* **19**, 1383–1391 (2017).

33. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

34. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

35. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).

36. L. Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).

37. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

38. Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).

39. Yang, H. D. *et al.* Oncogenic potential of histone-variant H2A.Z.1 and its regulatory role in cell cycle and epithelial-mesenchymal transition in liver cancer. *Oncotarget* **7**, 11412–11423 (2016).

40. Ma, F. *et al.* Dysregulated NF-κB signal promotes the hub gene PCLAF expression to facilitate nasopharyngeal carcinoma proliferation and metastasis. *Biomed. Pharmacother.* **125**, 109905 (2020).

41. Lipson, K. E., Wong, C., Teng, Y. & Spong, S. CTGF is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis. *Fibrogenesis Tissue Repair* **5**, S24 (2012).

42. Tsilimigras, D. I. *et al.* Stem Cell Therapy for Congenital Heart Disease. *Circulation* **136**, 2373–2385 (2017).

43. Wang, L. L.-W. *et al.* Cell therapies in the clinic. *Bioeng. Transl. Med.* **6**, e10214 (2021).

44. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

45. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

46. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

47. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

48. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

49. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0092867415006376?via%3Dihub.

50. Bausch-Fluck, D. *et al.* A Mass Spectrometric-Derived Cell Surface Protein Atlas. *PLOS ONE* **10**, e0121314 (2015).

51. Ding, Y. *et al.* Roles of Biomarkers in Myocardial Fibrosis. *Aging Dis.* **11**, 1157–1174 (2020).

52. Bouzeghrane, F., Reinhardt, D. P., Reudelhuber, T. L. & Thibault, G. Enhanced expression of fibrillin-1, a constituent of the myocardial extracellular matrix in fibrosis. *Am. J. Physiol. Heart Circ. Physiol.* **289**, H982-991 (2005).

53. Kahounová, Z. *et al.* The fibroblast surface markers FAP, anti-fibroblast, and FSP are expressed by cells of epithelial origin and may be altered during epithelial-to-mesenchymal transition. *Cytom. Part J. Int. Soc. Anal. Cytol.* **93**, 941–951 (2018).

54. Jechlinger, M. *et al.* Autocrine PDGFR signaling promotes mammary cancer metastasis. *J. Clin. Invest.* **116**, 1561–1570 (2006).

55. Barriere, G., Fici, P., Gallerani, G., Fabbri, F. & Rigaud, M. Epithelial Mesenchymal Transition: a double-edged sword. *Clin. Transl. Med.* **4**, 14 (2015).

56. Wu, Y. J., Pierre, D. P. L., Wu, J., Yee, A. J. & Yang, B. B. The interaction of versican with its binding partners. *Cell Res.* **15**, 483–494 (2005).

57. Xie, J., Wu, Z., Xu, X., Liang, G. & Xu, J. Screening and identification of key genes and pathways in metastatic uveal melanoma based on gene expression using bioinformatic analysis. *Medicine (Baltimore)* **99**, e22974 (2020).

58. Luecken, M. D. *et al. Benchmarking atlas-level data integration in single-cell genomics*. 2020.05.22.111161 https://www.biorxiv.org/content/10.1101/2020.05.22.111161v2 (2020) doi:10.1101/2020.05.22.111161.

59. Chen, W. *et al.* A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. *Nat. Biotechnol.* **39**, 1103–1114 (2021).

60. Tran, H. T. N. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**, 12 (2020).

61. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).

62. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).

63. Monsanto, M. M. *et al.* Enhancing myocardial repair with CardioClusters. *Nat. Commun.* **11**, 3955 (2020).