Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature: 

Jessica Reggan Hoffman 6/30/22

Name Date

Statistical Modeling of C-kit+ Progenitor Cell Extracellular Vesicles to Predict Clinical Trial Outcomes

By
Jessica Reggan Hoffman
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Molecular and Systems Pharmacology

_____
Michael E. Davis, Ph.D.
Advisor

_____
John Calvert, Ph.D.
Committee Member

_____
Manu Platt, Ph.D.
Committee Member

_____
Roy Sutliff, Ph.D.
Committee Member

_____
Samantha Yeligar, Ph.D.
Committee Member

Accepted:

_____
Kimberly Jacob Arriola, Ph.D., MPH
Dean of the James T. Laney School of Graduate Studies

_____
Date

Statistical Modeling of C-kit+ Progenitor Cell Extracellular Vesicles to Predict Clinical Trial Outcomes

By

Jessica Reggan Hoffman
B.S., University of North Carolina at Chapel Hill, 2018

Advisor: Michael E. Davis, PhD

An abstract of a dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate Division of Biological and Biomedical Science, Molecular and Systems Pharmacology

2022

Abstract

Statistical Modeling of C-kit+ Progenitor Cell Extracellular Vesicles to Predict Clinical Trial Outcomes

By

Jessica Reggan Hoffman

Congenital heart disease is the most common form of birth defect and affects nearly 1% of live births in the US, annually. Complex forms of congenital heart disease, including hypoplastic left heart syndrome, require surgical palliation and often lead to pediatric heart failure. Originally studied for use in adult populations, cardiac cell therapy is gaining traction for pediatric populations as a therapeutic strategy to address underlying damage and potentially repair and regenerate myocardium. Specifically, our group is involved in the CHILD clinical trial (NCT03406884), investigating the use of autologous cardiac-derived c-kit+ progenitor cells (CPCs). Unfortunately, cardiac cell therapy preclinical and clinical investigations have been hampered by mixed results, primarily too much variation in cell populations and patient outcomes. Finally, there has been a shift in our understanding of how cardiac cell therapy works: functional effects may be attributed to paracrine signaling and the release of extracellular vesicles (EVs), rather than direct cell engraftment and differentiation. The purpose of this dissertation is to understand sources of variability and determine the biological signals contributing to repair in CPCs and CPC-derived extracellular vesicles (CPC-EVs). We analyzed transcriptomic data (bulk and single cell) and used machine learning regression models to link our RNA-sequencing data to functional outcomes. The subsequent dissertation chapters are designed to explore (1) differences between neonate- and child-derived CPCs at the single cell level, (2) bulk transcriptomic differences between patient matched CPCs and CPC-EVs, and (3) the relationship between CPC-EV cargo and *in vitro*, cardiac-relevant outcomes. Overall, we uncovered a more heterogenous population in child CPCs, enriched in pro-fibrotic and inflammatory cell subpopulations. We determined that CPC-EVs contain different RNA cargo than EVs derived from other cell types, and CPC-EVs are particularly enriched in miRNAs involved in cardiac development and cell proliferation. Finally, we used machine learning models to link CPC-EV RNA cargo from the CHILD trial samples to *in vitro* functional outcomes for the purposes of building a predictive and informative clinical tool.

Statistical Modeling of C-kit+ Progenitor Cell Extracellular Vesicles to Predict Clinical Trial Outcomes

By

Jessica Reggan Hoffman
B.S., University of North Carolina at Chapel Hill, 2018

Advisor: Michael E. Davis, PhD

A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate Division of Biological and Biomedical Science, Molecular and Systems Pharmacology

2022

Acknowledgements

There is a long list of folks I'd like to thank for helping me succeed these past four years in graduate school. First, I'd like to thank my advisor, Mike Davis, for supporting me and giving me the flexibility and opportunity to explore my scientific interests. Mike provided several opportunities to collaborate with clinicians and scientists on interesting and rewarding projects. At every step, Mike supported my pursuit to learn and strengthen my bioinformatics skillset. I'd also like to thank my committee members: John Calvert, Manu Platt, Roy Sutliff, and Samantha Yeligar. John Calvert and Roy Sutliff offered insightful feedback on my F31 grant proposal and have provided helpful critiques on my qualifying exam and thesis committees. Manu Platt has given great feedback and help with partial least squares modeling, beginning with my first rotation project in 2018. Finally, Samantha Yeligar has provided invaluable feedback since the very beginning, from journal presentations to professional development and applications of bioinformatics. Dr. Yeligar has given me feedback along the way that has not only propelled my work forward, but also has allowed me to improve my soft skills and develop as a scientist.

I would also like to thank my previous advisors and mentors that have helped me along the way. Thank you to Dr. Naama Toledano Furman and Dr. Roberto Molinaro from Houston Methodist Research Institute for introducing me to biomedical research during my summer internships. Further, I would not be where I am today without the help and support of Ilona Jaspers, Meghan Rebuli, and Adam Speen at UNC Chapel Hill. Dr. Jaspers encouraged me to think critically about research, allowed me to pursue my own projects, and encouraged me to apply for conferences and fellowships. Dr. Rebuli and Dr. Speen helped me learn the essential skills of science: experiment planning, statistical design, abstract and manuscript writing, cell

culture, and many more. Above all, they helped me realize that collaborative biomedical research can be fun and exciting.

I would also like to thank my Davis lab mates, past and present: Aline, David, Matt, Don, Ben, Sruti, Preety, Hyun-Ji, Kenneth, Olga, Kimmai, and Arun. Thank you to our lab manager, Milton, who was always willing to lend a hand during times, including experiment issues and ordering problems. Thank you to Olga and Kenneth for always lending an ear, offering help around the lab, and being ready for a long chat or coffee run. Further, I would not have been able to finish this dissertation work without the help of Arun, Hyun-Ji, and Sruti. Arun was a joy to mentor and provided me with insightful ideas and sequencing support that were critical to our *Genomics* paper. A massive thank you to Hyun-Ji and Sruti for being the best teammates I could have hoped for. Hyun-Ji and Sruti helped me in countless ways: growing hundreds of millions of CPCs, running endless ultracentrifuge spins, offering outstanding feedback on my dissertation work, and more. I will miss our days in the cell chambé.

Thank you, Kaitlyn, for supporting me, providing motivation and countless laughs when I'm burnt out, and for always being ready to sing some musical theater. Thank you to Jonathan for everything, especially your encouragement through writing this dissertation and wrapping up graduate school. Thank you for your wonderful support and thank you for providing a bright spot to my days. Thank you to my MSP cohort. We've been through exceptional obstacles these few years. I appreciate the class help, as well the fun times.

Finally, I'd like to thank my family for their endless support. Thank you to Julie for being my role model and for always offering scientific (and life) advice and perspective. I am thankful for my relaxing time spent on "the ranch" writing manuscripts, chatting about science and careers. I'd like to thank my Atlanta family: Mark, Beth, Meg, Scott, and Matt. Thank you for

your support and for opening your home to me, especially during the early pandemic. I'd also like to thank my grandparents, Tom and Diane Hoffman, for their tremendous support over the years; the visits and encouraging phone calls have meant the world to me. Finally, I'd like to thank my parents for their love and support. Thank you for supporting my interests all these years, whether it is math or science or tap dance. You have given me everything and have gone above and beyond to help me. Thank you for the calls, visits, and even cat sits. Thank you for everything.

Table of Contents

List of Figures and Tables

List of Abbreviations

Abbreviation	Category	Description
NGS	Bioinformatics	Next-generation sequencing
scRNAseq	Bioinformatics	Single-cell RNA-sequencing
PLSR	Bioinformatics	Partial least squares regression
PCA	Bioinformatics	Principal component analysis
p	Bioinformatics	Variable
n	Bioinformatics	Sample
t-SNE	Bioinformatics	t-distributed stochastic neighbor embedding
UMAP	Bioinformatics	uniform manifold approximation and project
LASSO	Bioinformatics	Least absolute selection and shrinkage
SIMPLS	Bioinformatics	statistically inspired modification of partial least squares
RMSE	Bioinformatics	Root-mean-square error
CHD	Clinical	Congenital heart disease
HLHS	Clinical	Hypoplastic left heart syndrome
CHILD	Clinical	Autologous Cardiac Stem Cell Injection in Patients with Hypoplastic Left Heart Syndrome
CPC	Cell Biology	c-kit+ cardiac progenitor cell
CDC	Cell Biology	Cardiosphere-derived cells
MSC	Cell Biology	Mesenchymal stem (or stromal) cell
EV	Cell Biology	Extracellular vesicle
MVB	Cell Biology	Multivesicular body

Chapter 1. Introduction

1.1 Congenital heart disease

Congenital heart disease (CHD) is the most common type of birth defect, affecting nearly 1% of live births in the US annually. Approximately 25% of CHD patients require invasive treatment within the first year of life. Improvements in surgical outcomes and medical management has led to an aging CHD population: the median age of death from severe CHD has increased from 2 to 23 years between 1987 and 2005 (Khairy et al., 2010).

1.1.1 Forms

There are many forms of CHD ranging in severity and complexity, affecting the heart's chambers, septum, valves, arteries, and veins. Simple defects include atrial and ventricular septal defects, patent ductus arteriosus, and pulmonary stenosis. Many simple defects do not require intervention; openings may close and stenoses may improve over time. The mortality rate of simple CHDs has been estimated at 2.2 per 1,000 patient years (Buratto et al., 2016).

On the other hand, critical defects often require invasive intervention (**Table** 1-1). Critical CHD affects ~0.2% of newborns in the US annually and includes abnormalities like Ebstein anomaly, coarctation of the aorta, and hypoplastic left heart syndrome (HLHS), among others. Critical CHD cases may comprise of multiple defects. For example, tetralogy of Fallot is made up of a combination four defects: ventricle septal defect, pulmonary valve stenosis (or atresia), over-riding of the aorta, and right ventricular hypertrophy. Tetralogy of Fallot may be palliated with a shunt from the aorta to the pulmonary artery, balloon dilation, or stent in the right ventricle outflow. Additionally, a complete repair of this condition may be achieved by closing the ventricle septal defect and removing the pulmonary obstruction (Apitz et al., 2009). Other interventions for critical CHD conditions include cardiac catheterization and bandings.

Table 1-1. Critical congenital heart diseases.

Condition	Description	Occurrence (US births)	Treatment
Coarctation of the aorta	Narrow aorta	1 in 1,800	Cardiac catheterization, balloon angioplasty
Double-outlet right ventricle	Aorta connected to right ventricle instead of left ventricle	1 in 10,000	Biventricular repair (arterial switch) or intraventricular repair or Fontan operation
D-Transposition of the great arteries	Pulmonary artery and aorta are switched	1 in 3,400	Arterial switch operation (common) or atrial switch operation
Ebstein anomaly	Tricuspid valve and right ventricle are malformed	1 in 10,000	Tricuspid valve repair or replacement. Removal of atrial septal defect.
Hypoplastic left heart syndrome	Underdeveloped left ventricle (often with underdeveloped mitral valve, aortic valve, aorta ascending portion, and atrial septal defect)	1 in 3,800	Series of 3 surgeries: (1) Norwood procedure, (2) bi-directional Glenn shunt, (3) Fontan procedure
Interrupted aortic arch	Discontinuous aortic arch	1 in 50,000	Close aortic gap
Pulmonary atresia (with intact septum)	Pulmonary valve is not formed	1 in 7,100	Cardiac catheterization and balloon angioplasty
Total anomalous pulmonary venous return	Pulmonary veins return blood to the right atrium by an abnormal connection	1 in 7,800	Connect pulmonary veins to left atrium
Tetralogy of Fallot	4 defects: ventricle septal defect, pulmonary stenosis, enlarged aortic valve, right ventricular hypertrophy	1 in 2,500	Replace pulmonary valve and enlarge passage to pulmonary artery. Patch ventricular septal defect.
Tricuspid atresia	Tricuspid valve is not formed	1 in 9,750	Septostomy (enlarge atrial septal defect) Temporary pulmonary artery banding Sunt from aorta to pulmonary artery Bi-directional Glenn Fontan
Truncus arteriosus	Aorta and pulmonary artery are not properly separated	<1 in 10,000	Separation of the aorta and pulmonary artery

Sources: CDC.gov, clevlandclinic.org, (Mai et al., 2019)

Despite improvements in palliative and reparative surgery, many CHD patients require heart transplantation. Single ventricle (including HLHS), d-transposition of the great arteries, and

right ventricular outflow tract lesion conditions represent the top three diagnostic categories for patients undergoing heart transplantation (Lamour et al., 2009). Many issues remain regarding transplant, including donor availability and immune rejection. An estimated 15-20% of CHD patients on a transplant list die before receiving a donor heart, and 1-year survival rates after transplantation remain high, between 20 and 25% (Hsu and Lamour, 2015). Symptomatic heart failure is often an indication for heart transplantation in CHD patients. Unlike adult heart failure, however, there are fewer alternative treatment strategies: the mechanisms behind heart failure in children are unknown, and mechanical support devices (e.g. ventricular assist devices) are not designed for children and have rarely been tested in pediatric populations (Hsu and Lamour, 2015).

1.1.2 Etiology

The etiology for most CHDs is unknown. In fact, an estimated 90% of CHD is caused by multifactorial etiologies (Chung and Rajakumar, 2016). Several genetic and environmental, modifiable and non-modifiable risk factors have been identified. Modifiable risk factors contributing to CHDs include maternal dietary deficiency, substance abuse, obesity, diabetes, and air pollution. Conversely, non-modifiable factors include maternal rheumatologic disorders, genetics, medications, metabolic disorders, and viral infections (Mullen et al., 2021). Within the category of genetic causes, syndromic CHD can be caused by chromosomal aneuploidy, chromosomal translocation, chromosomal deletions, or single-gene defects. Chromosomal aneuploidy was the first discovered genetic cause and is associated with a large proportion of CHDs (Muntean et al., 2017). The well-studied single-gene disorders Alagille, Holt-Oram, Noonan, LEOPARD, Costello, Cardiofaciocutaneous, and CHARGE syndromes are also linked to CHDs (Muntean et al., 2017).

Non-syndromic CHD make up the majority of cases and are much more challenging to identify, due to genetic phenomena like genetic heterogeneity and polygenes. Currently, genes implicated in CHD include transcription factors (*GATA4*, *NKX2.5, TBX5, HAND2*), signaling genes (*FOXH1, NOTCH1, NOTCH2, JAG1*), and structural genes (*GJA1, ELN, FLNA*) (Pierpont et al., 2007). Importantly, next-generation sequencing (NGS) technology offers greater opportunity to identify novel CHD mutations. For example, single-cell RNA sequencing (scRNAseq) has further elucidated cardiac development mechanisms and given insight into particular progenitor cell types causing defects (de Soysa et al., 2019; Miao et al., 2020). Notably, racial minorities are disproportionately affected by CHDs, potentially due to environmental and genetic disparities (Oster et al., 2011). Recent efforts using induced pluripotent stem cell disease model may uncover genetic causes and connections to racial disparities.

1.1.3 Hypoplastic left heart syndrome

HLHS is a complex single ventricle CHD characterized by an underdeveloped left ventricle (**Figure** 1-1). Neonates with HLHS often display cyanosis, difficult breathing, poor feeding, cold extremities, and weak pulses. HLHS occurs in 2-3 per 10,000 live births in the US and accounts for ~1% of all CHD cases. Despite a lower prevalence rate, HLHS accounts for disproportionately high mortality: 25-40% of all neonatal cardiac mortality (Siffel et al., 2015). HLHS used to be a universally fatal diagnosis within the first week of life; an underdeveloped left ventricle has insufficient pumping capability to supply systemic circulation. However, HLHS is now palliated by a series of three operations which redirect blood flow such that the right ventricle sustains systemic circulation. Surgical intervention has provided better outcomes for HLHS patients with survival probabilities increasing from 0% before 1984 to 42% in 2005.

Currently, the highest mortality for HLHS patients occurs in the first year of life, before the stage II surgery, and children who survive to 1 year old have a long-term survival rate of 90% (Siffel et al., 2015).

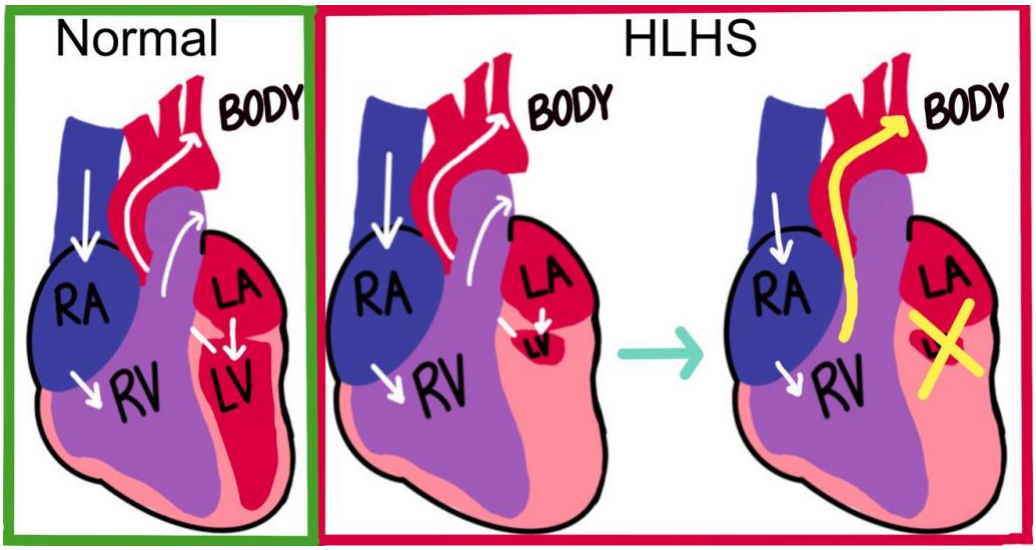


Figure 1-1. HLHS anatomy and circulation.
Comparison of blood flow in normal vs. HLHS heart. RA: right atrium; RV: right ventricle; LA: left atrium; LV: left ventricle.

The 3-stage palliation of HLHS is depicted in **Figure** 1-2. The stage I surgery, Norwood procedure, is performed within the first few weeks of life and begins the conversion of the right ventricle into the main ventricle providing for systemic circulation. The Norwood procedure redirects the right ventricle outflow to the aorta using a shunt either from the systemic artery to pulmonary artery (Blalock-Taussig shunt) or right ventricle to pulmonary artery (Sano shunt). The Norwood procedure causes right ventricle volume and pressure overload and necessitates the stage II surgery, the bidirectional Glenn procedure, performed four to six months later. The Glenn procedure redirects blood flow from the superior vena cava to the right pulmonary artery, reducing pulmonary vascular pressure and right ventricle volume overload. Finally, the stage III surgery, the Fontan procedure, is performed at around three years of age. The Fontan procedure redirects the remainder of systemic venous blood from the inferior vena cava to the right

pulmonary artery. The Fontan procedure further reduces the right ventricle volume load and completely separates oxygenated and deoxygenated blood.

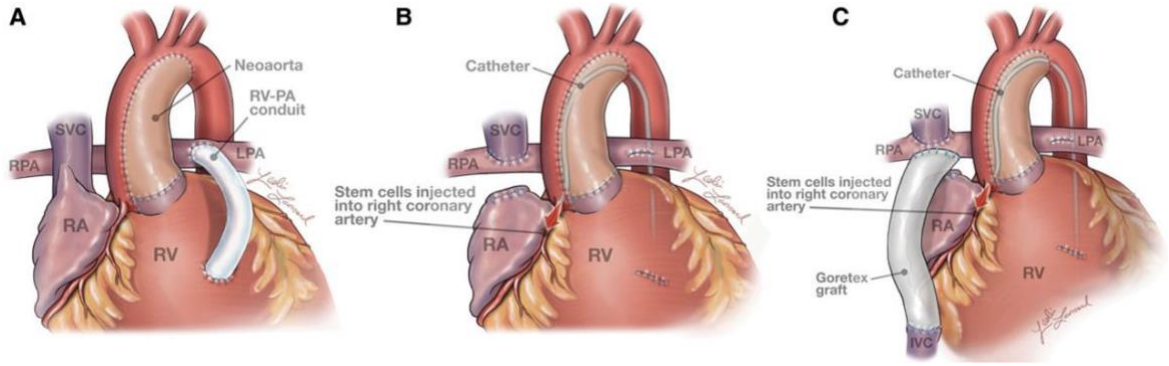


Figure 1-2. Three-stage palliation of HLHS.

Diagram depicting (**a**) stage I Norwood, (**b**) stage II bi-directional Glenn, and (**c**) Fontan operations. Illustration from (Wehman and Kaushal, 2015) shows the Transcoronary Infusion of Cardiac Progenitor Cells in Patients with Single Ventricle Physiology (TICAP) trial strategy.

Overall, surgical palliation supports single-ventricle physiology. However, the right ventricle is not physiologically adapted to sustain systemic circulation. In normal physiology, the right ventricle pumps blood into pulmonary circulation with low vascular resistance. Unlike the left ventricle, the right ventricle is crescent-shaped and contracts longitudinally without much twisting. The right ventricle has thinner walls, is more compliant, and has a greater volume than the left ventricle. Taken together, it is unsurprising that HLHS patients often develop RV dysfunction, adverse RV modeling, and eventual RV failure. RV dysfunction has been shown to predict mortality in HLHS patients: 18-month survival for patients with and without RV dysfunction are 35% and 70%, respectively (Altmann et al., 2000).

Currently, the only curative option of HLHS is heart transplant, and unfortunately, HLHS patients who have undergone surgical palliation have higher mortality rates after transplant than other CHD conditions (Everitt et al., 2012). Given the risks and limitations associated with heart transplant, therapeutic intervention to prevent RV dysfunction is necessary to improve outcomes

for HLHS patients. More specially, intervention between stage I and II operations, where mortality rates are highest, is required (Siffel et al., 2015).

1.2 Cardiac cell therapy for congenital heart disease

Traditionally viewed as a terminally differentiated organ, the heart has limited regenerative capacity and cardiac injury is often irreversible. Current pharmacological and interventional treatment strategies do not address the underlying damage of cardiovascular injury and transplant remains the only curative option. Given the aforementioned limitations of transplant, stem or progenitor cell therapies, which may induce endogenous repair mechanisms, have gained traction as a strategy for improving cardiac function, preventing adverse remodeling, and potentially reversing cardiovascular damage. A recent estimate suggests that more than 5,000 patients have received cell therapy for various cardiovascular diseases (Wehman and Kaushal, 2015). Initially investigated in adult populations, cell therapies for pediatric patients with congenital heart disease are now under clinical investigation. Notably, these trials primarily investigate cell therapies for critical single ventricle defects, including HLHS. The handful of cell therapy trials for CHD are summarized in Table 1-2.

Table 1-2. Cell therapy clinical trials for congenital heart disease.

Cell Type	Trial Name/ Identifier	Year	Description	Status/Results
CPC, autologous	CHILD: NCT03406884 *multi-center study with location at Children's Healthcare of Atlanta	2018-	Autologous CPC intramyocardial injection for with HLHS patients	Enrolling for phase II Phase I completed. Results not yet published. (Kaushal et al., 2022)
CDC, autologous	TICAP: NCT01273857	2011-2015	Transcoronary infusion of cardiac progenitor cells for	Manufacturing and delivery were feasible. No serious adverse events reported.

			single ventricle patients; Phase I	CDC-treated patients showed a higher right ventricular ejection fraction, improved somatic growth, and reduced heart failure status (Ishigami et al., 2015)
CDC, autologous	PERSEUS: NCT01829750	2013-2017	Cardiac progenitor cell infusion to treat univentricular heart disease; Phase II	CDC-treatment improved ventricular ejection fraction after 3 months (+6.4% vs. 1.3% control). Also, increases in somatic growth, quality of life, and reduced heart failure status. (Ishigami et al., 2017)
CDC, autologous	APOLLON: NCT02781922	2016-	Cardiac stem/ progenitor cell infusion in univentricular physiology; Phase III	Enrolling.
Bone marrow-derived MSCs, allogeneic	ELPIS: NCT03525418	2018-2022	Lomecel-B delivered during stage II surgery for HLHS; Phase I/II	Completed. Results not yet published. (Kaushal et al., 2017)
Bone marrow-derived MSCs, allogeneic	ELPIS II: NCT04925024	2021-	Lomecel-B delivered during stage II surgery for HLHS; Phase II (Next phase of ELPIS, MSCs now referred to as medicinal signaling cells)	Enrolling.
Mesoblasts: mesenchymal precursor cells (MPCs)	NCT03079401	2017-	Mesoblast stem cell therapy for single ventricle and borderline left ventricle patients	Active, not enrolling. Study participants are receiving the cell therapy; no new patients are being recruited or enrolled. Anticipated study completion: 01/2024
Umbilical cord blood-derived MNCs, autologous	NCT01883076	2013-2019	Autologous UCB-MNCs for HLHS patients; Phase I	Safety and feasibility: 0% operative mortality, one adverse event, no significant safety concerns over 6 months. No reduction in cardiac function related to intramyocardial injections of UCBs. (Burkhart et al., 2019)

Bone marrow-derived MNCs, autologous	NCT02549625	2015-2021	Intracoronary delivery of autologous bone marrow-derived MNCs for patients with single right ventricular failure; Phase I	Completed. Results not yet published. First patient demonstrated reduced ventricular size and ~5% increase in ejection fraction (Qureshi et al., 2017)
Umbilical cord blood-derived MNCs, autologous	NCT03779711	2019-	Intramyocardial injection of autologous UCB-derived MNCs during surgical repair of HLHS; Phase II	Active, not enrolling. Study participants are receiving the cell therapy; no new patients are being recruited or enrolled. Anticipated study completion: 02/2026
Bone marrow-derived MSCs, allogeneic	MedCaP: NCT04236479	2020-	MSC delivery via cardiopulmonary bypass for infants with CHD undergoing two-ventricle repair	Enrolling.

CPC: c-kit+ cardiac derived progenitor cells

CDC: cardiosphere-derived cells

MSC: mesenchymal stem, or stromal, cells

MNC: mononuclear cells

1.2.1 Stem and progenitor cell types for treatment of congenital heart disease

Several cell candidates are under investigation including mesenchymal stem, or stromal, cells (MSCs), mononuclear cells (MNCs), cardiosphere-derived cells (CDCs), and CPCs. These four cell types have differing characteristics which make them attractive in autologous (self) or allogenic (donor-derived) models. First, MSCs offer unique immunological properties that allow them to function as an ‘off-the-shelf’ allogenic product in cell therapy studies. MSCs are immunoprivileged: they lack major histocompatibility complex (MHC) class-II and costimulatory CD80, CD86, and CD40 molecules, and have reduced MHC class-I. In addition to evading the immune system, MSCs also function in an immunosuppressive manner. MSCs secrete anti-inflammatory growth factors and extracellular vesicles (EVs), which decrease various proinflammatory cytokine expression, suppress T helper and T cytotoxic cell

proliferation, and stop B cell and dendritic cell maturation (Karantalis et al., 2015). Furthermore, MSCs can self-replicate and transdifferentiate into various tissue lineages[1] (Quevedo et al., 2009). Our group is involved in the ELPIS trial examining the use of allogeneic bone marrow-derived MSCs during the stage II bi-directional Glenn operation in HLHS patients. The results of phase I of the ELPIS trial are not yet published, but allogeneic MSC use in patients with acute myocardial infarction and chronic ischemic cardiomyopathy did not trigger an immune response and showed promising results for cardiac improvement (Hare et al., 2012; Hare et al., 2009).

Second, autologous MNCs derived from umbilical cord blood and bone marrow have been examined for use in CHD pediatric patients. Notably, umbilical cord blood MNCs are the most prevalent cell source available, are isolated at birth, and can differentiate into non-blood cell types including cardiomyocyte and endothelial cells (Oommen et al., 2015). Recently, MNCs have been the subject of controversy with research suggesting that delivery of these cells induces an inflammatory-based wound healing response, responsible for observed MNC-efficacy[2] (Vagnozzi et al., 2020). Completed phase I trials of MNC treatment for HLHS patients have demonstrated this therapy to be safe and feasible (Burkhart et al., 2019; Qureshi et al., 2017).

Third, autologous CDCs have demonstrated promise in phase I and II trials for single ventricle patients (Ishigami et al., 2017; Ishigami et al., 2015). CDCs were developed in 2007 and are comprised of a heterogenous population of cells derived from cardiac tissue explants.

---

[1] Recognized primary trans differentiation of MSCs include adipocytic, chondrocytic, or osteocytic lineages. Differentiation into cardiomyocytes *in vivo* remains controversial. Malliaras K and Marban E (2011) Cardiac cell therapy: where we've been, where we are, and where we should be headed. *Br Med Bull* **98**:161-185.

[2] CPCs were also implicated in this study but were not included in most timepoints of the ischemia/reperfusion animal study. CPCs were only included in experiments testing differentiation into endothelial cells/cardiomyocytes and for functional effects 2 weeks injury.

CDCs are CD105+/CD45- and multipotent, differentiating into cardiomyocytes, endothelial cells, and smooth muscle cells. The first CDC trial, CADUCEUS, showed autologous CDC delivery were safe, feasible, and regenerative in adult patients with left ventricular dysfunction after myocardial infarction. These results spurred interest and resulted in the TICAP, PERSEUS, and APOLLON trials for single ventricle pediatric patients. A drawback to CDCs is the potential for populations to include fibroblasts. Given the demonstrated regenerative potential of CDCs, current and future work involves optimizing more translatable approaches. Ongoing studies investigating the use of the cell-free alternative, CDC-derived EVs, show promise (Marban, 2018).

Finally, our group is directly involved in the CHILD clinical trial examining use of autologous CPCs for HLHS during the stage II bi-directional Glenn operation. The subsequent section describes the design and rationale of this trial. Overall, CPCs represent another cardiac-derived cell option that is more homogenous than CDCs. CPCs are isolated via c-kit and are self-renewing, endothelial progenitors. Unfortunately, CPC research has been tainted by scientific misconduct and initial research indicating differentiation of CPCs into cardiomyocytes has been retracted (Bolli and Tang, 2022). However, several preclinical and clinical studies continue to demonstrate the therapeutic potential of CPCs and have now attributed their efficacy to the CPCs' secretome (Agarwal et al., 2017; Saha et al., 2019; Sharma et al., 2017; Trac et al., 2019a). Discussed in the following section, CPC-EVs are packaged with pro-reparative molecules and represent an attractive, alternative cell-free therapy.

1.2.2 CHILD clinical trial

As mentioned previously, mortality for HLHS patients remains high and is largely attributed to right ventricle dysfunction. Preclinical rat and swine studies investigating the use of CPCs in a

pulmonary artery banding model—a model of right ventricular dysfunction—have shown that CPCs may attenuate cardiac dysfunction, increase angiogenesis, decrease wall thickness, and reduce fibrosis (Agarwal et al., 2016; Wehman et al., 2017). The CHILD clinical trial (NCT03406884: autologous **c**ardiac stem cell injection in patients with **h**ypoplast**i**c **l**eft heart syn**d**rome) was designed to test the safety, feasibility, and potential efficacy of autologous CPC use in HLHS infants. Of note, the CPCs used in the trial are isolated from otherwise discarded right atrial appendage tissue, collected the stage I Norwood procedure. Phase I of CHILD is now complete, and phase II is currently enrolling HLHS patients to test efficacy of the CPC therapy (Kaushal et al., 2022).

The CHILD trial was designed to leverage (1) the potency of neonatal CPCs, and (2) the adaptive myocardia of young HLHS patients. First, several preclinical studies have demonstrated that CPCs derived from neonate patients (nCPCs) outperform cells derived from adults (aCPCs) (Agarwal et al., 2016; Mishra et al., 2011; Sharma et al., 2017; Simpson et al., 2012). Analysis of the right atrial cardiac tissue show that CPC density declines with age: 9% in neonates and 3% in older children (Mishra et al., 2011). Additionally, nCPCs are shown to be more reparative than aCPCs due to their retention of pluripotent gene expression and enhanced secretion of cardioprotective factors, including EVs (Sharma et al., 2017). In the CHILD trial, biopsies are collected from patients ~ 1 week old, and the cells used in trial may represent the most reparative CPC population possible. Second, HLHS infants may represent an ideal candidate for cardiac cell therapy, as their myocardium may be more adaptive and responsive. Previous research has shown that cardiomyocyte renewal and cell activity reduce with age (Bergmann et al., 2009; Mollova et al., 2013). Therefore, the delivery of nCPCs to HLHS infants may represent an ideal cardiac cell therapy strategy.

For the modeling work described in chapter four, we grew and expanded CPCs from the first seven lead-in phase I patients. We collected EVs from the conditioned media of cultured CPCs and sequenced both the patient-derived CPCs and their respective EVs. For now, we have constructed models combining CPC-EV RNA sequencing data with cardiac-relevant *in vitro* experimental outcomes. In the future, we aim to correlate these *in vitro* outcomes to the clinical cardiac functional outcomes (e.g. right ventricle ejection fraction, stroke volume) to determine the most useful and representative *in vitro* experiments, as well as the EV RNAs driving therapeutic efficacy.

1.3 Extracellular vesicles

The work contained in this section is adapted from our 2020 review published in Nanomedicine.
Bheri S, **Hoffman JR**, Park HJ, Davis ME. Biomimetic nanovesicle design for cardiac tissue repair. Nanomedicine (Lond). 2020 Aug;15(19):1873-1896. doi: 10.2217/nnm-2020-0097.

Extracellular vesicles (EVs) are lipid bilayer vesicles, containing protein and nuclear cargo (Saha et al., 2019). EVs are released from all cell types and the cell source is referred to as the 'parent cell'. The International Society of Extracellular Vesicles classifies EVs on physical characteristics including size, density, and origin (Thery et al., 2018). Exosomes and microvesicles are derived from viable cells, whereas apoptotic vesicles are derived from apoptotic bodies (**Figure** 1-3). Within these classes, there exists large heterogeneity and no current consensus markers.

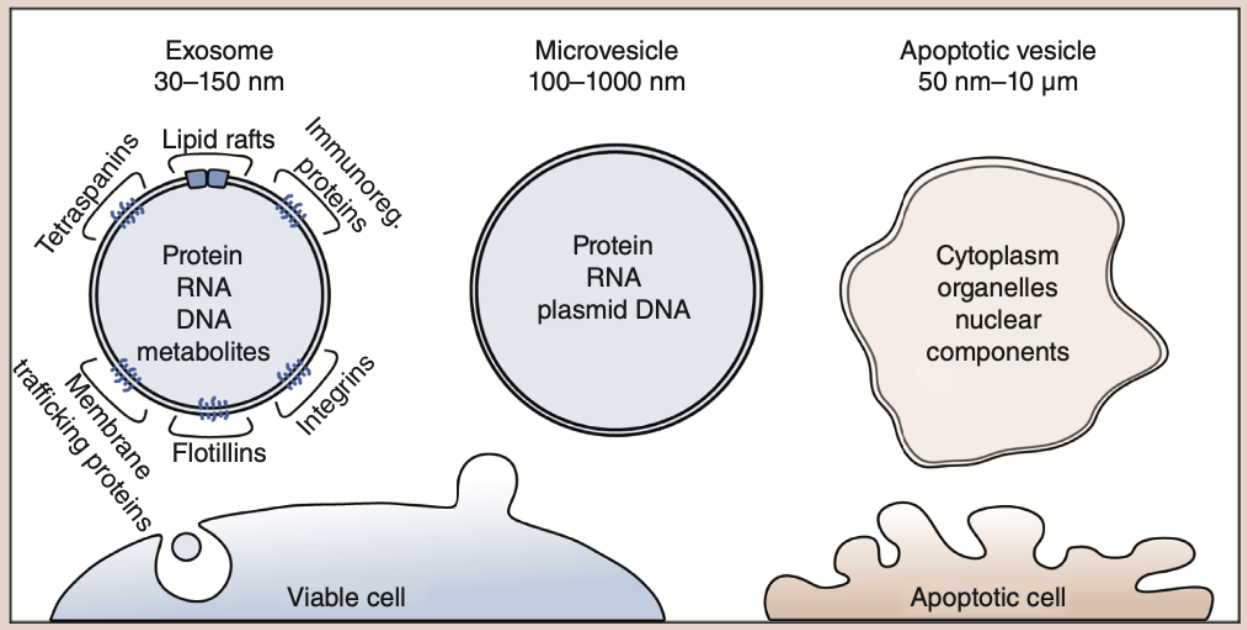


Figure 1-3. Schematic of subpopulations of extracellular vesicles.

Exosomes and microvesicles are derived from viable cells. Apoptotic vesicles are derived from apoptotic cells. Exosomes are the smallest vesicles surrounded by a lipid bilayer membrane with lipid rafts, tetraspanins, immunoregulatory proteins, membrane trafficking proteins, integrins and flotillins embedded in it. Microvesicles are slightly larger than exosomes and apoptotic bodies can have the largest vesicles. Of the three types of extracellular vesicles, exosomes are known to be cardio protective and reparative. (Bheri, Hoffman, Park, Davis, 2020).

Small EVs, historically referred to as exosomes, are less than 150-nm vesicles, contain protein and/or nuclear cargo, and are the most well-studied type of EV. These vesicles form in the cytoplasm from the inward budding of endosomes. They were initially discovered in sheep reticulocytes and were labeled as nothing more than carriers for waste export (Johnstone et al., 1987). After several decades of EV-related research, they have now been accepted as major players in paracrine signaling and potential biomarkers for several disease (Barile et al., 2017a). Small EVs traffic mRNA, miRNA, DNA, and proteins between cells. As cell-free components, these vesicles are often enriched in specific cargo, as compared to their parent cells, allowing EV-based therapies to become a potential alternative to cell therapies (Haraszti et al., 2019).

Next, medium or large EVs, historically referred to as microveicles, range from 100 to 1000 nm in size. These EVs were originally thought to be ‘dust material’ derived from platelets and their role in cellular interaction was only uncovered more recently (Wolf, 1967).

Microvesicles form from the outward budding of a cell's plasma membrane (Williams et al., 2007). Like small EVs, they also play a role in cell-cell communication with the transport of miRNA, mRNA, DNA, and proteins between cells (Camussi et al., 2010). The precise differences between small EVs and microvesicle-mediated transport are still unclear, but being larger in size, microvesicles can successfully carry plasmid DNA (Lamichhane et al., 2015). Despite this, microvesicles remain poorly explored as a therapeutic for tissue repair and regeneration.

Finally, medium or large EVs produced during apoptosis, referred to as apoptotic vesicles, can range in size from 50 nm to 10 μm, depending on the parent cell type. Apoptotic vesicles are formed by indiscriminate blebbing of a cell membrane during apoptosis (Saraste and Pulkki, 2000). Consequently, apoptotic vesicle cargo typically consists of parent cell remnants, including cytoplasm, organelles, and nuclear content. Apoptotic vesicles could play an important role in regulating local disease microenvironments as the frequency of cellular apoptosis is higher in disease states. For example, apoptotic vesicles from mature endothelial cells induced differentiation in endothelial progenitor cells (Hristov et al., 2004). Nonetheless, the detailed role of apoptotic vesicles and their cargo in tissue homeostasis and repair remains unclear.

1.3.1 Biogenesis

Small EV biogenesis and release occurs through a series of endocytic steps: inward budding of the plasma membrane and formation of the early endosome; transformation of the early endosome into the late endosome; inward budding of the late endosome or multivesicular body (MVB); and fusion of the MVB with the plasma membrane for release of small EVs into the extracellular space (**Figure** 1-4). This generalized process results in EV release and subsequent

signaling. However, the MVB may also fuse with lysosome, instead of the plasma membrane, resulting in vesicle degradation (Pan et al., 1985; Tian et al., 2013).

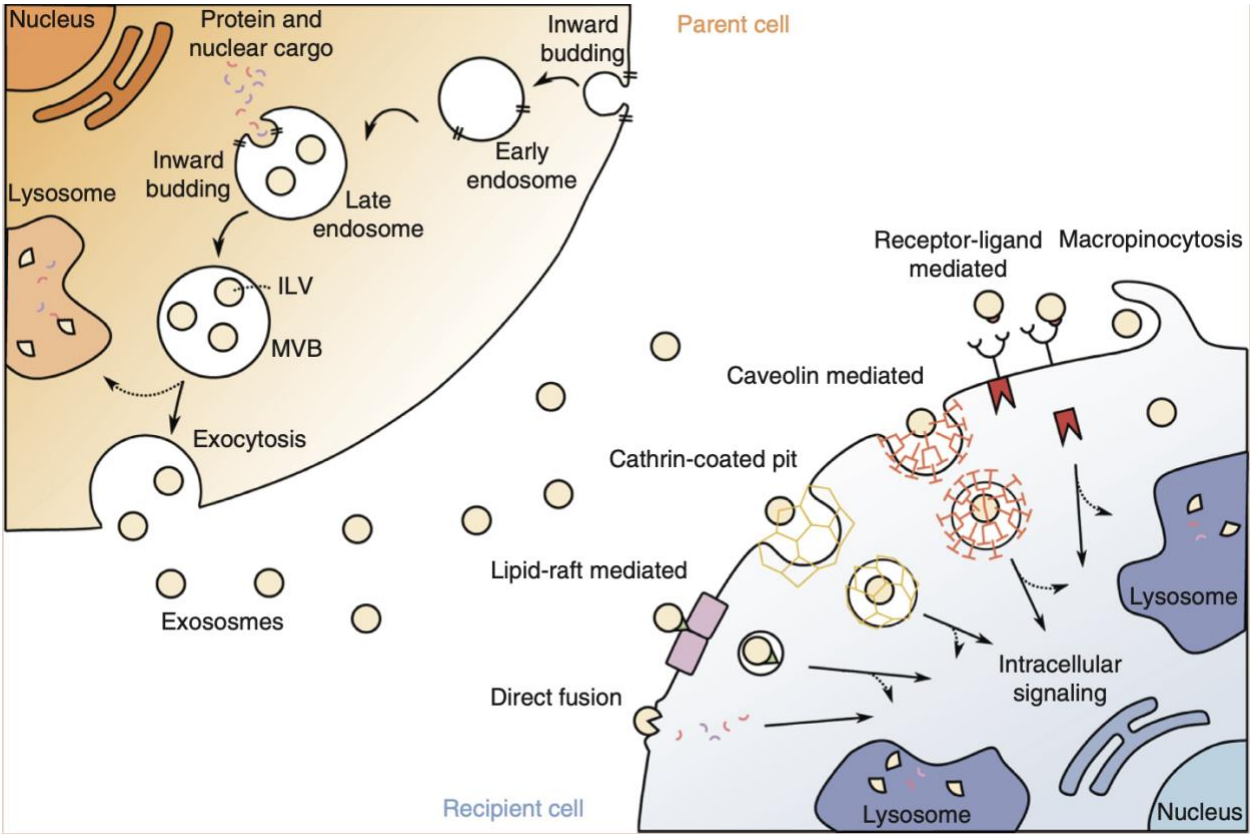


Figure 1-4. Small EV biogenesis and uptake.

Small EV biogenesis may occur in the parent cell (orange) through a series of endocytic steps, which conclude with EV release from the MVB. Several protein complexes (not shown) are involved at each stage of biogenesis (e.g., ESCRT proteins). During inward budding, some membrane proteins and lipids are also incorporated (not shown). Once EVs are released into the extracellular fluid, they get trafficked to the recipient cell (blue). Here, EV uptake occurs through one of many methods: macropinocytosis, receptor–ligand mediated uptake, lipid raft mediated uptake (including clathrin-coated pits and caveolin) or direct fusion. In the recipient cell, the EV is either degraded in the lysosome or activates intracellular signaling. ESCRT: Endosomal sorting complex required for transport; ILV: Intraluminal vesicle; MVB: Multivesicular body. (Bheri, Hoffman, Park, Davis, 2020).

Each step in the biosynthesis of small EVs allows the vesicles to acquire certain cargo and membrane components. EVs contain membrane proteins, including tetraspanins and integrins, that originate from the cell's plasma membrane. When the initial endocytosis of this membrane occurs, these proteins become incorporated into the early endosome and are carried along during the formation of EVs (Stoorvogel et al., 1991). Additionally, during the process of inward budding of the late endosomal membrane, cytosolic proteins, RNAs and other molecules

are encapsulated, forming cargo-loaded ILVs inside the MVB. These ILVs are eventually released as small EVs when the MVB fuses to the cell's plasma membrane (Jaiswal et al., 2002).

Once released into the extracellular space, EVs may be taken up by recipient cells via endocytic processes including lipid-raft-based uptake (including clathrin-coated pits, and caveolin), direct membrane fusion, macropinocytosis or through receptor–ligand interactions (**Figure** 1-4) (Costa Verdera et al., 2017; Hemler, 2003; Svensson et al., 2013; Vinas et al., 2018). Furthermore, there are known membrane proteins that play a role in the uptake process. Tetraspanins are highly enriched in EV membranes and are involved in cell fusion and penetration events (Hemler, 2003). Additionally, tetraspanin-rich domains are implicated in sorting receptors and intracellular molecules into EVs. Heat-shock proteins (including HSP70 and HSP90) are implicated in antigen binding and presentation in exosomes, and annexins and Rab play a part in membrane fusion (Gastpar et al., 2005; Jeppesen et al., 2019).

Finally, once internalized by recipient cells, EVs either remain functional and transfer materials into recipient cells or are shuttled to the lysosome for degradation (Tian et al., 2013). In the former case, EVs fuse to endosomes, allowing for horizontal genetic transfer of the cargo to the recipient cell's cytoplasm (Pegtel et al., 2010). Here, the delivered intravesicular molecules can be involved in epigenetic reprogramming of recipient cells through the delivery of proteins, lipids and RNAs. Nonetheless, the effects of EVs are not limited to their cargo; EVs may also affect recipient cells via receptor–ligand-mediated signaling and the transfer of receptors to the cell surface (Al-Nedawi et al., 2008). It is evident that EVs are highly tuned to be efficient messengers between cells and regulating such nanovesicles could be impactful for cell therapies.

1.3.2 Structure and function

An EV can be split into two main regions: the membrane and the internal core. First, the EV's bilayer membrane is composed predominately of amphiphilic molecules, lipid rafts, and membrane proteins. The lipid classes typically found in an EV membrane include cholesterol, sphingomyelin, phosphatidylcholine, phosphatidylserine, and phosphatidylethanolamine (Llorente et al., 2013). The lipid distribution in the bilayer's inner and outer leaflets is asymmetric with sphingomyelin and phosphatidylcholine enriched in the outer leaflet. The orientation and distribution of the lipids is associated with their packing parameter, dependent on the shape of the fatty acid tails (Llorente et al., 2013). Previous research has shown that membrane lipids play an important role in vesicle trafficking (Huijbregts et al., 2000; Ikonen, 2001). Interestingly, the ratio of lipid classes in EVs is dependent on cellular microenvironment conditions and is different to that of parent cells, suggesting active packing of lipids (Haraszti et al., 2019). The exact rationale and advantage of certain lipid enrichments in EVs remains unclear.

EV proteins may be found on the membrane or in the core. EV membrane proteins include tetraspanins, immunoregulatory proteins, membrane trafficking proteins, integrins, and flotillins. EV core proteins include chaperones, cytoskeleton proteins, enzymes, signal transduction proteins, and exosome biogenesis proteins (Conigliaro et al., 2017). Small EVs formed through the endocytic process acquire general vesicle proteins (EV markers), as well as parent cell-specific proteins (origin markers). Common EV markers include tetraspanins CD9, CD63, and CD81, as well as ALIX, TSG101, and HSP70 (Thery et al., 2018). EV marker proteins may serve important roles in the biogenesis of EVs. ALIX and TSG101 are members of the endosomal sorting complex required for transport, the four-complex mechanism involved

with cargo sorting and EV biogenesis (Li et al., 2018). EV origin markers tend to be lineage and disease-specific, dependent on the proteins expressed on the parent cell. For example, EVs derived from hypoxic cardiomyocytes often include HSP60. Like EV lipids, EV proteins are present in different ratios than parent cells, suggesting selective or active protein loading (Haraszti et al., 2016).

EV nucleic acids may be found in the core. In fact, a large part of exosome cargo consists of nucleic acids such as miRNA, mRNA, long noncoding RNA (lncRNA), and DNA. Nucleic acids are loaded into EVs from the cell cytoplasm during the formation of intraluminal vesicles. EV nucleic cargo composition is highly dependent on parent cell state and milieu (Agarwal et al., 2017; Gray et al., 2015).

1.3.3 Role of stem/progenitor cell-derived extracellular vesicles in cardiac repair

Cell therapy studies have shown that reparative effects may be attributed to paracrine signaling—intercellular communication wherein cell-free components are trafficked to nearby cells (Bao et al., 2017; Barile et al., 2017a; Lee et al., 2012; Ong et al., 2015). In addition to soluble growth factors and cytokines, these signals include EVs. EV cargo is dependent on parent cell type and environment, and stem/progenitor cell-derived EVs have been shown to mimic the tissue reparative effects of their parent cells. Studies have shown that inhibition of EV secretion reduces the effectiveness of cardiac cell therapy (Kishore and Khan, 2016; Lang et al., 2016; Marban, 2018). Combined with the previously mentioned disputes surrounding stem cell differentiation potential, transplanted cells also exhibit low retention. These limitations, combined with the lack of immunogenicity associated with EVs, make stem and progenitor cell-derived EVs an attractive alternative strategy to cardiac cell therapy.

A few studies of stem/progenitor cell EV cargo have identified beneficial molecules and follow up preclinical trials are underway directly overexpressing these RNAs. A handful of EV RNA involved in cardiac repair are listed in **Table** 1-3. Importantly, we and others have shown that it is groups of covarying EV cargo molecules, rather than single molecules, which contribute to repair *in vivo*. Given the inherent heterogeneity of EV cargo and dependency on the parent cell niche, we need more quantitative studies to understand the full landscape of stem/progenitor cell-derived EV-induce cardiac repair.

Table 1-3. Representative EV RNAs involved in cardiac repair.

RNA	Cell source	Outcome
miR-146a	CDC	Promote cardiac regeneration and enhance sytosolic function in mouse model of myocardial infarction (Ibrahim et al., 2014)
miR-21a-5p	MSC	Reduce pro-cell death target genes; cardioprotective (Luther et al., 2018)
miR-125b-5p	MSC	Protection against ischemia-reperfusion injury (Varga et al., 2014)
miR-126 and miR-130a	CD34+ hematopoietic stem cells	Promote angiogenesis (Sahoo et al., 2011)
miR-292	CPC	Anti-fibrotic (Gray et al., 2015)
miR-451	CPCs from cardiospheres	Protection against ischemia-reperfusion injury Reduce fibrotic effect in cardiac fibrosis, promote angiogenesis, increase cardiomyocyte survival (Chen et al., 2013b; Zhang et al., 2010)
PI3K/AKT pathway-associated mRNA	Endothelial progenitor cell	Increased endothelial cell angiogenesis (Deregibus et al., 2007)
MALAT1 lncRNA	UCB-MSC	Reduced age-induced cardiac dysfunction (Zhu et al., 2019)
NEAT1 lncRNA	MSC	Inhibit cardiomyocyte apoptosis (Chen et al., 2020)

1.4 Machine learning – unsupervised vs. supervised learning

An unfortunate consequence of next-generation sequencing (NGS) experiments is that the number of variables (p) far exceeds the number of samples (n). In the context of bulk total RNAseq, >15,000 RNAs are measured for typically a handful of samples. Considering we cannot view or model thousands of dimensions, or variables, machine learning strategies can help us make sense of large data sets, reducing the complexity so that we may visualize and interpret results. Fortunately, the collinearity of variables, inherent to sets as large at RNAseq data, allow one to implement machine learning techniques. The overall goal of machine learning is to build or train models to make decisions without being explicitly being told to do so. There are two main types of machine learning methods: unsupervised learning and supervised learning (**Figure** 1-5)[3]. Unsupervised learning involves learning patterns and structure from unlabeled, or uncategorized, data. Supervised learning involves learning functions that map inputs to outputs based on labeled data.

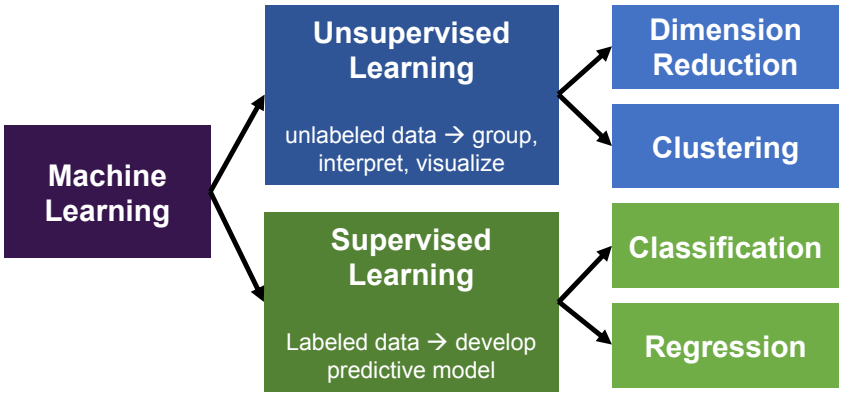


Figure 1-5. Machine learning: unsupervised and supervised learning.

Unsupervised learning strategies utilize unlabeled data and identify structures and patterns. Supervised learning strategies utilize labeled data and generate predictions.

[3] Two other types include semi-supervised learning and reinforcement learning. As the name suggests, semi-supervised learning (ex. low density separation models) combines unlabeled data with a small amount of labeled data, a mixture of unsupervised and supervised learning. In reinforcement learning (ex. Monte Carlo methods, Heuristic methods) agents take action to maximize reward in a feedback system.

1.4.1 Dimension reduction and clustering

Unsupervised learning can be broken into two main forms: dimension reduction and clustering. A main goal of dimension reduction strategies is to project variables into lower dimensional space, either transforming data in a linear or nonlinear manner. The most commonly used method in bioinformatics is principal component analysis (PCA). In PCA, variables are linearly mapped into lower dimensional space to describe, or represent, the greatest amount of variance. The principal components are eigenvectors of the covariance matrix—the square matrix holding the joint variability between each pair of variables. The PCA algorithm computes components orthogonal to the previous component, thereby maximizing the variance described by each subsequent component. In bioinformatics, data are typically displayed in 2 or 3 dimensional plots with each axis representing a principal component. PCA allow us to visualize the similarity of samples as well as determine the "weight" or correlation of each variable and each principal component. Other examples of dimensionality reduction methods include t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and project (UMAP), nonlinear techniques often used in single cell RNA sequencing analyses.

Clustering involves grouping similar objects. In bioinformatics, clustering algorithms can reveal co-expressed genes or identify similar samples. Some clustering algorithms include hierarchical clustering, k-means or centroid-based clustering, distribution-based clustering, and density-based clustering. As the name suggests, hierarchical clustering seeks to build a hierarchy of clusters. In agglomerative "bottom-up" approaches, items start in their own cluster and pairs of clusters are merged. In divisive "top-down" approaches, all items begin in one cluster and clusters are split recursively. Hierarchical clustering can be used in pathway analyses to understand gene set enrichment in similar gene ontology terms (Harris et al., 2004). K-means

clustering involves randomly setting k-number of central vectors, assigning items to a center, and iteratively updating the centers to minimize the distances of items. In k-means clustering, users may set the number of clusters to identify in an RNA sequencing data set and determine k-many clusters of similarly expressed genes (Kolde, 2012). Finally, shared nearest neighbor clustering is an example of density-based clustering algorithm. Shared nearest neighbor clustering is implemented in the Seurat package, used to identify cell clusters in single cell RNA sequencing data (Satija et al., 2015).

1.4.2 Weighted correlation network analysis

Weighted correlation network analysis (WGCNA) is an unsupervised learning algorithm originally designed to provide a user-friendly correlation network method. WGCNA has many applications, but the focus of this approach is the construction of co-expression gene networks. WGCNA was written as an R package and designed for use in microarray experiments that suffer from the p>n problem (Langfelder and Horvath, 2008). WGCNA achieves multiple goals, it (1) finds clusters of highly correlated genes, and (2) summarizes clusters with eigengenes, or hub genes, to (3) relate clusters to sample traits (e.g. disease status, age, survival time, etc.) and (4) relate clusters to each other.

WGCNA begins by computing the adjacency matrix of the dataset and then transforming it to form the topological overlap matrix (TOM). Then, the corresponding dissimilarity matrix, 1-TOM, is used to perform hierarchical clustering and detect co-expression clusters. Cluster eigengenes are computed and used to test correlation to sample traits. Unlike traditional differential expression analysis, WGCNA is free from the multiple comparisons problem: the approach forms a discrete number of clusters before testing for eigengene significance, or the correlation between clusters and sample traits. For example, in chapter four, we constructed a

network and bin >12,000 RNAs into 33 clusters. These clusters were then tested for correlation to *in vitro* outcomes. Once a network is formed and clusters are determined, gene clusters may then be related to biological processes (e.g. gene ontology) with pathway analysis.

1.4.3 Regularized regression models

Supervised learning can be broken into classification and regression-type problems. In classification tasks, models assign labels to a set of data. A classic example includes identifying healthy controls vs. diseased samples. Classification algorithms include decision trees, naïve Bayes classifiers, and logistic regression, among others. For the purposes of this work, I will pay attention to regression methods to estimate relations between RNA sequencing data and experimental or clinical outcomes. For example, this dissertation is concerned with estimating improvements in patient ejection fraction rather than classifying patients. In fact, phase I of the CHILD clinical trial does not include a control patient group.

Within regression algorithms, the simplest type, ordinary least-squares linear regression, is ill-suited for RNA sequencing data sets given that $p>n$; this strategy yields infinitely many solutions! To address this issue, one may implement regularized, or penalized, regression models, including least absolute selection and shrinkage (lasso) regression, ridge regression, and elastic net regression. Regularized regression involves invoking a penalty to minimize the weight of variables. All regularized regression models use hyperparameters to control the learning process. Hyperparameters are set a priori by the user but may be tuned using cross validation to minimize error.

Lasso regression shrinks variable coefficients towards 0 and allows for coefficients to equal 0. Variables with a coefficient equal to 0 are dropped from the multiple regression model, and in this way, lasso performs feature selection and often creates sparse models. Lasso uses L1

regularization; it adds a penalty that is equal to the absolute value of the magnitude of the coefficients. The hyperparameter used in lasso regression, λ, dictates the amount of shrinkage. When $\lambda = 0$, all variables are considered, and linear regression is performed. When $\lambda \rightarrow \infty$, all coefficients are set to 0 and no variables are considered. As λ increases, bias increases (under-fitting); As λ decreases, variance increases (over-fitting). As with all regularized regression models, the goal is to find an optimized value for λ that balances bias and variance. Lasso regression is a particularly useful approach for problems with only a few informative or predictive variables, cases where models will benefit from dropping variables entirely. However, a considerable drawback of this method is that it in problems with collinear variable sets, lasso will retain one variable and drop the others. In this way, information may be lost. Pairing lasso with other methods like WGCNA and ridge regression helps bolster and validate lasso results.

Ridge regression shrinks variable coefficients towards 0 as well, but coefficients cannot equal 0. In ridge regression, variables are not dropped, extreme coefficients are only minimized. Ridge regression uses L2 regularization which adds a penalty equal to the square of the magnitude of the coefficients. Ridge regression also has one hyperparameter, λ, controlling shrinkage in a similar manner as lasso. Ridge regression is useful when predicting power is spread over many variables (typically > n). In this situation, explained variance is not lost by dropping variables.

Finally, elastic net regression combines lasso and ridge, using both L1 and L2 regularization. Elastic net has two hyperparameters: λ controls shrinkage (similar to lasso and ridge), and α controls the ratio of L1 and L2 penalties. When $\alpha = 0$, only L2/ridge regression is used. When $\alpha = 1$, only L1/lasso is used. Elastic net is a more complex algorithm and thus more

computationally intensive. Elastic net requires testing many combinations of λ and α to minimize error.

1.4.4 Partial least squares regression

Partial least squares regression (PLSR) is a method that achieves both dimension reduction and regression. Using a latent variable approach, PLSR projects both the X and Y data into new spaces to model the covariance between the two. Simply put, PLSR creates components in X space that maximize the variance in the Y space. In this way, PLSR can be thought as a hybrid of multiple linear regression (correlating X and Y) and PCA (capturing maximum X variance). This method is well suited for bioinformatics due to the dimension reduction aspect of the algorithm: X variables in RNAseq data are highly collinear and are captured well in lower dimensional space. The general form for PLS is written in below:

$$\mathrm{X} = \mathrm{TP}^{T} + \mathrm{E}_{X}$$

$$\mathrm{Y} = \mathrm{UQ}^{T} + E_{Y}$$

In these equations, X is a matrix of predictor variables of size *n x p* (samples x RNA), and Y is a matrix of response variables of size *n x m* (samples x outcomes/y). E represents the error term for X and Y. P and Q are the loading matrices for X and Y. T and U are the score matrices (or latent variables) of X and Y. T and U can be written as follows:

$$\mathrm{T} = \mathrm{XW}$$

$$\mathrm{U} = \mathrm{YC}$$

Here, W and C are the weight matrices that can be computed by singular value decomposition (a quick and efficient method to factorize matrices). The goal is to maximize to the covariance of T and U, or the scores of X and Y. There are a few different algorithms for

PLSR including, nonlinear iterative partial least squares (NIPALS), orthogonal projections to latent structures, and the statistically inspired modification of partial least squares (SIMPLS). The original algorithm, NIPALS is numerical and requires matrix deflation, a computationally intensive technique whereby the covariance matrix is modified to remove eigenvectors. On the other hand, the SIMPLS algorithm is a statistical method (similar to PCA) with the objective to maximize covariance. SIMPLS does not require matrix deflation and derives T directly as linear combinations of X variables. SIMPLS solutions are merely linear combinations of the original variables and are thus more easily interpretable (De Jong, 1993). The work presented in this dissertation uses SIMPLS as it handles multivariate Y's and is the most efficient.

As with other regression models PLSR has a hyperparameter to tune: the number of components in the model. Selecting the number of components is a matter of balancing performance with complexity: the best, most robust models do not over nor underfit! Identifying the number of components to include can be achieved by using cross-validation techniques and examining error measurements, like root-mean-square error (RMSE).

1.4.5 Random forest regression

Random forest is an ensemble machine learning method[4] that can perform classification and regression. Random forests are constructed from several decision trees, a method with a flowchart-like structure which maps decisions and outcomes. The basic steps of random forest include (1) random sampling features (RNAs) and data points (experimental samples), (2) constructing decision trees for each random sample and computing predictions for each, and (3) aggregating results to take the mean or median of the tree outputs (**Figure** 1-6). By aggregating

[4] As the name suggests, ensemble methods combine multiple models to improve predictability and produce one optimal model.

several models, random forest constructs a better model with less bias. Random forest can also model nonlinear relationships, a distinct advantage over the aforementioned regularized regression methods.

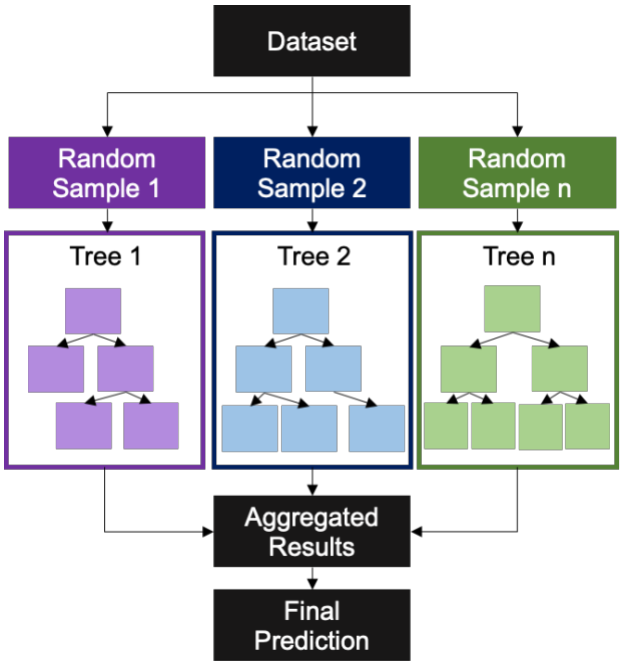


Figure 1-6. Random forest diagram.
In the ensemble learning technique, random forest regression, decision trees are generated from random samples of the dataset. Decision trees are ultimately aggregated to form final predictions.

Nevertheless, there are drawbacks associated with this method. Random forest is a "blackbox" approach that gives reasonable predictions but has little interpretability. Additionally, random forest has several hyperparameters which need to be tuned with cross validation. The number of variables randomly sampled at each tree node (denoted 'mtry' in the R ranger package) is one of the most important parameters to tune. As a rule of thumb, this parameter may be set to the square root or one-third of the number of total features for classification problems and regression problems, respectively. For noisy data, this parameter may be set higher to increase the probability of sampling an informative variable that is predictable; Increasing this parameter decreases bias. However, as with all hyperparameter tuning, increasing this parameter

(and decreasing bias) also increases variance. Additional hyperparameters involved in random forest include the number of decision trees, the maximum depth of the individual trees, criteria to split trees (e.g. error estimates), etc. Although random forest contains many more hyperparameters, it is generally resilient to over-fitting due to the ensemble nature of aggregating many models.

1.5 Research objectives

The purpose of this dissertation research was to investigate patient variability of CPCs for the improvement of cardiac cell therapy. Previous research has shown that CPC reparative function may be attributed to the release of EVs. Here, we explore the two main sources of variability, likely responsible for mixed clinical results: CPCs and CPC-derived EVs (CPC-EVs). We analyzed transcriptomic data (bulk and single cell) and use machine learning regression models to link our RNAseq data to functional outcomes. The subsequent dissertation chapters are designed to explore (1) differences between neonate- and child-derived CPCs at the single cell level, (2) bulk transcriptomic differences between patient matched CPCs and CPC-EVs, and (3) the relationship between CPC-EV cargo and *in vitro*, cardiac-relevant outcomes.

Chapter 2 is a submitted manuscript, co-authored with Arun Jayaraman, which identifies differences between neonate- and child-derived CPCs with scRNAseq. In this work, we observed a more heterogenous population in child samples and uncovered pro-fibrotic and inflammatory cell clusters, enriched with child CPCs. We identified markers of the pro-fibrotic cell cluster and validated their expression with flow cytometry. Chapter 3 is a published manuscript in *Genomics* that investigates the differences between patient matched CPCs and CPC-EVs. In this work, we identified miRNAs involved in cardiac development and cell proliferation, which may be selectively exported to EVs. We used data mining to compare enriched CPC-EV miRNAs to

EVs derived from other cell types and distinguish generic EV-biogenesis related miRNAs from potentially pro-reparative miRNAs. We also examined bulk transcriptomic differences in neonate- and child- derived CPCs with competitive endogenous RNA networks and discovered overall enrichment of non-coding RNAs in child CPCs. Finally, chapter 4 is a manuscript preparation which uses various machine learning regression methods to link CPC-EV RNAseq to experimental outcomes. This work contains samples from the CHILD clinical trial and the overall goal is build an *in vitro* model, predictive of clinical outcomes.

Chapter 2: Single cell RNA sequencing reveals distinct c-kit+ progenitor cell populations

Arun R. Jayaraman*[1], Jessica R. Hoffman*[1,2], Michael E. Davis[1,2,3]

* co-first authors

[1]Wallace H. Coulter Department of Biomedical Engineering, Emory University School of Medicine and Georgia Institute of Technology, Atlanta, GA, USA.

[2]Molecular and Systems Pharmacology Graduate Training Program, Graduate Division of Biological and Biomedical Sciences, Laney Graduate School, Emory University, Atlanta, GA, USA.

[3]Children's Heart Research and Outcomes (HeRO) Center, Children's Healthcare of Atlanta and Emory University, Atlanta, GA, USA.

The work contained in this chapter was submitted for publication in June 2022.

2.1 Abstract

Human c-kit+ cardiac progenitor cells (CPCs) have demonstrated efficacy in preclinical trials for the treatment of heart failure and myocardial dysfunction. Unfortunately, large variability in patient outcomes and cell populations remains a problem. Previous research has demonstrated that the reparative capacity of CPCs may be linked to the age of the cells: CPCs derived from neonate patients increase cardiac function and reduce fibrosis. However, age-dependent differences between CPC populations have primarily been explored with bulk sequencing methods. In this work, we hypothesized that differences in CPC populations, and subsequent cell therapy outcomes, may arise from differing cell subtypes within donor CPC samples. We performed single cell RNA-sequencing on four neonatal CPC (nCPC) and five child CPC (cCPC) samples. Subcluster analysis revealed cCPC-enriched clusters upregulated in several fibrosis- and immune response-related genes. Module-based analysis identified upregulation of chemotaxis and ribosomal activity related genes in nCPCs and upregulation of immune response and fiber synthesis genes in cCPCs. Further, we identified versican and integrin alpha 2 as

potential markers for a fibrotic cell subtype. By investigating differences in patient-derived CPC populations at the single-cell level, this research aims to identify and characterize CPC subtypes to better optimize CPC based therapy and improve patient outcomes.

2.2 Introduction

Cell therapy has emerged as a promising therapeutic strategy for the treatment of disease, including auto-immune disease, blood disorders, cancer, neurodegenerative disease and cardiovascular disease. Various tissue-specific cells, blood cells, and stem cells have been clinically approved, including the use of autologous mesenchymal stem (or stromal) cells for acute myocardial infarction. Unfortunately, cell therapy has been hampered by mixed results, in part due to high cell heterogeneity. Unlike small molecule drugs, cells are highly variable, adaptive to biological cues, and complex in their mechanisms of action. Inconsistencies in cell therapy trials may be explained by batch-to-batch or patient-to-patient variation. We and others have demonstrated that cell donor characteristics affect the phenotype and resulting efficacy of cells. Age and disease have been shown to negatively impact cell efficacy, reducing the effectiveness of cardiac-derived progenitor cells (Agarwal et al., 2016; Sharma et al., 2017; Shoja-Taheri et al., 2019), adipose stem cells (Efimenko et al., 2014; Jumabay et al., 2015), and mesenchymal stem cells (Fan et al., 2010; Khong et al., 2019; Kim et al., 2015; Stolzing et al., 2008), among others (Vasa et al., 2001). Given, the high heterogeneity of cell populations, emphasis has been placed on identifying mechanisms of repair and markers of "good" cells. Recent studies have leveraged single cell RNA-sequencing to identify subpopulations of cells which may be driving therapeutic efficacy. By identifying potential cell surface markers of reparative cells, researchers will be able to isolate and/or enrich for optimal cell populations.

Currently, autologous human cardiac-derived c-kit+ stromal cells (CPCs) are under investigation in the CHILD clinical trial for the treatment of hypoplastic left heart syndrome, a critical congenital heart disease (NCT03406884).(Kaushal et al., 2022) Preclinical results indicate CPCs induce repair in damaged myocardium.(Agarwal et al., 2016; Saha et al., 2019; Sharma et al., 2017; Tang et al., 2010) Additionally, recent results from the phase II CONCERT-HF trial (NCT02501811) suggest that a combination of CPCs and MSCs improve clinical outcomes from patients with ischemic heart failure.(Bolli et al., 2021) Previous research investigating CPC heterogeneity has demonstrated that cell culture conditions (e.g. hypoxia (Hernandez et al., 2018; Tang et al., 2009; Yan et al., 2012) and cell aggregation (Trac et al., 2019b)), as well as donor age (Agarwal et al., 2016) and disease status affect CPC composition and therapeutic potential. We and others have shown that CPC reparative outcomes may be linked to age: CPCs derived from neonate patients (nCPCs) outperform cells derived from older patients (Agarwal et al., 2016; Fuentes et al., 2013; Hernandez et al., 2018; Mishra et al., 2011; Shoja-Taheri et al., 2019). For example, nCPCs possess greater anti-fibrotic signaling, reduced immune response, and increased chemotaxis capabilities, in comparison to child CPCs (cCPCs)(Agarwal et al., 2016).

Nevertheless, differences among CPC populations have been primarily investigated using bulk sequencing methods which treat patient-derived cells as a homogenous sample. Here, we hypothesized that variance in patient outcomes may be driven by differences in cell subtypes or subpopulations, and that CPCs transition to reduced reparative states as patients age. To address this hypothesis, we used single-cell RNA sequencing to (1) identify potentially phenotypically different cell subpopulations and (2) map transcriptomic trajectories of cells from CPCs of neonate (n=4) and child (n=5) congenital heart disease patients as shown in **Figure** 2-1. Overall,

we uncovered a more heterogenous cell population among older patient samples and identified fibrotic and inflammatory cell subpopulations within these samples, which may explain differences in therapeutic outcomes. We identified and confirmed surface expression of versican and integrin subunit alpha 2, markers of a fibrotic cell subpopulation. Furthermore, our analyses unveiled major differences and traced trajectories from cells belonging to a fibrotic cell cluster to cells belonging to clusters enriched in cell cycle and cell proliferation processes. Ultimately, by identifying and selecting for pro-reparative CPC populations, it may be possible to improve therapeutic outcomes.

2.3 Methods

2.3.1 C-kit+ progenitor cell (CPC) culture and expansion

Cells collected from the right atrial appendage of five neonatal (< 1 month) and five child (3.43 years ± 2.6 years) patients with congenital heart disease were separated for c-kit+ CPCs using magnetic cell sorting. Cells were cultured in Ham's F-12 medium (Corning Cellgro®, Corning, NY, USA) with 10% fetal bovine serum, 1% penicillin-streptomycin, 1% L-glutamine and 0.04% human fibroblast growth factor-β. Cells were used between passage 5 and 15. To identify the functional differences between neonate and child CPCs, sorted cells were expanded in culture and submitted for single cell RNA sequencing (10x Genomics). Patient characteristics for samples used in the study are listed in **Supplemental Table 2-1**.

2.3.2 Cell sorting of CPC subpopulation

Flow cytometry was utilized for characterization of CSC subpopulations based on expression of the versican and integrin alpha 2 surface proteins. Anti-versican (Creative Biolabs, CBMAB-C9301-LY) and anti-integrin alpha 2 (R&D Systems, FAB1233P) antibodies conjugated to

Alexa Fluor 647 and PE, respectively, were selected for analysis. Zombie Yellow™ dye (Biolegend) was used to assess cell viability. Pooled child CSCs from patients 926, 938, and 902 (4-year-old patient with atrial septal defect) were cultured, collected, resuspended in flow buffer (2% FBS in PBS), and stained at room temperature, using the manufacturer's suggested concentration. Samples (including unstained and single stain controls) were analyzed with an Aurora Flow Cytometer (Cytek). The compensations were computed using FlowJo software.

2.3.3 Computational Methods

Raw reads from single cell sequencing were processed using CellRanger (10x Genomics, v6.0.0) (Zheng et al., 2017). The doublets were filtered using Scrublet and Scanpy and the raw counts data was processed using the Seurat package in R (Hao et al., 2021; Wolf et al., 2018; Wolock et al., 2019). Cells with 1,000-7,000 distinctly expressed genes and mitochondrial gene fraction totalling <5% of total transcript counts were kept. One neonatal patient sample (Patient 985) was removed due to low transcript counts (<4,000) and a small number of distinctly expressed genes (<2,000).

Data from patient samples were integrated by first normalizing counts using the SCTransform method with variation due to mitochondrial gene fraction regressed out of the datasets (Hafemeister and Satija, 2019). The datasets were combined 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.

Trajectories were constructed using Monocle 3 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.

The dataset was batch corrected using the Batchelor alignment methodology implemented in Monocle. Co-expression of genes was computed along trajectories using the Moran's I statistic as implemented in Monocle and highly co-expressed genes with a q-value <0.05 were clustered into 21 gene modules using the Leiden community detection algorithm (Cao et al., 2019; Haghverdi et al., 2018; Levine et al., 2015; Qiu et al., 2017; Traag et al., 2019; Trapnell et al., 2014). A summary of the analysis pipeline is shown in **Figure** 2-1.

Surface proteins were identified using the cell surface protein atlas validated surface proteomes dataset (Bausch-Fluck et al., 2015). 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 (CD) proteins for better identification of cell surface proteins. The dataset was analyzed for differentially expressed genes that are conserved across donor samples within 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.

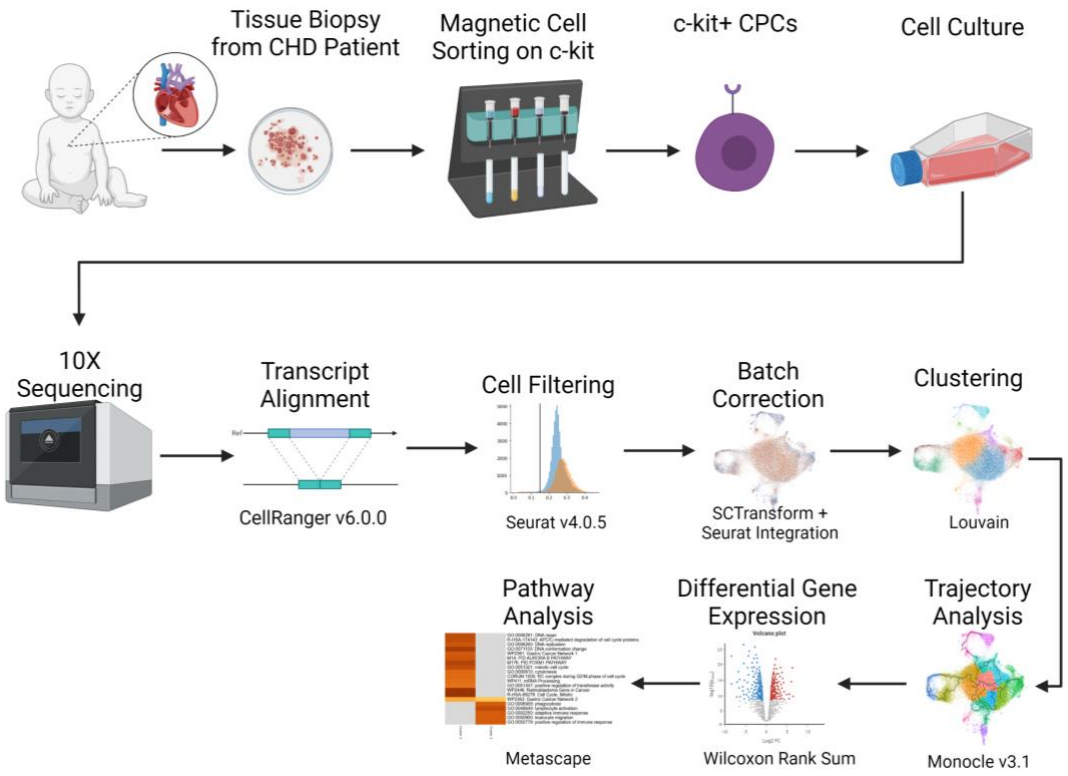


Figure 2-1. Single-cell analysis pipeline.

Top sequence: CPCs were isolated via-kit+ magnetic sorting from tissue biopsies of CHD patients and grown in 2D culture. Bottom sequence: Live cells were analyzed on the 10X Genomics platform. Single cell RNA-sequencing data were filtered, batch corrected, and clustered. Downstream, cells were analyzed with trajectory, differential expression, and pathway analysis tools. Figure generated using BioRender. CPC: c-kit+ cardiac progenitor cell; CHD: congenital heart disease.

2.4 Results

2.4.1 Clustering and compositional analysis reveal differences in neonate and child CPCs

Initial clustering with Louvain identified twelve CPC subpopulations (**Figure** 2-2 **a and b**). Neonate-derived samples were largely enriched in clusters 0 and 1, while child-derived samples were enriched in clusters 3, 6, 8, and 9 (**Figure** 2-2 **c and d**). Notably, we observed a higher level of sample-to-sample variability in child-derived samples. Patients 896 and 926 possessed a more neonate-like clustering profile, whereas patients 938, 1048, and 1092 produced a more dissimilar clustering profile with less cells represented in clusters 0 and 1 and more cells represented in clusters 3, 6, 8, and 9 (**Figure** 2-2 **d**). Pathway analysis indicated that nCPC-enriched clusters 0 and 1 are involved in tissue reparative processes with upregulated genes

connected to angiogenesis and fiber organization (**Figure** 2-2 **e**). The cCPC-enriched cluster 6 is related to supramolecular fiber organization. Finally, clusters 2 and 5 (representative of both nCPCs and cCPCs) are highly enriched in cell proliferation and cell cycle processes.

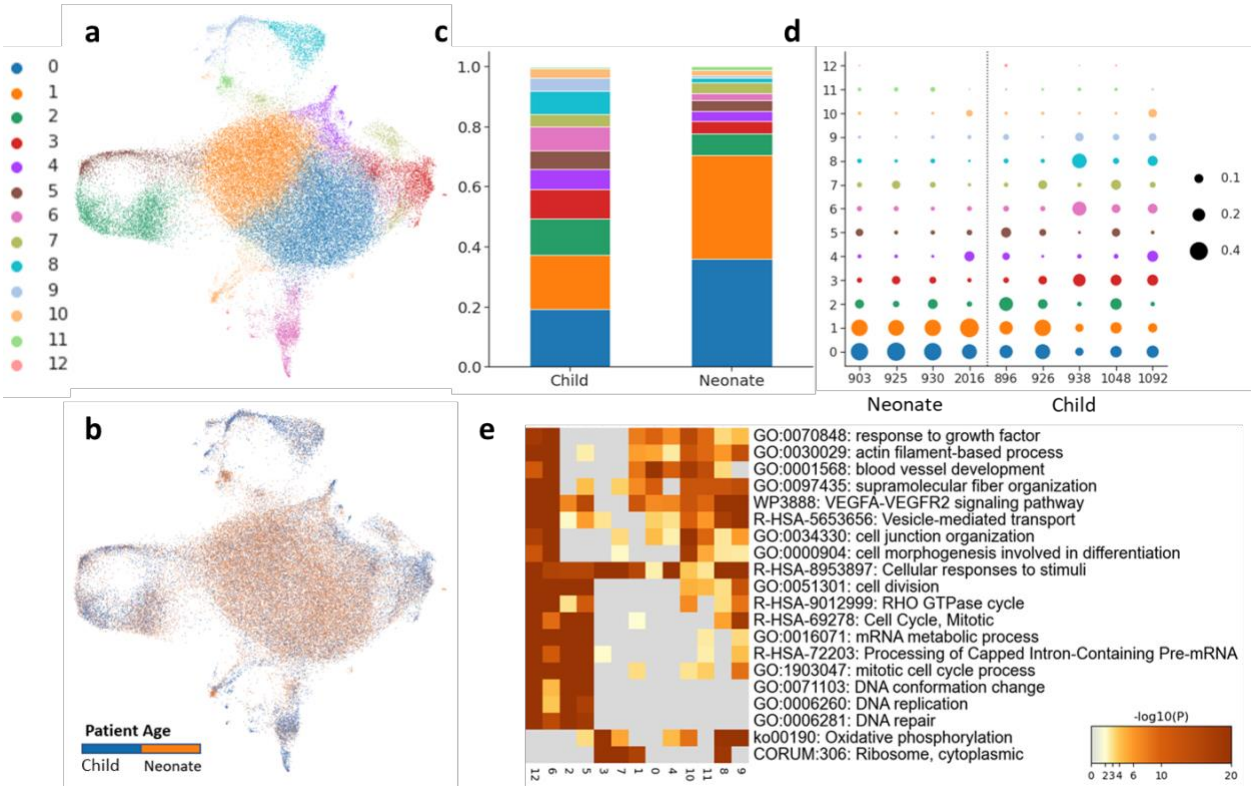


Figure 2-2. Cluster compositions of nCPCs and cCPCs.

UMAP projections display all patient-derived CPCs colored by (**a**) cell cluster and (**b**) age group in 2-dimmensional space. Cell clustering was determined by the Louvain community finding algorithm. Cluster composition as grouped by (**c**) age group and (**d**) patient sample show differential composition of cell subpopulations. (**e**) Pathway analysis of upregulated genes in each cluster shows enrichment of growth factor, blood vessel development, cell cycle, and vesicle transport-related pathways. CPCs: c-kit+ cardiac progenitor cell; UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction

2.4.2 Trajectory analysis identifies co-expressed genes within CPC populations

To understand how transcriptomic profiles change as cells move between CPC subpopulations, we performed trajectory analysis with Monocle 3 (**Figure** 2-3 **a**). We computed psudotimes using various clusters as the starting, or root, node. Notably, pseudotimes computed using cluster 2 cells (enriched in proliferative and cell cycle processes) as the root node resulted in the highest pseudotimes in cluster 8 cells (enriched in processes associated with oxidative stress and

stimuli), indicating the transcriptomic profiles of these cells to be the most distinct from the cluster 2 cells (**Figure** 2-3 **b**). Alternatively, pseudotimes computed using cluster 6 (enriched in fiber organization) as the root node resulted in the largest pseudotimes at cluster 2 (Error! Reference source not found.**b**).

Next, to relate the previously determined cell clusters to gene sets, we computed co-expressed gene modules from our trajectory analysis. Co-expression of genes were computed along trajectories categorized based on the Moran's I statistic computed in Monocle, where a higher value indicates a higher level of co-expression with cells in similar positions of the trajectory. Highly co-expressed genes were clustered using the Leiden algorithm into 21 gene modules (**Figure** 2-3 **c**). Some modules corresponded strongly with certain cell 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.

Relating gene modules to CPC age groups, we determined that nCPCs are highly upregulated in genes belonging to modules 9, 13, 14, and 21 and cCPCs are upregulated in genes belonging to modules 1, 3, 8, 15, and 16. Pathway analysis of module 13 and 21 genes (upregulated in nCPCs) indicates enrichment of pathways related to small molecule biosynthesis and ribosomal activity, respectively (**Figure** 2-3 **d and e**). Module 8 genes (upregulated in cCPCs) contained several immune-related cytokines including *IL6* and *IL1B*. Additionally, module 9 contained notable gene members *CD34* and *PDGFB* and module 9 contained *FGF2*. Across both age groups, modules 8 and 9 contained genes associated with chemotaxis and modules 3, 9, 13, and 15 contained genes linked to ECM organization.

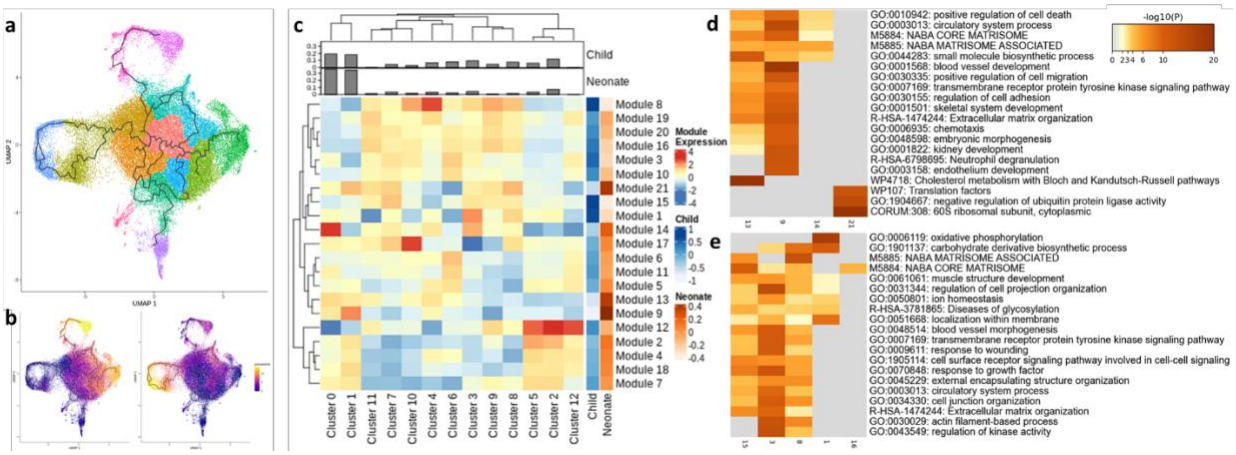


Figure 2-3. Trajectory analysis and gene modules.

UMAP projections with trajectories determined by Monocle colored by (**a**) monocle clusters and (**b**) pseudotime with root nodes set to cluster 2 (left) and cluster 6 (right) from the Seurat analysis. (**c**) gene module expression heatmap by Seurat cluster. Modules were determined through a Leiden clustering of highly co-expressed genes along trajectories. Module expression levels computed by age group and are shown in the two rightmost columns in the heatmap. Cluster proportions by age group illustrated by bar charts on top of the heatmap. Pathway analysis for highest expressing modules among (**d**) neonates and (**e**) children highlights important biological differences between the age groups. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

2.4.3 Cell cluster four is upregulated in cytokines

Differential gene expression analysis identified several cytokines upregulated in cluster 4 cells such as *IL1B, IL6,* and *IL33* (**Figure** 2-4 **a**). Pathway analysis indicates strong enrichment of immune-related signalling pathways, including genes involved in the IL-10 and IL-17 signalling pathways (**Figure** 2-4 **b**). In addition, cluster 4 is enriched in apoptotic signalling and negative regulation of cell proliferation 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. This cell subpopulation is enriched in cCPCs; however, analysis of donor-specific clustering profiles indicates one nCPC sample (Patient 2016) has a high proportion of these cells (**Figure** 2-2 **d**).

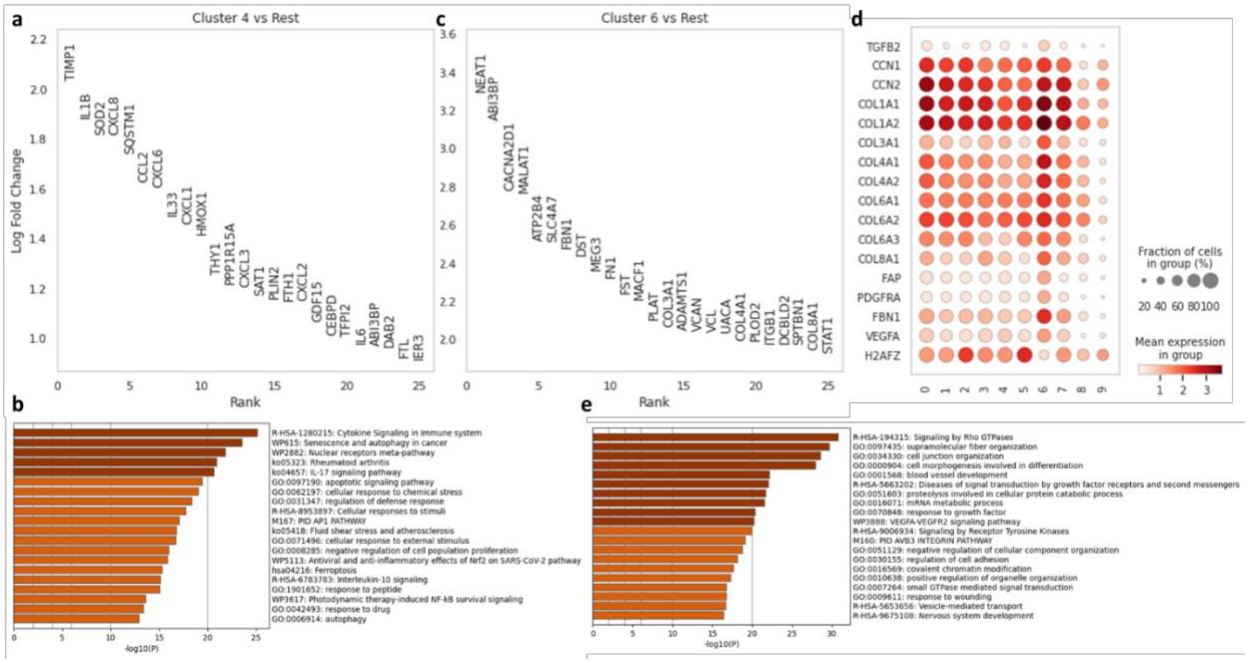


Figure 2-4. Characteristics of clusters 4 and 6.

(**a**) Top 25 differentially expressed genes ordered by log fold change between cluster 4 cells and non-cluster 4 cells. (**b**) Pathway analysis barplot of upregulated differentially expressed genes for cluster 4 cells. (**c**) Top 25 differentially expressed genes ordered by log fold change between cluster 6 cells and non-cluster 6 cells. (**d**) Dotplot of selected genes relating to fibrosis, angiogenesis, and proliferation. (**e**) Pathway analysis barplot of upregulated differentially expressed genes for cluster 6 cells.

2.4.4 Cell cluster six 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. This population is enriched in cCPCs and has high expression of several different types of collagen and genes associated with fibrosis including *TGFB2*, *CCN1*, *CCN2*, and *FBN1* (**Figure** 2-4 **c and d**) (Bouzeghrane et al., 2005; Ding et al., 2020). In addition, the upregulated genes *PDGFRA* and *FAP* are known fibroblast markers and correlate with an epithelial-to-mesenchymal transition (Jechlinger et al., 2006; Kahounova et al., 2018). 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 (**Figure** 2-4 **e**). Pathways associated

with TGF-β signaling were also enriched in this cluster (not shown). Differentially expressed genes associated with this cluster co-cluster heavily within gene modules 3 and 6.

2.4.5 Confirmation of cluster six surface proteins

Differentially expressed surface proteins were selected using the cell surface protein atlas database (**Figure** 2-5 **a**) (Bausch-Fluck et al., 2015). Primary anti-versican and anti-ITGA2 antibodies were selected for characterization of pooled child CSC subpopulations using flow cytometry. The child CSCs were pooled previously and are comprised of patients 926, 938, and 902 (4-year-old patient with atrial septal defect). Cell viability was confirmed with a Zombie Yellow™ dye (85.7%, data not shown). A subpopulation of cells with high versican and ITGA2 expression make up approximately 14% of the pooled child CSCs (**Figure** 2-5 b).

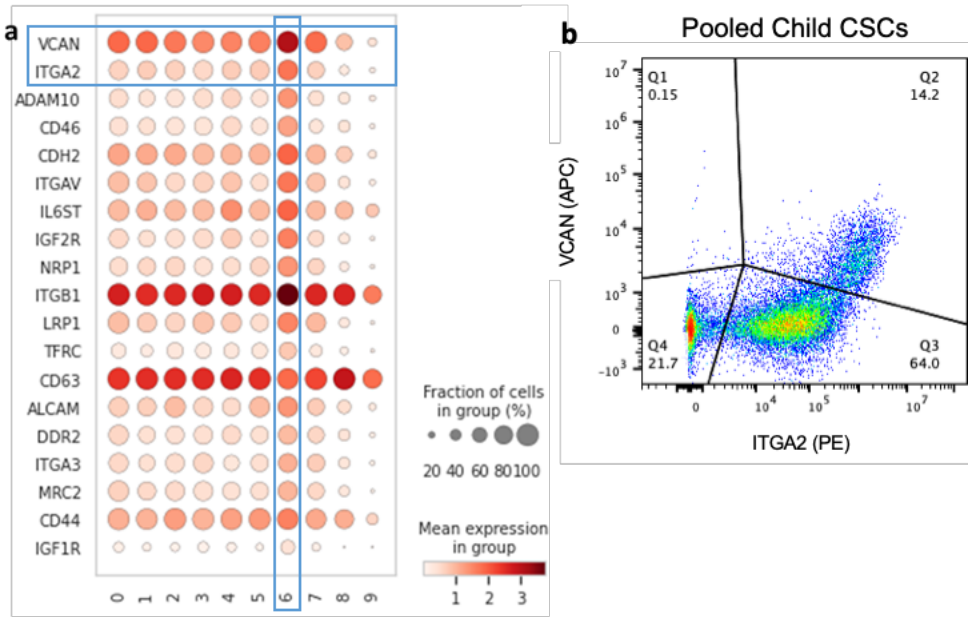


Figure 2-5. Cluster six surface protein confirmation.

(**a**) Transcriptional expression of conserved differentially expressed surface proteins in cluster 6 cells. Proteins were identified from the surface proteome dataset. (**b**) Flow cytometry analysis identified the population of interest in pooled child CPCs: ITGA2+, VCAN+. CPC: c-kit+ cardiac progenitor cell; ITGA2: Integrin alpha-2; VCAN: versican.

2.5 Discussion

There are clear age-dependent therapeutic differences in neonate and child CPC populations demonstrated by our prior studies.(Agarwal et al., 2016; Shoja-Taheri et al., 2019) In comparison to cCPCs, nCPCs demonstate greater anti-fibrotic potential, cell proliferation and chemotaxis, and enhanced secretion of cardioprotective paracrine factors.(Agarwal et al., 2016) Importantly, previous studies have isolated CPCs from patient cardiac biopsies using c-kit+ selection and explored age-dependent differences between nCPCs and cCPCs using bulk RNA sequencing. This approach, however, masks the identity of potential cell subpopulations and attributes sample variance to patient variables. Here, we aimed to understand how these macroscopic dynamics present at the single cell level, and whether we would be able to discern CPC subpopulations for selection with cell surface markers. To do so, we computed initial cell clusters—our CPC subpopulations. Then, to determine the major differences between cell clusters we examined both differentially expressed genes and enrichment of co-expressed gene modules. By combining multiple single-cell analysis methods, we uncovered potential phenotypes of CPC subpopulations which may explain CPC variability.

Given previous research, we expected to first find major differences in nCPCs and cCPCs. Indeed, nCPCs largely clustered among clusters 0-5, whereas a considerable portion of cCPCs clustered in the offshoot branches of the UMAP projection, namely clusters 3, 4, 6, 8, and 9 (**Figure** 2-2 **a-c**). Furthermore, given the demonstrated reduced performance of cCPCs, we hypothesized that cCPC samples are more heterogenous and may represent cells transitioning to a less reparative state. Here, we identified a high level of sample-to-sample variability among child patients, with some cells having more neonate-like clustering profiles than others (**Figure** 2-2 **c and d**). Most obvious, the two cCPC samples with neonate-like clustering profiles also

corresponded to the youngest of the child patient cohort (Patients 896 and 926, 12 months and 14 months old). Based on this observation, we ran a quasi-poisson regression model to assess gene expression variability dependent on patient age. Ultimately, the results of the model mirrored the results from the clustering-based analysis (**Supplemental Figure 2-1**; **Supplemental Table 2-2**). Genes such as *ABI3BP* and *CXCL6* that were upregulated in cCPC-enriched clusters also expressed highly in older patients, while genes like *CXCL12* and *CXCR4* that are upregulated in nCPC-enriched clusters had higher expression in younger patients.

Interestingly, we found evidence for enrichment of pro-inflammatory cell subpopulations and gene modules in cCPC samples, as compared to nCPC samples. First, cluster 4 cells, showed high expression of several inflammation- and immune-related cytokines including *IL1β*, *CXCL8, CCL2, CXCL6, IL33, CXCL1-3,* and *IL6* (**Figure** 2-4 **a**). While some CPCs from neonate patient 2016 were found in this cluster, cluster 4 was overall enriched in cCPCs (**Figure** 2-2 **e**). Second, we identified age-related differences among the composition of cytokine gene modules 8 and 9, determined with trajectory analysis. nCPC-enriched module 9 cytokines were more strongly associated with chemotaxis, whereas cCPC-enriched module 8 included inflammatory-related cytokines (**Figure** 2-3 **c-e**). Furthermore, we found a high positive correlation between cluster 4 and gene module 8. These results contradict a recent study by Vagnozzi et al. which challenged the efficacy of CPCs, attributing reparative function to an acute inflammatory-based wound healing response after cell delivery.(Vagnozzi et al., 2020) Of note, this work was completed in a mouse model of ischemia-reperfusion injury with murine CPCs, and CPC efficacy was evaluated 2 weeks after injection. Nevertheless, other studies investigating human CPCs from neonate patients corroborate our single-cell results. The results reported here are consistent with our previous research indicating cCPCs drive an increased immune response and nCPCs induce

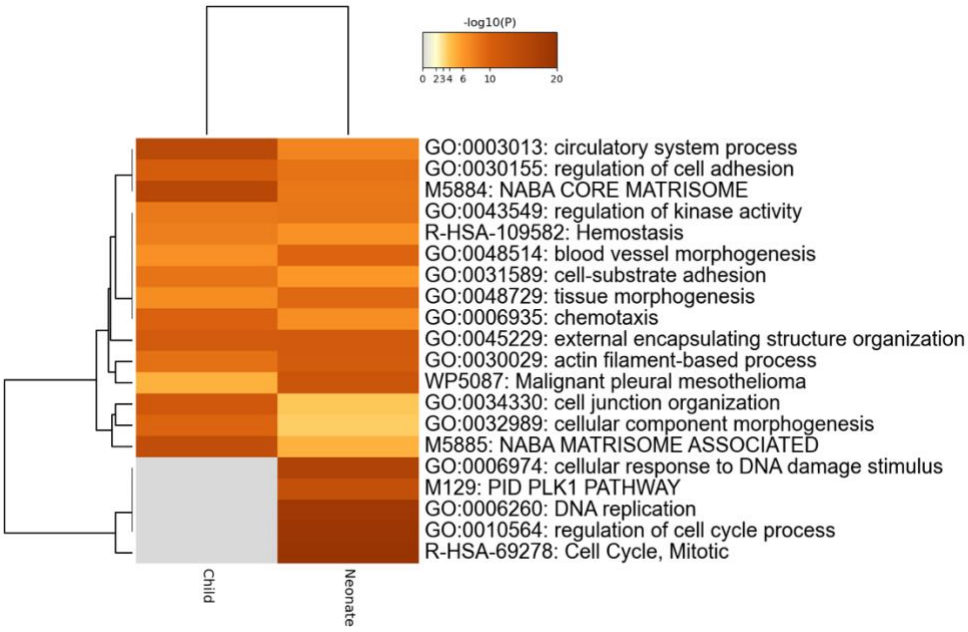
higher levels of mesenchymal stem cell chemotaxis.(Agarwal et al., 2016) Pathway analysis of gene array data comparing nCPCs and cCPCs demonstrated enrichment of anti-inflammatory response in nCPCs.(Agarwal et al., 2016) Additionally, in a rat model of myocardial infarction, nCPCs reduced macrophage infiltration in the myocardium post-injury, compared to adult CPCs.(Sharma et al., 2017)

Another cCPC-enriched subpopulation, cluster 6 showed high expression of genes related to fibrosis and angiogenesis, including *ITGB1, FBN1, DST, FN1, FST, COL3A1*/*4A1*/*8A1,* and *ADAMTS1* (**Figure** 2-4 **b and c**). Known fibroblast markers, *FAP* and *PDGFRA*, as well β-catenin were also upregulated in this cell cluster. Well-studied long non-coding RNAs *NEAT1*, *MEG3*, and *MALAT1* are among the most upregulated RNAs in cell cluster 6 and have been shown to contribute to myocardial injury and adverse remodeling. Multiple groups have determined MALAT1 and NEAT1 play a role in cardiac fibrosis.(Ge et al., 2022; Huang et al., 2019) Specifically, MALAT1 expression is upregulated in mouse myocardial infarction models and siRNA knockdown of MALAT1 attenuates infarction-induced fibrosis.(Huang et al., 2019) Further, MEG3 has been shown to promote myocardial damage in a rat model of cardiac ischemia-reperfusion and overexpression of MEG3 enhances myocyte apoptosis and decreases cell proliferation.(Zou et al., 2019) Interestingly, this cluster also has very high levels of fibronectin expression, which has been implicated as a critical protein during cardiac repair.(Konstandin et al., 2013) Given the strong connection of cluster 6 to adverse remodeling processes and upregulation of fibrotic genes, we sought to identify markers for this cell subpopulation. We identified *ITGA2* and *VCAN*, a proteoglycan extracellular matrix regulator, as candidate markers for cluster 6 cells. We used flow cytometry to confirm surface protein

expression on pooled child CSCs and demonstrated that the population of interest is discernable with from ITGA2 and VCAN surface protein expression (**Figure** 2-5 **b**).

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. Future work would allow for genetic profiles to be linked to protein expression of interest. Future CITE-seq experiments would allow for more rapid detection of subpopulation surface markers which when combined with cell sorting, would aid in compositional optimization of CSC samples. The combination of the RNA sequencing data presented in this study along with protein expression and in vitro or in vivo testing could be used to build predictive models 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 CSC-based therapies.

2.6 Supplemental Information



Supplemental Figure 2-1. Pathway analysis for genes from patient age regression.

Quasi-poisson regression was used to assess gene expression variability dependent on patient age. The results recapitulate our categorical age analyses: neonate CPCs are related to cell cycle and blood-vessel development, whereas child CPCs are closely related to cell adhesion processes.

Supplemental Table 2-1. Patient characteristics for single cell analysis.

Sample	*Patient*	*Age Group*	*Passage*	*Age*	*Heart Disease*	*Sex*
1	903	Neonate	11	< 1 Week	Hypoplastic left heart syndrome	F
2	925	Neonate	5	< 1 Week	Total anomalous pulmonary venous return	F
3	930	Neonate	5	< 1 Week	Total anomalous pulmonary venous return	M
4	926	Child	6	14 Months	Ventricular septal defect	M
5	1048	Child	12	6 Years	Atrial septal defect	F
6	896	Child	9	12 Months	Ventricular septal defect	F
7	938	Child	6	5 Years	Subaortic Membrane Resection	M
8	1092	Child	15	5 Years	Atrial septal defect	F
9	985	Neonate	7	2 Weeks	Interrupted aortic arch	M
10	2016	Neonate	12	1-2 Weeks	Atrial septal defect	F

Supplemental Table 2-2. Regression analysis of gene expression and patient age.

Top 20 genes from each age group using the quasi-poisson regression model implemented in Monocle to assess variability of gene expression based on patient age. Negative estimate values correspond to genes whose expression values that decrease with age, while positive values indicate genes with expression values that increase with age. The model formula also accounted for variability caused by batch effects. Values were sorted first on q-value, then by estimate.

Gene	Estimate	Age Group
LIPG	-0.76	Neonate
FIBIN	-0.73	Neonate
PCOLCE	-0.52	Neonate
TNNT1	-0.52	Neonate
AEBP1	-0.52	Neonate
CXCL12	-0.50	Neonate
KCNG1	-0.48	Neonate
BGN	-0.48	Neonate
PHGDH	-0.46	Neonate
COL6A3	-0.42	Neonate
HRH1	-0.42	Neonate
EDN1	-0.41	Neonate
PSAT1	-0.41	Neonate
PPP1R13L	-0.41	Neonate
NCK2	-0.40	Neonate
SLC1A5	-0.40	Neonate
INHBA	-0.38	Neonate
ANGPTL4	-0.37	Neonate
SOX17	-0.37	Neonate
U2AF1	-0.35	Neonate
CXCL6	0.70	Child
PCOLCE2	0.70	Child
RDH10	0.71	Child
BMPER	0.71	Child
ADIRF	0.72	Child
CACNA2D1	0.73	Child
TSLP	0.77	Child
MGARP	0.77	Child
F3	0.78	Child
APOA1	0.79	Child
EPDR1	0.84	Child
CPA4	0.87	Child
SLIT3	0.89	Child
ABI3BP	0.90	Child
RARRES1	0.98	Child
MEST	1.00	Child
NTN4	1.05	Child
PTGIS	1.07	Child
RGS5	1.09	Child
LOXL4	1.10	Child

Chapter 3: Comparative computational RNA analysis of c-kit+ progenitor cells and their extracellular vesicles

Jessica R. Hoffman[1,2], Hyun-Ji Park[1], Sruti Bheri[1], Arun R. Jayaraman[1], Michael E. Davis[1,2,3]

[1]Wallace H. Coulter Department of Biomedical Engineering, Emory University School of Medicine and Georgia Institute of Technology, Atlanta, GA, USA.

[2]Molecular and Systems Pharmacology Graduate Training Program, Graduate Division of Biological and Biomedical Sciences, Laney Graduate School, Emory University, Atlanta, GA, USA.

[3]Children's Heart Research and Outcomes (HeRO) Center, Children's Healthcare of Atlanta and Emory University, Atlanta, GA, USA.

The work contained in this chapter was published in Genomics in March 2022.

Citation: **Hoffman JR**, Park HJ, Bheri S, Jayaraman AR, Davis ME. Comparative computational RNA analysis of cardiac-derived progenitor cells and their extracellular vesicles. Genomics. 2022 Mar 26;114(3):110349. doi: 10.1016/j.ygeno.2022.110349. Epub ahead of print. PMID: 35346780.

3.1 Abstract

Stem/progenitor cells, including cardiac-derived c-kit+ progenitor cells (CPCs), are under clinical evaluation for treatment of cardiac disease. Therapeutic efficacy of cardiac cell therapy can be attributed to paracrine signaling and the release of extracellular vesicles (EVs) carrying diverse cargo molecules. Despite some successes and demonstrated safety, large variation in cell populations and preclinical/clinical outcomes remains a problem. Here, we investigated this variability by sequencing coding and non-coding RNAs of CPCs and CPC-EVs from 30 congenital heart disease patients and used unsupervised learning methods to determine potential mechanistic insights. CPCs retained RNAs related to extracellular matrix organization and exported RNAs related to various signaling pathways to CPC-EVs. CPC-EVs are enriched in miRNA clusters related to cell proliferation and angiogenesis. With network analyses, we identified differences in non-coding RNAs which give insight into age-dependent functionality

of CPCs. By taking a quantitative computational approach, we aimed to uncover sources of CPC cell therapy variability.

3.2 Introduction

Heart disease remains the leading cause of morbidity and mortality in the US, and often results in irreversible damage to the myocardium (Kochanek et al., 2020). Initially treated by surgery, drugs, or transplant, cardiac cell therapy emerged in the early 2000s with the goal to regenerate healthy myocardium after injury or disease. Over the years, several stem or progenitor cell types have been investigated for the treatment of various ischemic and congenital heart diseases (Bittle et al., 2018; Marban, 2018). In particular, previous preclinical research has shown that cardiac-derived c-kit+ cells (CPCs) repair the myocardium after injury (Agarwal et al., 2016; Saha et al., 2019; Trac et al., 2019b). Of note, our group is involved in a current phase I clinical trial investigating CPCs for treatment of hypoplastic left heart syndrome, a complex single ventricle congenital heart disease (NCT03406884).

Despite demonstrated safety and some efficacy in cardiac cell therapy preclinical and clinical trials, large variation in cell populations and patient outcomes remains a significant problem for further developing larger-scale, reliable therapies (Marban, 2018; Nguyen et al., 2016). Some of the variance in cell populations can be attributed to donor age. Additionally, once isolated, in vitro cell culture conditions or manipulations, like hypoxia and cell aggregation, also affect cell reparative effects (Gray et al., 2015; Trac et al., 2019b). Considering the high variability of cell populations, it is thus important to identify specific mechanisms of action in order to enhance cell therapy efficacy for patients. To this extent, our group has demonstrated that CPCs can repair rat right ventricle failure in an age-dependent manner, with neonatal CPCs having the greatest reparative capacity (Agarwal et al., 2017; Agarwal et al., 2016; Shoja-Taheri et al.,

2019). However, there is a dearth of quantitative studies investigating the underlying RNA cues driving CPC therapeutic efficacy. We have previously investigated the molecular basis for the differences between reparative and non-reparative CPCs with RNAseq experiments (Agarwal et al., 2016; Shoja-Taheri et al., 2019), however these studies were not comprehensive, with low sample sizes and consideration of only one or two types of RNA.

Originally, transplanted cells for cardiac disease treatment were thought to function via engraftment, proliferation, and differentiation. However, transplanted cellular retention is low and much of the therapeutic benefit is now attributed to paracrine signaling, including the release of extracellular vesicles (EVs) (Marban, 2018). EVs are lipid-bilayer vesicles released from cells via exocytosis or budding of the plasma membrane into the extracellular space. Once released, neighboring recipient cells may internalize EVs via endocytic processes, including direct membrane fusion, lipid-raft based uptake, and receptor-ligand interactions (Bheri et al., 2020). Importantly, EVs carry and protect diverse molecules, including RNAs, proteins, and lipids. The crucial role of EVs in cell therapy has been highlighted in previous work demonstrating that the inhibition of EV release diminishes the reparative effect of stem and progenitor cells (Ibrahim et al., 2014). In the context of cardiac cell therapy, this suggests CPCs themselves are not the only source of variable RNA signals contributing to repair. Studies have shown that the uptake of EVs by resident cardiac cells allows for the transfer of stem or progenitor cell EV cargo and stimulates repair in the injured tissue (Marban, 2018). Despite the well-established link between EV release/uptake and repair, our understanding of the signals or cargo molecules contributing to these effects is poor, especially given that EV cargo is highly heterogeneous (Willms et al., 2016). Importantly, CPCs and CPC-EV RNA content have been studied separately, but never together, comparatively.

Given the current limitations surrounding cardiac cell therapy variability, we aimed to investigate these discrepancies by performing bulk sequencing of primary CPCs from pediatric heart patients and their released EVs. First, we performed conventional differential expression analyses to identify key differences between cells and EVs and used unsupervised learning methods to reduce dataset dimensionality and summarize the data. Then, we systematically investigated the biological significance of our RNA-seq experiments with pathway and competitive endogenous RNA network analyses. We utilized publicly available datasets to provide greater context and determine the specificity of our CPC-EV results, in comparison to other cell types and EVs. Our results highlight the need for more quantitative investigations of cardiac cell therapy, and more personalized medicine approaches to cell therapy.

3.3 Methods

3.3.1 Isolation and Culture of c-kit+ Progenitor Cells (CPCs)

This study was approved by the Institutional Review Board at Children's Healthcare of Atlanta and Emory University. CPCs were isolated by c-kit magnetic bead sorting from cardiac biopsies of congenital heart disease patients (Supplementary Table 1) as previously described(Agarwal et al., 2016; French and Davis, 2014). CPCs were cultured using Hams F-12 medium with 10% fetal bovine serum, 1% Penicillin-Streptonmycin, 1% L-glutamine, and 0.04% fibroblast growth factor 2.

3.3.2 Extracellular Vesicle (EV) Collection and Characterization

EVs were successfully isolated from 27 of the 30 CPC populations. Briefly, CPCs were grown to 90% confluency, washed with PBS, and quiesced with serum free medium for 24 hours. Conditioned media was collected and subjected to sequential centrifugation: 3000 g for 10 min to

remove cells, 28,000 g for 30 min to remove cell debris, and 118,000 g for 1 hr 54 min to pellet EVs (Optima XPN-100 ultracentrifuge; Beckman Coulter SW 41 Ti rotor). EV size and concentration was determined by NanoSight NS300. Samples were diluted 1:10 in PBS, and three, 60-second video images were captured per sample and analyzed by NanoSight NTA 3.4 software.

3.3.3 Next Generation Sequencing

RNA from CPCs and CPC-EVs were isolated with the miRNeasy Mini Kit (Qiagen), according to manufacturer's instructions. Purified RNA was analyzed (2100 Bioanalyzer and TapeStation Controller, Agilent Genomics) for miRNA and RNA size, quality, and quantity. RNA library preparation and sequencing was conducted by Novogene Co., Ltd (Illumina NovaSeq 6000 with PE150 platform) or the Emory Yerkes Nonhuman Primate Genomics core (Illumina HiSeq 3000). Sequencing source information for each CPC patient is displayed in Supplementary Table 1.

Small RNA sequencing performed at Novogene were trimmed and filtered in the FASTQ Toolkit Illumina Basespace app. Truseq adapters (AGATCGGAAGAGC) were trimmed with an adapter trim stringency set to 0.90. Reads were filtered to 18-51 length with reads passing the FASTQ Toolkit filter. Then, reads were mapped with Bowtie aligner and hg19 and miRBase v21 references, and mature miRNA hits were determined using the small RNA Illumina Basespace app. Small RNA sequencing performed at Emory Yerkes Nonhuman Primate Genomics core were aligned and hits were determined using the Qiagen GeneGlobe console with QIAseq miRNA Quantification tool. Default parameters were used: 3' adapters were trimmed using cutadapt, reads with less than 16 base pair insert sequences or less than 10 base pair unique

molecular index sequences were removed, reads were aligned with Bowtie aligner, GRCh38 and miRBase v21 references.

Total RNAseq files from Novogene and Emory Yerkes Nonhuman Primate Genomics core were aligned, and gene counts were determined with the STAR aligner in the Illumina BaseSpace app, RNA-Seq Alignment. Reads were aligned with hg19 reference genome. Biotypes were matched to alignment results using the Ensembl based annotation package (EnsDb.Hsapiens.v79). miRNAs were considered only from small RNA sequencing and were thus removed from the total RNA sequencing set. All of CPC lncRNAseq was performed with Novogene. lncRNAs were aligned using STAR with the quantMode GeneCounts option and GRCh38 reference (Dobin et al., 2013). Total RNA sequencing counts after filtering are presented in Supplementary Tables 11 and 12; miRNA sequencing counts after filtering are presented in Supplementary Tables 13 and 14. RNA sequencing and alignment metrics are presented in Supplementary Tables 15 and 16.

3.3.4 RNA Sequencing Data Analysis

Data analysis was completed in R. First, raw aligned RNA counts for CPC and CPC-EVs were filtered: we removed RNA with zero count entries in ten or more samples and used edgeR package's 'filterByExpr' function using the default parameter settings. RNA counts were normalized using the edgeR weighted trimmed mean of M-values method (default parameters), and transformed into $\log_2$ counts per million (logCPM). Principal component analyses (PCA) were performed using the 'prcomp' built in function. Heatmaps were generated using the pheatmap package with Manhattan distance calculations and ward.D2 clustering method.

Batch correction of sequencing data (in logCPM) was implemented using the sva package's 'ComBat' function to correct for sequencing performed at separate sites: Novogene and Emory

Yerkes Nonhuman Primate Genomics Core (Supplementary Table 1)(Leek et al., 2012). PCA plots before and after batch correction are displayed in **Supplemental Figure 3-3 a and b**. Variance explained by each covariate (patient, sex, age (in $\log_{10}$months), type (cell or EV), and sequencing site) was assessed before and after batch correction with variancePartition package (**Supplemental Figure 3-3 c and d**) (Hoffman and Schadt, 2016). Age was treated as a fixed effect and all other categorical covariates were treated as random effects.

For differential expression analysis of CPC and CPC-EV RNA content, we used the dream (differential expression for repeated measures) linear mixed modeling approach from the variancePartition package to account for patient matched cell and EV data (Hoffman and Roussos, 2021). Patient was treated as a random effect and the following co-variates were considered: patient, sex, age (in $\log_{10}$months), type (cell or EV), batch (sequencing site). Weights were estimated with the voomWithDreamWeights function and the cell vs. EV hypothesis test was conducted with the dream function (Satterthwaite approximation method). Cell vs. EV differential expression analyses were conducted for n=26 patients: Patient 894 was removed because sex was not known, and cell sequencing from patients 938, 1048, and 1092, were not included as EVs were not sequenced from these patients.

Volcano plots were constructed from linear mixed model results using the EnhancedVolcano package. Correspondence between PCA variable loadings and differential expression analysis results, presented in volcano plots, is displayed in **Supplemental Figure 3-4**. miRTarBase was used to identify miRNAs with known targets (validated by at least three assays) (Hsu et al., 2011). Biological pathway enrichment analyses were performed using Metascape (Zhou et al., 2019).

3.3.5 Data Mining

We probed previously published datasets containing miRNA data from parent cells and their EVs (Kanlikilicer et al., 2016; Melo et al., 2014; Mittelbrunn et al., 2011; Ono et al., 2014; Tadokoro et al., 2013; Umezu et al., 2014; Vignard et al., 2020; Yokoi et al., 2018). GEO2R was used to determine the differential expression of EV vs. cell miRNAs (Benjamini Hochberg p-value adjustment)(Barrett et al., 2013). miRNAs in each dataset were ranked from largest to smallest $\log_2$fold-change (EV/Cell). To account for different sized datasets, ranks were scaled within their respective study using the 'smoothPalette' function in the tagcloud package, before plotting with the scales package in R. Data mining results are presented in Supplementary Table 17.

3.3.6 ceRNA Network Construction

Differentially expressed RNAs (neonate vs. child CPCs) were determined using edgeR and limma/voom method. Reads from total RNA, lncRNA, and miRNA CPC sequencing were filtered: we removed RNA with zero count entries in ten or more samples and used edgeR package's 'filterByExpr' function with the default parameter settings. Total RNA, miRNA, and lncRNA models were built using all CPC data (neonate, infant, and child) with the following covariates: age group, sex, and batch (sequencing site). Counts were transformed to logCPM values and weights for linear modeling were computed using the limma voom function. Linear models were fit. Then, contrast matrices for age groups were created (neonate – child, neonate – infant, infant – child), coefficients were estimated, and moderated statistics were computed with empirical Bayes moderation. Differentially expressed RNAs between neonate and child groups were identified using the topTable function. Differential expression results for lncRNA sequencing are presented in Supplementary Table 18. The differentially expressed miRNAs (18)

were matched to differentially expressed lncRNAs (134) by miRcode and differentially expressed mRNAs (505) by miRTarBase (|fold-change|>2 and $p<0.05$). The resulting RNAs that were matched within the set were displayed in a network with Cytoscape (Shannon et al., 2003). The full network (107 nodes, 144 edges) was reduced to the most highly connected hubs (36 nodes, 75 edges) using the MCODE plug in.

Furthermore, we investigated the age-related differences in CPC RNAs in a quantitative approach by transforming patient age to $\log_{10}$months. We constructed linear models with age (continuous), sex, and batch (if applicable) covariates, as before. We compared the differentially expressed RNAs ($p<0.05$) determined by this quantitative method with the differentially expressed RNAs determined by the categorical method above (age as neonate, infant, or child). The results are displayed in **Supplemental Figure. 3-5**.

3.4 Results

3.4.1 Characterization of EVs from neonate, infant, and child CPCs

To characterize CPC-EVs, CPC populations were grown and expanded in 2D culture and EVs were isolated from the conditioned media via differential ultracentrifugation (**Figure** 3-1 **a**). CPCs were previously isolated from cardiac biopsies of neonate (<1 week, n=9), infant (1 week – 1 year, n=13), and child (>1 year, n=8) congenital heart disease patients (Supplementary Table 1). Bioanalyzer profiles of CPC and CPC-EV RNA revealed distinct 18S and 28S ribosomal peaks in the CPC RNA, but not in the CPC-EV RNA, confirming successful isolation of EVs without cellular contamination (**Figure** 3-1 **b**). Furthermore, total RNA from EVs was enriched in small RNAs with a peak ~22nt, the size of miRNAs. CPC-EVs were imaged using transmission electron microscopy and analyzed for size and concentration with nanoparticle

tracking analysis. Independent of patient cell source, all CPC-EVs were 100-140 nm, characteristic of exosomes or small EVs (**Figure** 3-1 **c and d**).

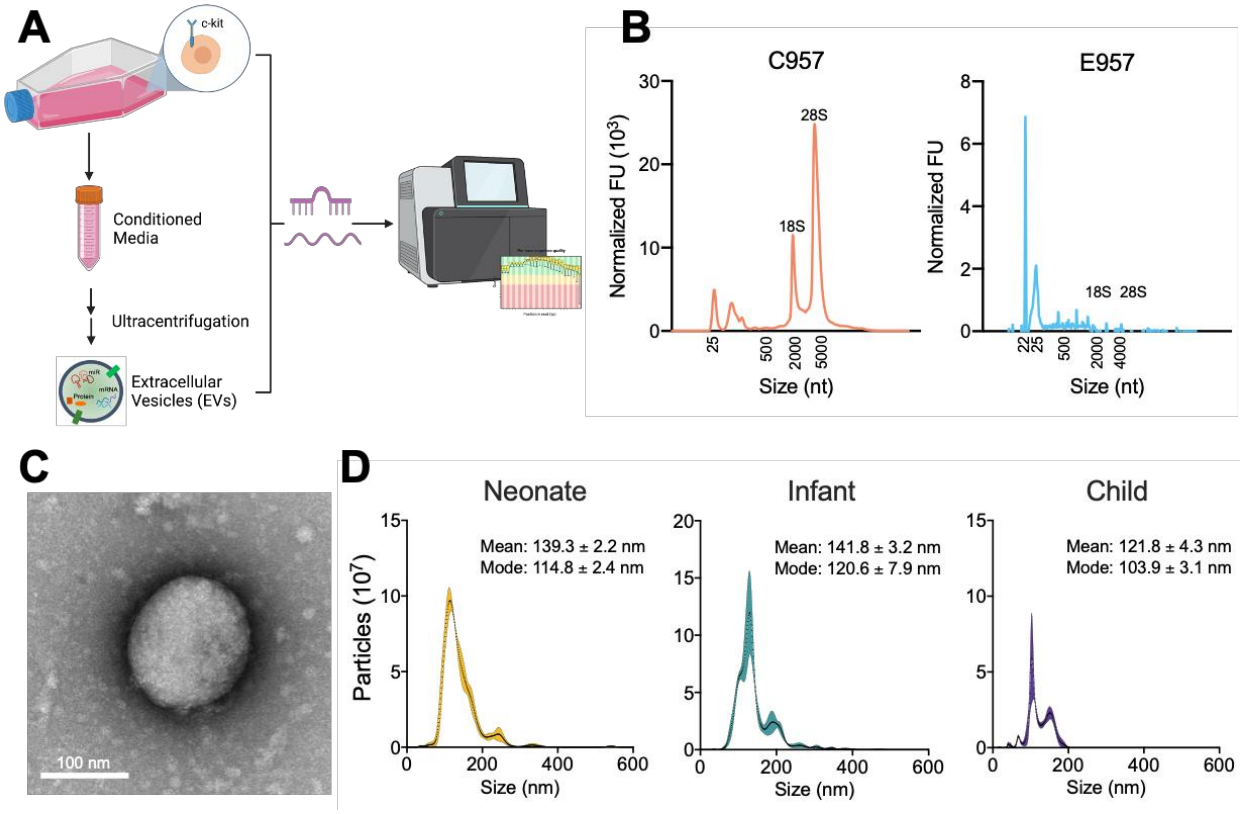


Figure 3-1. CPC-EV isolation and characterization.

a Schematic of study: CPCs and CPC EV miRNA and total RNA are sequenced and analyzed (created from Biorender.com).
b Bioanalyzer profile for patient matched CPC and CPC-EV RNA content.
c Transmission electron microscopy image of CPC EV. Scale bar, 100 nm.
d Vesicle size distribution histogram by nanoparticle tracking analysis in neonate, infant, and child CPC-derived EVs. Shaded region indicates standard error.

3.4.2 CPCs retain ECM-related RNAs and export signaling pathway-related RNAs to EVs

Total RNA sequencing results identified 13,718 and 8,718 expressed RNAs in CPCs and CPC-EVs, after filtering out lowly expressed RNAs (**Figure** 3-2 **a**) (Robinson et al., 2010). Dimension reduction with principal component analysis (PCA) of the 8,563 commonly expressed RNAs revealed distinctive separation of CPC and CPC-EV samples across the first two components, cumulatively representing 38% of the total variance (**Figure** 3-2 **b**). Furthermore, an unbiased

heatmap of the top 1% variable RNAs revealed hierarchical clustering by source: CPC vs. CPC-EV (**Figure** 3-2 **c**).

Given that CPC and CPC-EV samples were matched, derived from each patient, we used the dream (differential expression for repeated measures) approach to determine differentially expressed RNAs between CPCs and CPC-EVs. In total, 4,898 RNAs, or 57% of the commonly expressed RNAs, were differentially expressed with adjusted p-values < 0.05 (**Figure** 3-2 **d**, Supplementary Table 2)(Hoffman and Roussos, 2021). The top differentially expressed RNAs in CPCs and CPC-EVs are displayed in

Table 3-1**.** Notably, collagen type IV and VIII chains, integrin alpha V, dystroglycan 1, and growth arrest – specific 6 are upregulated in CPCs, and Ras-related protein Rab-13, dexamethasone-induced Ras-related protein 1, colony stimulating factor 1 receptor, and interleukins 33 and 16 are upregulated in CPC-EVs. To further determine the biological significance of these differentially expressed RNAs, we performed pathway analysis of the top 250 RNAs upregulated in CPCs and CPC-EVs (ranked by fold-change). Metascape pathway analysis revealed CPCs were enriched in RNAs associated with extracellular matrix (ECM) organization, ECM-receptor interaction, insulin-like growth factor transport and immune responses; whereas CPC-EVs were enriched in RNAs involved in peptide chain elongation, RNA splicing, and MAPK and G alpha (q) signaling (**Figure** 3-2 **e**, Supplementary Tables 3 and 4) (Zhou et al., 2019).

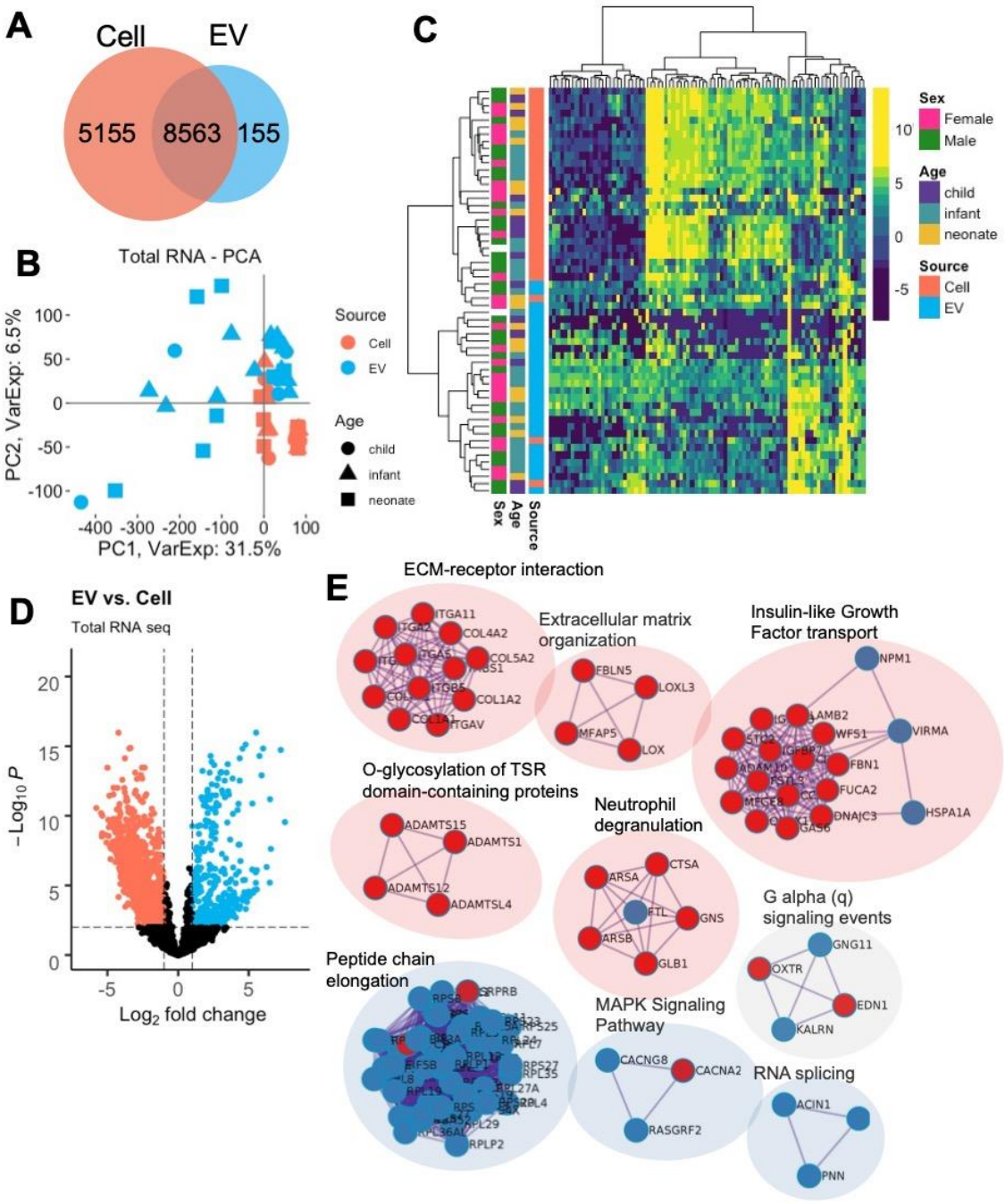


Figure 3-2. Differential expression of CPC and CPC-EV total RNAseq.

a RNAs expressed in CPCs and CPC EVs.
b PCA plot of CPC and CPC EVs show clustering by RNA source across PC1.
c Heatmap of top 1% variable RNAs show clustering of samples by source: cell and EV.
d Volcano plot of differentially expressed RNAs in CPCs and CPC EVs.
e Network of top enriched terms from differentially expressed RNAs upregulated in CPCs (red) vs. EVs (blue) using the Metascape tool.

Table 3-1. RNAs preferentially released to EV and retained in cell

RNA	Log2(EV/Cell)	padj	RNA	Log2(Cell/EV)	padj
CA11	7.56	**	CPA4	5.78	***
RAB13	7.26	***	OSMR	5.44	***
RASD1	6.56	**	COL8A1	5.40	**
PLEKHA4	6.51	***	OXTR	5.37	**
NEFM	6.49	**	FSTL3	5.36	***
C1orf115	6.26	**	GDF6	5.15	**
ANP32B	6.01	***	TNFRSF11B	5.07	**
NET1	5.90	***	COL4A1	5.00	***
C22orf46	5.76	***	RECK	4.94	**
CSF1R	5.57	**	CHPF	4.91	**
KIF21B	5.57	*	SLC16A3	4.89	**
TRAK2	5.52	***	P3H4	4.87	***
ERG	5.50	**	EDIL3	4.82	**
CDC42BPG	5.50	**	UGCG	4.81	**
IL33	5.42	*	TRAM2	4.79	**
IL16	5.31	*	ANGPTL4	4.78	**
DNAH10	5.31	*	MEGF6	4.75	**
CGNL1	5.26	***	GAS6	4.69	**
PITPNM3	5.25	**	IGFBP3	4.65	**
SPTBN4	5.06	*	B4GALT1	4.64	**
ROCK1P1	5.00	*	ARSJ	4.63	**
GUCY1A2	4.88	*	CD248	4.60	**
KIAA1211	4.87	*	PKD1	4.59	**
CCDC88C	4.79	*	ITGA5	4.58	**
FLG	4.79	*	DAG1	4.58	**

***p < 0.01, **p < 1e-5, ***p < 1e-10**

3.4.3 CPC-EVs are enriched in miRNAs involved in cardiac development and cell signaling

We then performed similar miRNAseq analyses: miRNAseq revealed 206 and 641 expressed miRNAs in CPCs and CPC-EVs, respectively, after filtering out lowly expressed miRNAs (**Figure** 3-3 **a**). PCA of the 193 commonly expressed miRNAs showed clear separation of CPC and CPC-EV samples across principal component 1, representing 27.9% of total variance (**Figure** 3-3 **b**). Linear regression using the dream approach was used to determine the differentially expressed miRNAs between CPCs and CPC-EVs. In total, 126 of 193, or 65% of miRNAs were differentially expressed with adjusted p-values <0.05 (**Figure** 3-3 **c**, **Supplemental Table 3-5**). The top differentially expressed miRNAs are displayed in

Table 3-2. Next, we found gene targets for the differentially expressed miRNAs using miRTarBase with the criteria that targets be experimentally validated by at least 3 methods (**Supplemental Table 3-6**). Pathway analysis of the gene targets revealed enrichment of gene ontology (GO) pathways involved in vasculature and heart development, VEGFA-VEGFR2, and TGF-beta signaling, as well as cell adhesion, differentiation, and apoptosis (**Figure** 3-3 **d, Supplemental Tables 3-7 and 3-8**).

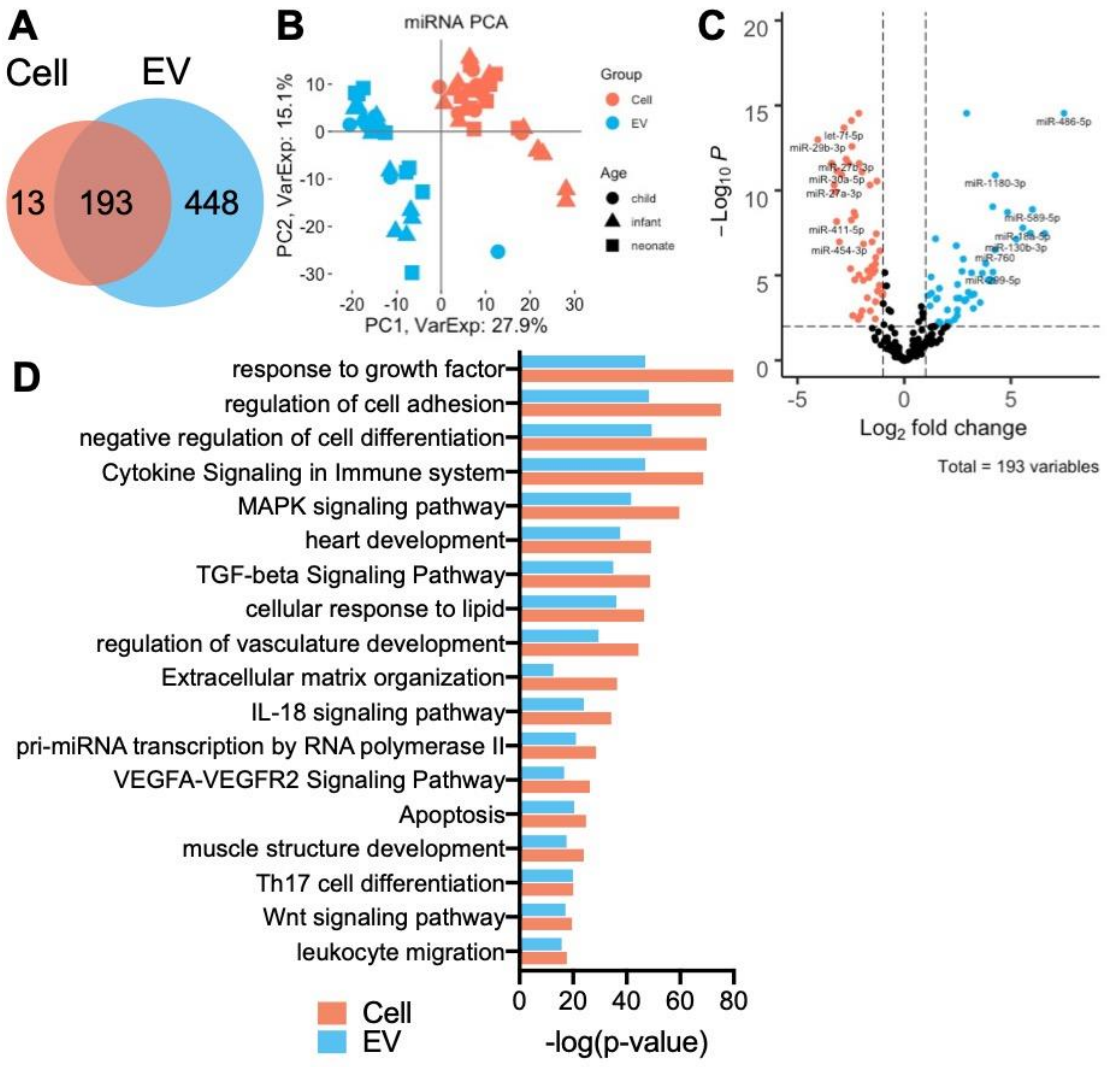


Figure 3-3. Differential expression of CPC and CPC EV miRNAseq.

a miRNAs expressed in CPCs and CPC EVs.
b PCA plot of CPC and CPC EVs show clustering by RNA source across PC1.
c Volcano plot of differentially expressed miRNAs in CPCs and CPC EVs.

d GO pathway enriched terms from differentially expressed miRNA gene targets, as determined by miRTarBase.

Table 3-2. miRNAs preferentially released to EV and retained in cell

miRNA	Log2(EV/Cell)	padj	RNA	Log2(Cell/EV)	padj
hsa-miR-486-5p	7.49	***	hsa-miR-29b-3p	4.05	***
hsa-miR-203a-3p	6.58	**	hsa-miR-379-5p	3.39	***
hsa-miR-589-5p	6.01	**	hsa-miR-27a-3p	3.28	***
hsa-miR-122-5p	5.92	**	hsa-miR-31-5p	3.24	**
hsa-miR-18a-5p	5.56	**	hsa-miR-411-5p	3.17	**
hsa-miR-130b-3p	5.25	**	hsa-miR-30a-5p	3.16	***
hsa-miR-320c	4.84	**	hsa-miR-454-3p	3.04	**
hsa-miR-1180-3p	4.27	***	hsa-let-7e-5p	2.93	***
hsa-miR-760	4.26	**	hsa-let-7f-5p	2.82	***
hsa-miR-299-5p	4.16	**	hsa-miR-27b-3p	2.73	***
hsa-miR-92b-5p	4.15	**	hsa-miR-29a-3p	2.61	***
hsa-miR-339-3p	4.15	*	hsa-miR-143-5p	2.52	**
hsa-miR-125a-3p	3.93	*	hsa-miR-30e-5p	2.48	**
hsa-miR-494-3p	3.82	**	hsa-miR-21-5p	2.48	***
hsa-miR-501-3p	3.66	**	hsa-miR-103a-3p	2.45	***
hsa-miR-500a-3p	3.56	*	hsa-miR-27a-5p	2.41	*
hsa-miR-184	3.26	*	hsa-let-7a-5p	2.35	**
hsa-miR-127-5p	3.23	*	hsa-miR-134-5p	2.32	*
hsa-miR-4661-5p	3.15	*	hsa-miR-143-3p	2.30	**
hsa-miR-197-3p	3.15	**	hsa-miR-431-5p	2.16	*
hsa-miR-214-3p	3.00	*	hsa-miR-24-3p	2.13	***
hsa-let-7d-3p	2.91	***	hsa-miR-493-3p	2.12	**
hsa-miR-1287-5p	2.90	*	hsa-miR-16-5p	2.12	***
hsa-miR-320b	2.82	*	hsa-miR-450b-5p	2.06	*
hsa-miR-92b-3p	2.76	**	hsa-miR-152-3p	1.98	***

***p < 0.01, **p < 1e-5, ***p < 1e-10**

3.4.4 CPC-EVs contain vesicle biosynthesis and cell cycle-related miRNAs

We investigated the enrichment of well-studied miRNA families and cardiac-related miRNAs, with known functions in CPC-EVs (**Figure** 3-4 **a**). Most notably, members of the miRNA 17/92 cluster (miR-18a-5p and miR-92a-3p) involved in development and cell proliferation are upregulated in CPC-EVs (Chen et al., 2013a; He et al., 2005; Lu et al., 2007). Additionally, members of the miRNA 99/100 family (miR-99a-5p and miR-99b-3p) involved in hematopoietic stem cell renewal are also upregulated in CPC-EVs (Emmrich et al., 2014).

To understand the specificity of our CPC vs. CPC-EV results, we searched the Gene Expression Omnibus (GEO) database for previously published datasets with miRNAseq from EVs and their parent cells. We found eleven datasets comprised of many cell types, including, various cancer cells, immune cells, and bone marrow mesenchymal stem cells (Kanlikilicer et al., 2016; Melo et al., 2014; Mittelbrunn et al., 2011; Ono et al., 2014; Tadokoro et al., 2013; Umezu et al., 2014; Vignard et al., 2020; Yokoi et al., 2018). We calculated fold-change values (EV/Cell) for each data set and ranked the miRNAs in order of decreasing fold-change value (**Figure** 3-4 **b**). We found our data set was congruent with other data sets: our CPC-EVs were enriched in some miRNAs upregulated in other EV types, and our CPCs retained some similar miRNAs to other cell types. Specifically, well-studied miR-122-5p, which has been implicated in miRNA EV cargo sorting, was top ranked in multiple data sets (Temoche-Diaz et al., 2019). Nevertheless, there were a handful of unique miRNAs, enriched in CPC-EVs, but not in EVs from other cell types. These miRNAs included miR-18a-5p and miR-130b-3p. Evidence suggests that miR-18a-5p is anti-apoptotic (Li et al., 2017; Mogilyansky and Rigoutsos, 2013), while miR-130b-3p is pro-apoptotic, targets insulin-like growth factor 1, and may be cardiac harmful (Gan et al., 2020; Li et al., 2016b).

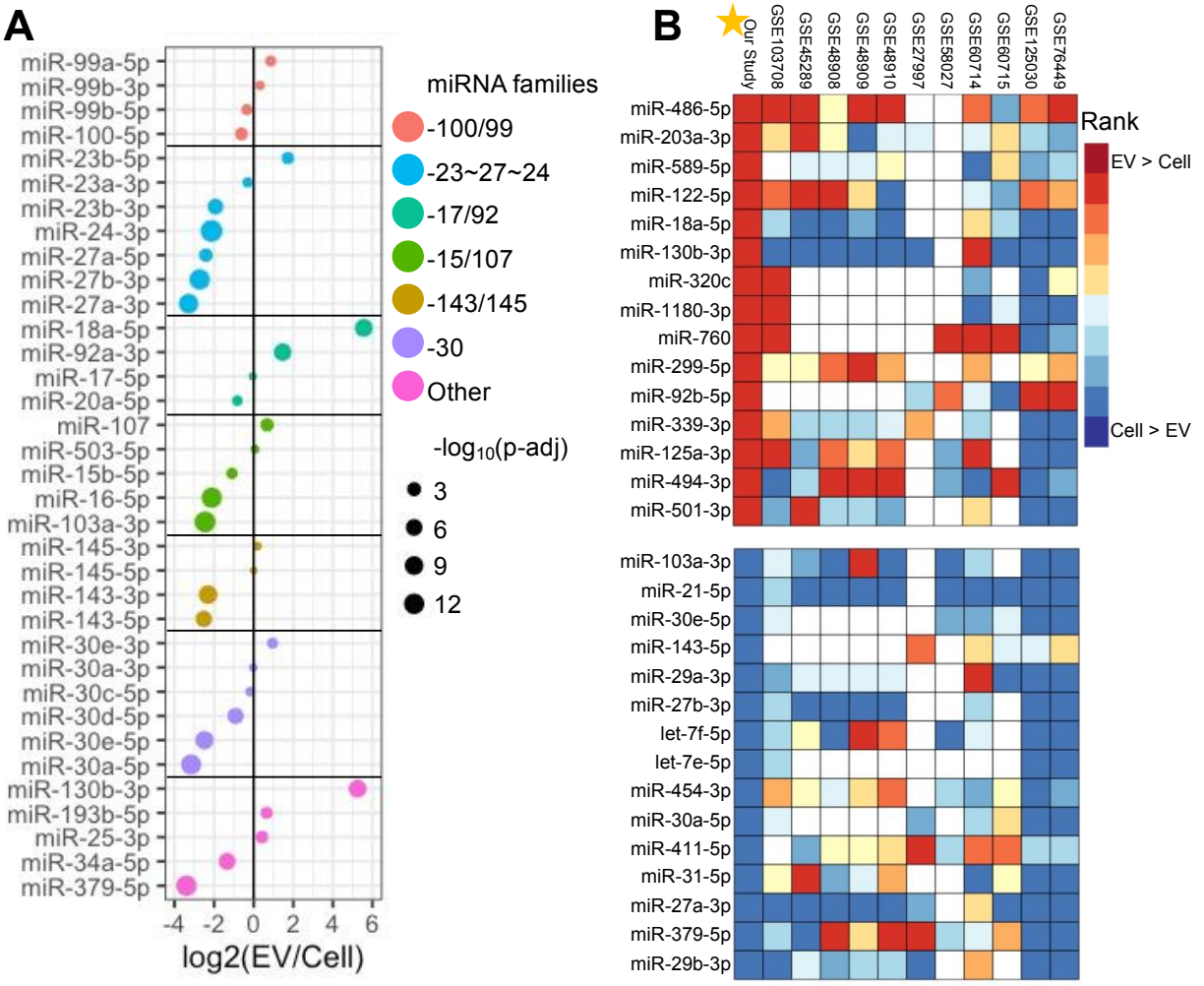


Figure 3-4. Differential expression of well-studied CPC and CPC-EV miRNAs and miRNA clusters.

a Differential expression of well-studied miRNAs in CPCs and CPC-EVs.
b Our study's top 15 miRNAs up (red) and down (blue) regulated in EV samples plotted with rank of top enriched miRNAs from 11 publicly available databases. GEO database numbers listed on top; white color indicates no available miRNA expression (NA).

3.4.5 Construction of ceRNA network

Evidence suggests that lncRNAs act as competing endogenous RNAs (ceRNAs) and play a key role in regulating RNA expression (Ma et al., 2020; Qi et al., 2015; Salmena et al., 2011). Furthermore, previous work has demonstrated that CPCs function in an age-dependent manner: CPCs from neonate patients are more reparative than CPCs from older patients (Agarwal et al., 2016; Sharma et al., 2017). To gain a comprehensive level of understanding of age-dependent

CPC ceRNA interactions, we investigated the differentially expressed mRNAs, lncRNAs, and miRNAs between neonate and child CPCs. Overall, child and neonate CPC RNA profiles were the most dissimilar (**Supplemental Fig. 3-1**). PCA of coding and non-coding RNAs showed separation of neonate and child CPCs across the first three components (**Figure** 3-5 **a**).

We used edgeR/limma to find the top differentially expressed miRNAs (18), lncRNAs (134), and mRNAs (505) (|fold-change|>2 and $p<0.05$) between neonate and child CPCs (**Supplemental Table 3-9**). Overall, child CPCs were enriched in various non-coding RNAs, as compared to neonate CPCs (**Figure** 3-5 **b and c**). We then matched differentially expressed miRNAs to differentially expressed lncRNAs and mRNAs by putative target sites using miRcode and miRTarBase as shown in **Figure** 3-5 **d** (Hsu et al., 2011; Jeggari et al., 2012). The resulting ceRNA network consisted of 107 nodes and 144 edges (**Figure** 3-5 **e**). The most highly connected nodes included miRNAs: miR-218-5p and -877-3p upregulated in child CPCs, and -23a-3p, -23b-3p, and -301a-3p, upregulated in neonate CPCs. (

Table 3-3**, Supplemental Fig. 3-2**). Metascape enrichment analysis revealed these nodes are enriched in pathways including, blood vessel development, positive regulation of cell cycle, and regulation of Wnt signaling pathway (Supplementary Table 10). We analyzed the hub genes in the full ceRNA (**Figure** 3-5 **e**) network using the Cytoscape MCODE plug-in. The full network was reduced to the most highly connected 36 hub nodes, including miRNAs: miR-23a-3p, -23b-3p, -148a-3p, -181a-5p, -218-5p, -301a-3p, and -877-3p (**Figure** 3-5 **f**).

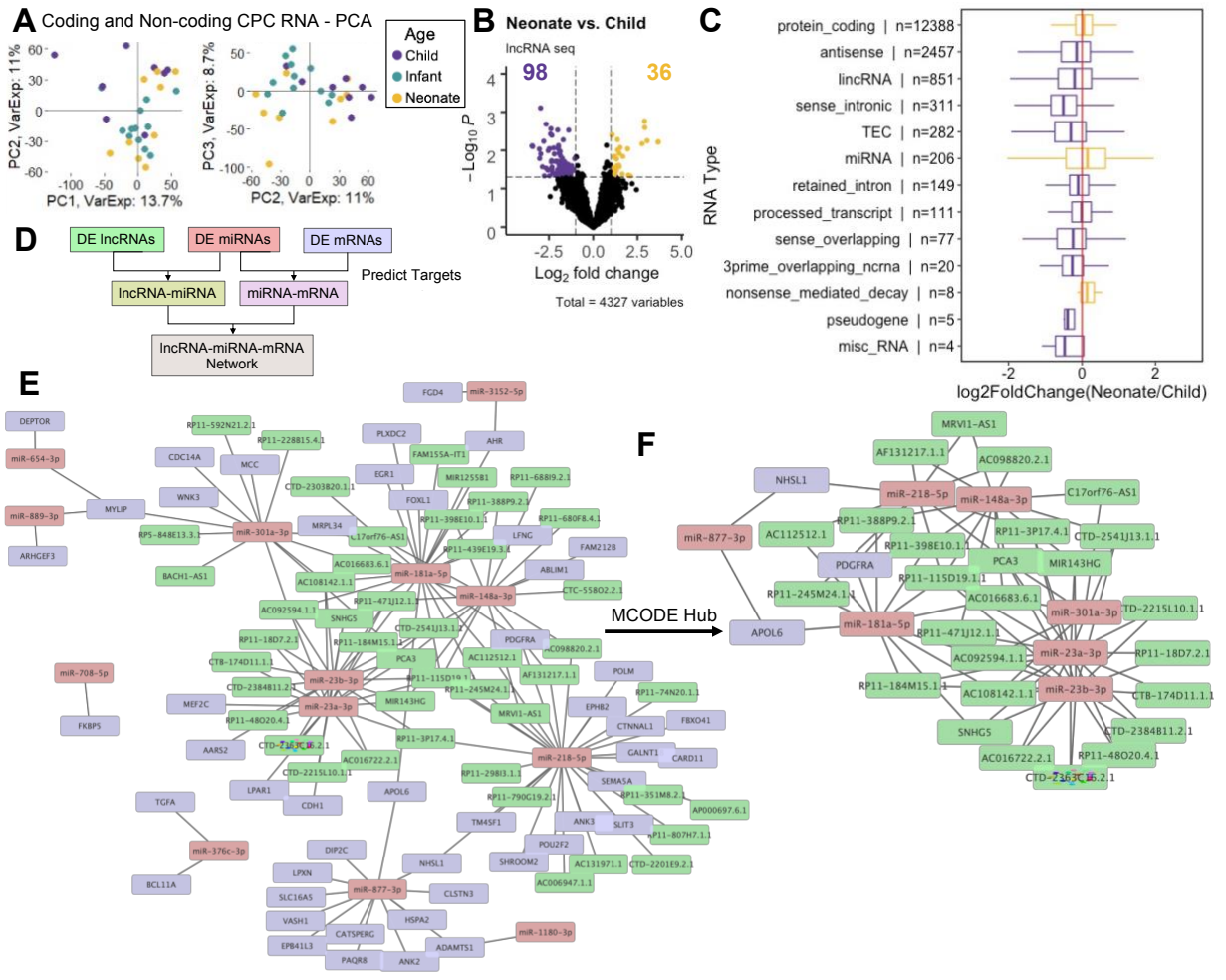


Figure 3-5. ceRNA network of neonate and child CPCs.

a PCA plot of CPC coding and non-coding CPC RNAs (PC1, PC2, PC3) show clustering by patient age group.
b Volcano plot of differentially expressed non-coding RNAs between neonate and child CPCs. Thirty-six and 98 RNAs are upregulated in neonate and child cells, respectively.
c Fold-change values for each RNA in various categories. RNAs upregulated in child CPCs (purple), compared to neonate CPCs (yellow). N represents number of measured RNAs in each category.

d Schematic for ceRNA network construction: differentially expressed RNAs between neonate and child CPCs were matched in miRcode and miRTarBase by putative target sites.
e Full ceRNA network of differentially expressed RNAs (neonate vs. child CPCs) with 107 nodes and 144 edges.
f Reduced MCODE network of the most highly connected RNA nodes.

Table 3-3. ceRNA network hub connectivity

RNA Type	Name	Node Degree
miRNA	miR-218-5p	31
miRNA	miR-181a-5p	22
miRNA	miR-23a-3p	20
miRNA	miR-23b-3p	18
miRNA	miR-148a-3p	17
miRNA	miR-877-3p	13
miRNA	miR-301a-3p	13
lncRNA	RP11-115D19.1.1	5
lncRNA	PCA3	5
lncRNA	AC092594.1.1	4
lncRNA	AC108142.1.1	4
mRNA	MYLIP	3
lncRNA	CTD-2541J13.1.1	3
lncRNA	MIR143HG	3
lncRNA	RP11-184M15.1.1	3
lncRNA	RP11-3P17.4.1	3
lncRNA	RP11-471J12.1.1	3
lncRNA	SNHG5	3
lncRNA	AC016683.6.1	3

3.5 Discussion

Several cell types are being clinically evaluated for use in cardiac cell therapy, including CPCs (NCT02501811, NCT03351400, NCT03406884). Despite promising preclinical results, there remain concerns over variability from different cell populations and cell therapy outcomes (Rosen et al., 2014). Specifically, several groups have shown that CPC therapeutic efficacy is dependent on patient age and disease, as well as cell culture and expansion conditions (Agarwal et al., 2016; Mishra et al., 2011; Trac et al., 2019b; Yan et al., 2012). Furthermore, research has shown that cardiac cell therapy efficacy can be attributed to paracrine signaling and the release

of EVs, rather than cellular engraftment, proliferation, and differentiation (Marban, 2018). Therefore, to investigate sources of CPC-therapy variability, we used next generation sequencing methods to determine differences among pediatric primary CPCs (n=30 patients) and their respective EVs released in vitro. We analyzed CPC and CPC-EV coding and non-coding RNA content and investigated the differences (1) between patient matched CPCs and CPC-EVs, and (2) across CPC patient age groups. Previous research has investigated differences between stem cells and their EVs, but these studies were limited in sample size and focused on miRNA only (Adamiak et al., 2018; Xin et al., 2020). By taking a comprehensive computational approach, we gained insight into potential mechanisms of action and sources of variability of CPC cell therapy.

First, we determined the differences between CPC and CPC-EV RNA content. For this study, we profiled coding and long non-coding RNA using total RNAseq and miRNA with small RNAseq. As expected, we found that the greatest contributor to variance across the entire dataset was source (cell vs. EV), rather than patient age or sex (**Figure** 3-2 **b and c**; **Figure** 3-3 **b**; **Supplemental Figure 3-3 c and d**). Using differential expression analysis, we discovered that a large portion of RNAs and miRNAs were expressed in only cells or EVs. Further, of the intersecting set of RNAs expressed in both cells and EVs, many were differentially expressed (adjusted p-values < 0.05, **Figure** 3-2 **a and d; Figure** 3-3 **a and c**). In total, CPCs retained RNAs involved in extracellular matrix organization and exported RNAs to EVs involved in various signaling pathways (**Figure** 3-2 **e; Figure** 3-3 **d**).

Importantly, cell type and cell environment affect EV cargo (Agarwal et al., 2017; Sahoo and Losordo, 2014; Trac et al., 2019a). In the context of cardiac cell therapy, stem and progenitor cell-EVs have been shown to carry RNAs with beneficial pleiotropic effects—immunomodulatory, anti-fibrotic, anti-apoptotic, pro-angiogenic, pro-migratory, and pro-

proliferative—as compared to their non-progenitor cardiac cell type counterparts (Barile et al., 2017b; Ibrahim et al., 2014; Marban, 2018). Additionally, manipulations affecting parent stem and progenitor cell environment, like hypoxia and cell aggregation, affect EV RNA cargo and resulting EV efficacy (Agarwal et al., 2017; Trac et al., 2019a). Special attention has been paid to EV miRNA cargo, as it plays an important role in cardiac repair: signaling between stem or progenitor cells and resident cardiac cells (Emanueli et al., 2015). Circulating EVs protect internal cargo from degradation and are thus a rich source of circulating miRNAs. Our results indicate that CPC-EVs are upregulated in several important and well-studied miRNA clusters (**Figure** 3-4 **a**). In particular, members of miRNA cluster 17/92 are upregulated in CPC-EVs. The miRNA 17/92 cluster was initially discovered as an oncogene and has been shown to promote cardiomyocyte proliferation (Chen et al., 2013a; He et al., 2005). Furthermore, miR-92a is highly expressed in endothelial cells and is upregulated in CPC-EVs (log_2fold-change = 1.46, as compared to CPC). Previous studies have shown that increasing cellular expression of miR-92a-3p specifically via EV delivery is pro-angiogenic—promoting cell cycle progression and endothelial-to-mesenchymal transition in endothelial cells—whereas direct cell overexpression of miR-92a-3p may be anti-angiogenic (Liu et al., 2019; Umezu et al., 2013; Yamada et al., 2019). Additionally, members of the miR-23/-24/-27 cluster are implicated in both positive and negative regulation of neovascularization and are differentially expressed in CPC and CPC-EVs (Kesidou et al., 2020; Li et al., 2016a). Overall, identifying well-studied miRNA clusters with known biological roles in our CPC-EVs provides greater insight into their potential mechanisms of action in vitro or in vivo.

Furthermore, we aimed to differentiate non-specific, EV biogenesis-related RNA cargo from CPC-specific, potentially pro-reparative cargo. To do so, we compared the top 15 miRNAs

enriched in CPC-EVs and CPCs to publicly available data sets from various cell types—immune cells, cancer cells, and mesenchymal stem cells—and we found both similarities and differences in the top ranked miRNAs. The 4th most upregulated miRNA in CPC-EVs, miR-122-5p (log_2fold-change = 5.92, **Table 3-2**), is upregulated in EVs from other cell types (**Figure** 3-4 **b**) and is loaded into EVs via Lupus La protein binding (Temoche-Diaz et al., 2019). Several other miRNA EV sorting mechanisms have been identified, including other RNA-binding protein mechanisms and membrane proteins involved in EV biogenesis. For example, three of the miRNAs upregulated in CPC-EVs, miR-320a, -193a, and -92a (log_2fold-change = 1.25, 0.80, 1.46) have been shown to be actively loaded into EVs via Argonaute 2, major vault protein, and vacuolar protein sorting-associated protein 4 binding, respectively (Jackson et al., 2017; McKenzie et al., 2016; Teng et al., 2017).

On the other hand, in our comparison to other datasets, we also identified miRNAs enriched in CPC-EVs that are not enriched in EVs from other cell types. Most notably, miR-18a-5p is upregulated in CPC-EVs (log_2fold-change = 5.56) and is a member of the pro-proliferative and anti-apoptotic 17/92 miRNA cluster. In contrast to some other cell type EVs, miR-501-3p was also identified as a top upregulated miRNA in CPC-EVs. A previous report determined that macrophage-derived exosome miR-501-3p promoted pancreatic ductal adenocarcinoma cell migration and proliferation (Yin et al., 2019). Considering treatment strategies for cardiac repair strive to promote cell proliferation, miR-18a-5p and miR-501-3p may be potent, progenitor cell-specific EV signals driving therapeutic response. Further investigation of these miRNAs is warranted. Overall, when understanding the function of released EVs in cell therapy, it will be important to discriminate between non-specific, machinery-related RNA cargo vs. CPC-specific,

potentially beneficial RNA. Elucidating CPC-specific EV cargo molecules will help us understand which RNA molecules are driving cardiac cell therapy efficacy.

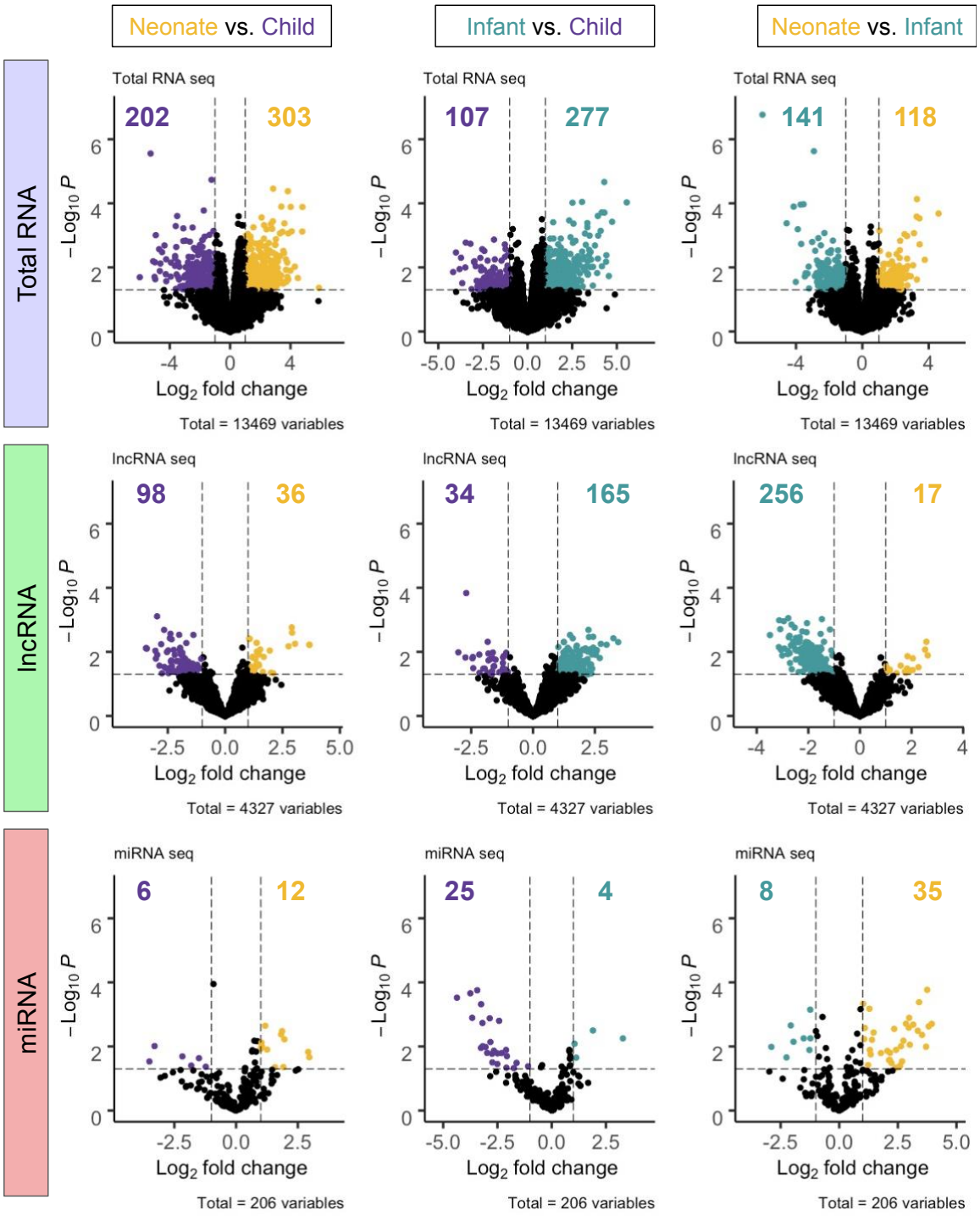
We have previously examined and reported differential CPC mRNA expression across neonate, infant, and child age groups in a smaller sample size (Shoja-Taheri et al., 2019). We have not, however, investigated the full RNA landscape and potential interactions of coding and non-coding RNAs in CPCs. Importantly, in 2011, Salmena et al. introduced the competing endogenous RNA (ceRNA) hypothesis, suggesting that RNA transcripts "talk" to each other via miRNA response elements, forming a large-scale transcriptome regulatory network (Salmena et al., 2011). A growing body of evidence suggests that lncRNAs play a key role in protein-coding gene regulation by acting as miRNA sponges (Ma et al., 2020; Qi et al., 2015; Salmena et al., 2011). Furthermore, non-coding RNAs play an important role in cardiac development (Frank et al., 2016), and CPCs lose their therapeutic functionality as they age (Agarwal et al., 2016; Sharma et al., 2017). Thus, to gain a comprehensive level of understanding of age-dependent CPC ceRNA interactions, we investigated the differentially expressed miRNAs, lncRNAs, and mRNAs between neonate and child CPCs. We determined that neonate and child CPCs had the greatest differences in RNA content. Infant CPCs had an "intermediate" RNA profile and were not included in the pairwise ceRNA network analysis (**Figure** 3-5 **a; Supplemental Figure. 3-1**) Largely, child CPCs had higher expression of non-coding RNAs, compared to neonate CPCs (**Figure** 3-5 **c**). Using the miRcode and miRTarBase target prediction databases, we mapped a lncRNA-miRNA-mRNA network of differentially expressed RNAs between neonate and child CPCs (**Figure** 3-5 **d-f**). The resulting network highlighted the importance of highly connected miR-218-5p, -181a-5p, -23a-3p, and -23b-3p (**Table 3-3**).

Our findings presented in this report are constrained by CPC in vitro culture conditions. Expansion of CPCs for the collection of EVs demands tens of millions of cells, requiring in vitro passaging of cells and potentially introducing transcriptome drift. To limit this issue, we kept cultures less than or equal to passage nine. The interpretation of these results is also limited due to the CPC two-dimensional culture experimental design, which is not ideal to recapitulate the in vivo cellular environment. Furthermore, CPC-EVs released from in vitro cell cultures may vary from CPC-EVs released from transplanted cells in in vivo and clinical models. EV cargo is highly heterogenous and affected by parent cell conditions and environment. Previous research in allogenic cardiac cell therapy has addressed this issue by collecting exosomes released from human CPCs in rat plasma, after cell transplantation, via major histocompatibility complex class I (Saha et al., 2019). Future efforts to identify CPC-EV markers for autologous and allogenic transplant models will assist our efforts to understand CPC-EV cardiac repair mechanisms of action.

In conclusion, we sequenced patient-derived CPCs and CPC-EVs and examined the RNA profiles with unsupervised learning methods to explore differences in coding and non-coding RNAs. We determined that CPC and CPC-EV RNA profiles differ and that there are age-dependent differences in non-coding RNAs of CPCs. More specifically, CPCs retain cell adhesion-related RNAs and export both generic EV transport-related RNAs and potential progenitor cell-specific and pro-reparative RNAs involved in cell proliferation and neovascularization. Further, child CPCs contain elevated levels of non-coding RNAs, compared to neonate CPCs. With this study, we hope to highlight the value of using unbiased methods as "precursors" to quickly hone in on potentially important CPC and CPC-EV RNAs so that more targeted experimental tests may be performed. Cell therapy for children is currently in early

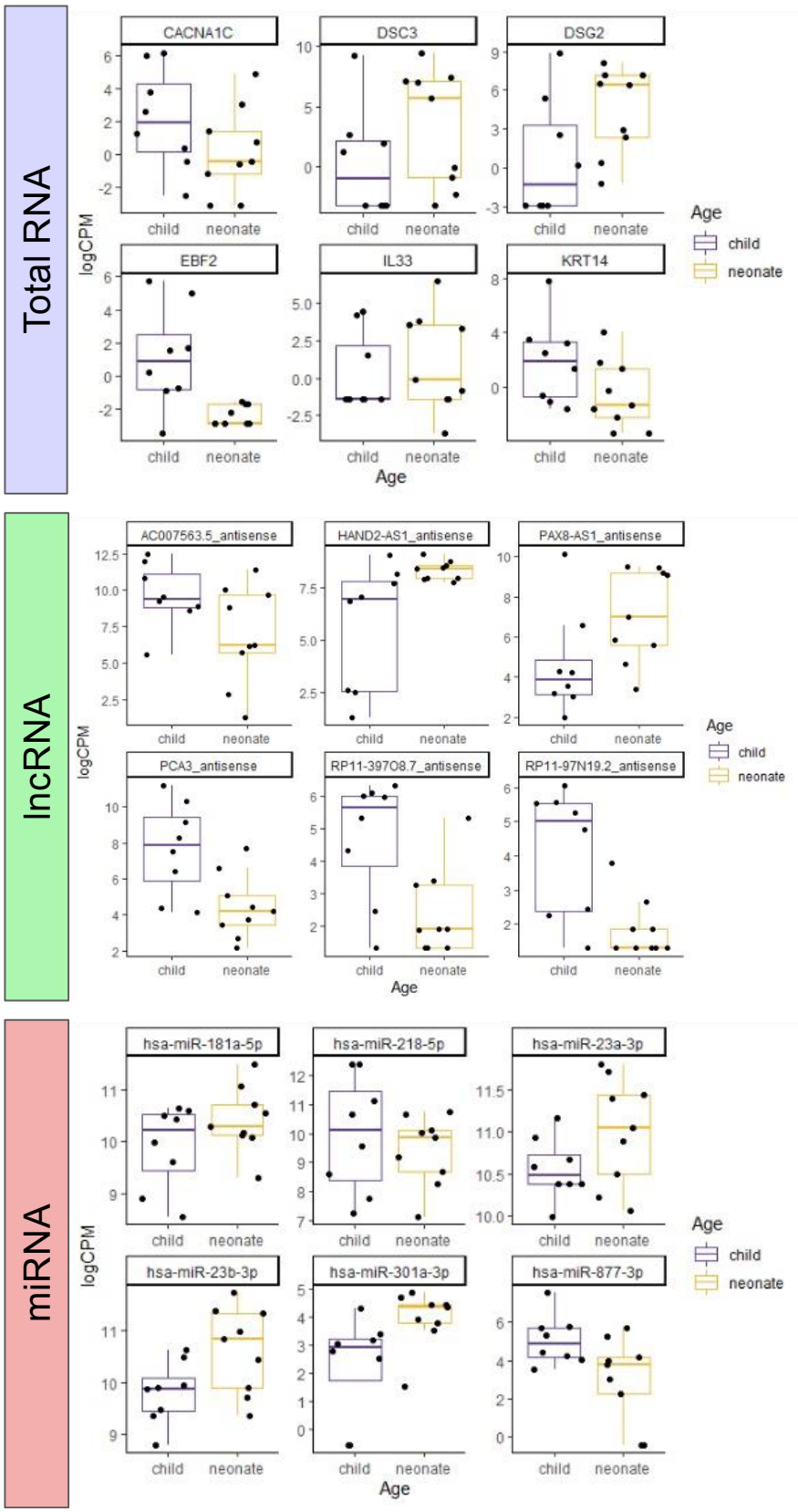
clinical trials, and data from our laboratory and others show that cells and EVs may independently, or in concert, repair the damaged myocardium (Ibrahim et al., 2014; Saha et al., 2019). Using bulk sequencing to develop tools to computationally assess mechanisms and biomarkers in an unbiased manner could improve the outcomes of this promising approach. Our work provides further perspective for understanding the mechanism of action of CPCs, which is valuable for addressing clinical trial variability.

3.6 Supplemental Information



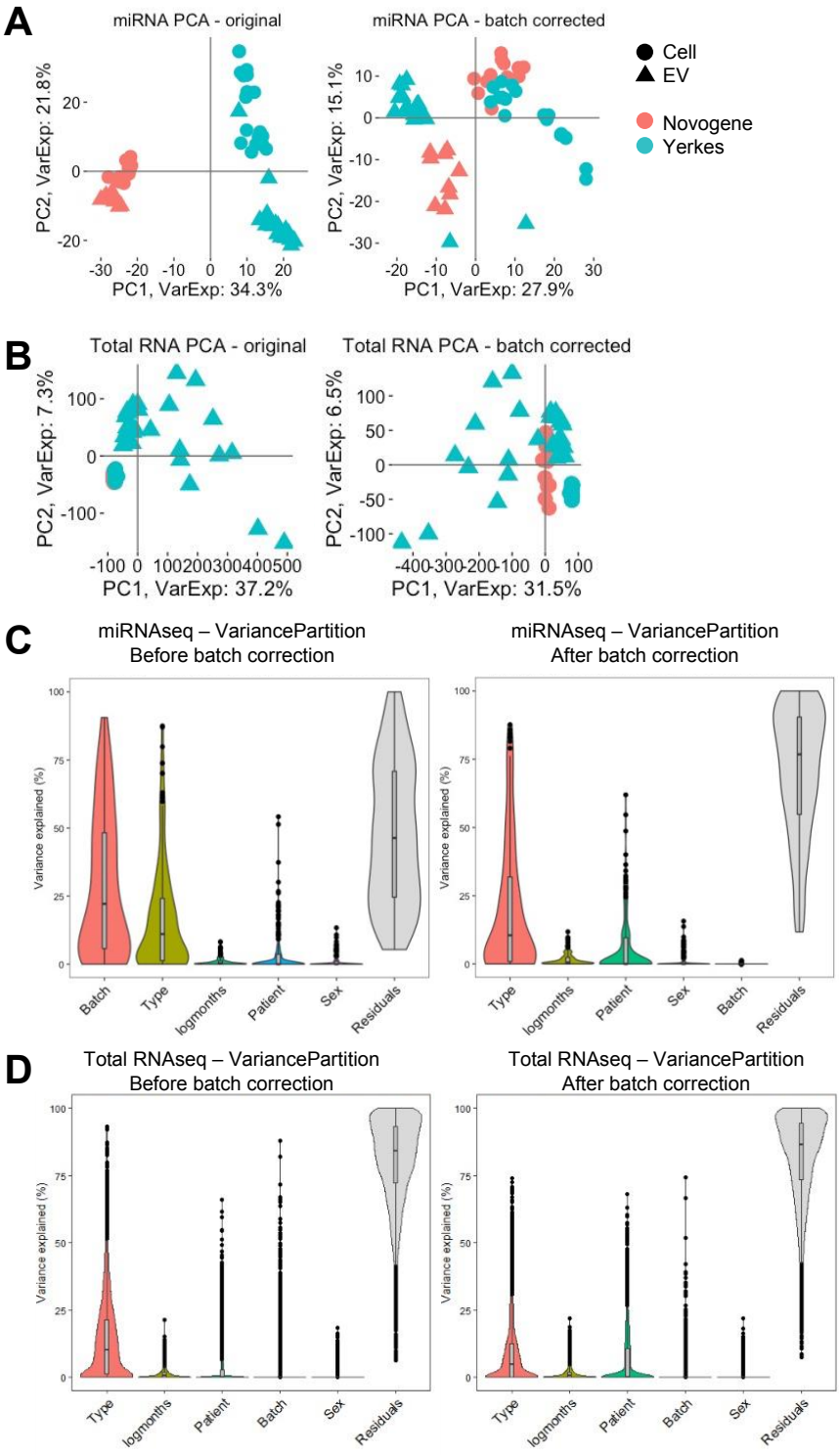
Supplemental Figure 3-1. Differentially expressed RNAs in neonate, infant, and child CPCs.

Volcano plots of differentially expressed RNAs from total RNAseq, lncRNAseq, and miRNAseq between different age CPC groups. p-value cutoff = 0.05, fold-change cutoff = 2.



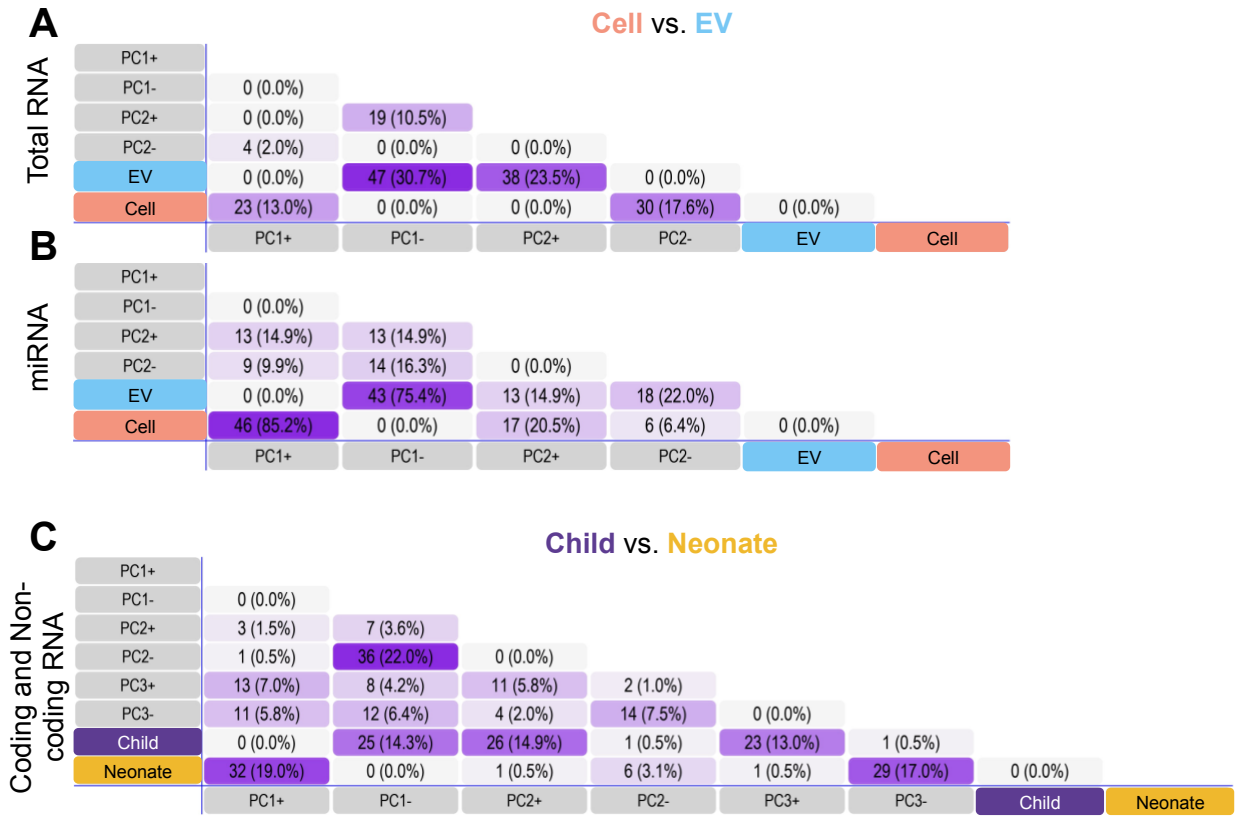
Supplemental Figure 3-1. Differentially expressed RNAs in neonate and child CPCs.

LogCPM values for top differentially expressed mRNAs, lncRNAs, and miRNAs.



Supplemental Figure 3-2. miRNAseq and Total RNAseq after batch correction.

a PCA plot of miRNAseq data (logCPM) before and after ComBat batch correction
b PCA plot of total RNAseq data (logCPM) before and after ComBat batch correction
c Violin plot of miRNAseq variance explained by each covariate before and after batch correction. Variance calculated for each miRNA.
d Violin plot of total RNAseq variance explained by each covariate before and after batch correction. Variance calculated for each RNA.

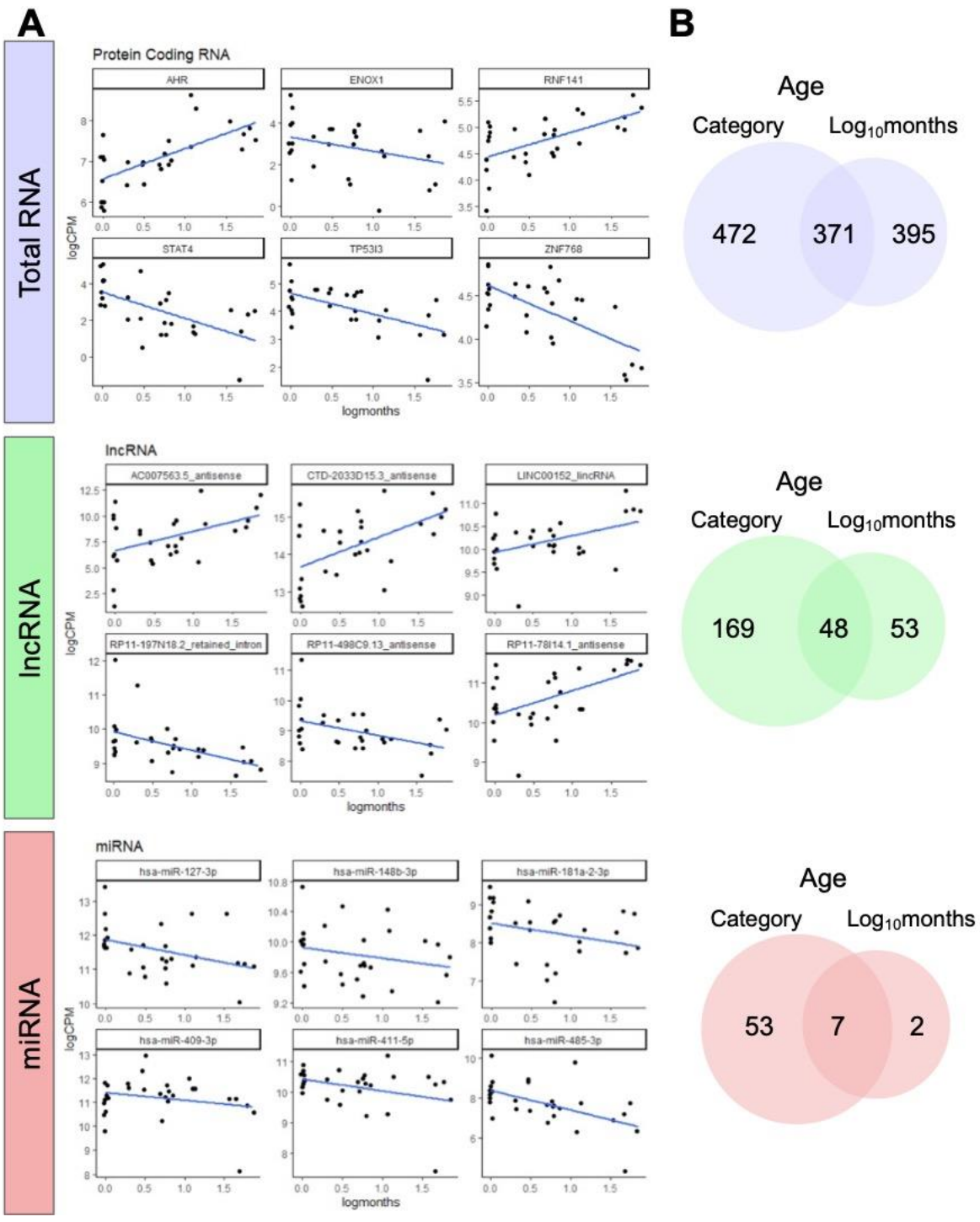


Supplemental Figure 3-3. PCA results correspond to differential expression analyses

a Top 100 RNAs that load onto principal component (PC) 1 and 2, both positively and negatively, were compared to top 100 differentially expressed RNAs (Cell vs. EV, by fold-change). Results correspond to PCA plot in Figure 3-2b: RNAs upregulated in EVs match RNAs heavily loaded on PC1- and PC2+, and RNAs upregulated in cells match RNAs heavily loaded on PC1+ and PC2-.

b Top 50 miRNAs that load onto PC1 and PC2 were compared to the top 50 differentially expressed miRNAs (Cell vs. EV, by fold-change). Results correspond to PCA plot in Figure 3-3b: miRNAs upregulated in cells and EVs match miRNAs heavily loaded on PC1+ and PC1-, respectively.

c Top 100 coding/non-coding RNAs that load onto PC1, PC2, and PC3 were compared to the top 100 differentially expressed CPC RNAs (child vs. neonate, by fold-change). Results correspond to PCA plot in Figure 3-5a: RNAs upregulated in child CPCs match RNAs loaded on PC1-, PC2+, PC3+; RNAs upregulated in neonate CPCs match RNAs loaded on PC1+ and PC3-.



Supplemental Figure 3-4. Differentially expressed CPC RNAs by patient age.

a Top six differentially expressed CPC protein coding RNAs, lncRNAs, and miRNAs by patient age in log10months. **b** Overlap of differentially expressed RNAs ($p<0.05$) as determined by categorical ages (neonate, infant, and child) and continuous age (log10months). CPC models constructed with the covariates age, sex, and batch (if applicable).

Chapter 4: Statistical Modeling of Extracellular Vesicle Cargo to Predict Clinical Trial Outcomes for Hypoplastic Left Heart Syndrome

Jessica R. Hoffman[1,2], Hyun-Ji Park[1], Sruti Bheri[1], Michael E. Davis[1,2,3]

[1]Wallace H. Coulter Department of Biomedical Engineering, Emory University School of Medicine and Georgia Institute of Technology, Atlanta, GA, USA.

[2]Molecular and Systems Pharmacology Graduate Training Program, Graduate Division of Biological and Biomedical Sciences, Laney Graduate School, Emory University, Atlanta, GA, USA.

[3]Children's Heart Research and Outcomes (HeRO) Center, Children's Healthcare of Atlanta and Emory University, Atlanta, GA, USA.

The work contained in this chapter will be submitted for publication summer 2022.

4.1 Introduction

Congenital heart disease (CHD) affects nearly 1% of births in the United States, with conditions ranging in severity. Hypoplastic left heart syndrome (HLHS) is one of the most complex forms of CHD and is characterized by an underdeveloped left ventricle (Bittle et al., 2018; Feinstein et al., 2012). HLHS is palliated by a series of three surgeries which redirects blood blow such that the right ventricle sustains systemic circulation. Despite the short-term improvements offered by surgical palliation, HLHS has one of the highest mortality rates among all CHD conditions (Saraf et al., 2019). In particular, HLHS patients suffer from right ventricle failure due to ischemia and pressure/volume overload. Therefore, to address right ventricle dysfunction and improve HLHS patient outcomes, cell-based therapies have been explored in several preclinical and clinical trials (Bittle et al., 2018). Notably, our group is currently investigating the use of autologous c-kit+ progenitor cells (CPCs) for the treatment of HLHS in a Phase I/II clinical trial (CHILD: NCT03406884) (Kaushal et al., 2022).

CPCs may be isolated from the heart tissue, including the otherwise discarded right atrial appendage, and expanded in culture for preclinical or clinical use. Multiple studies have shown that CPCs derived from neonatal patients outperform CPCs derived from older children and adults, in part due to their differences in CPC secreted factors (Agarwal et al., 2017; Agarwal et al., 2016; Mishra et al., 2011; Sharma et al., 2017; Simpson et al., 2012). Originally, cardiac cell therapy was thought to function in a direct manner: transplanted cells engraft, proliferate, and form new healthy tissue. However, many groups have now shown that transplanted cells function in a more indirect manner via paracrine signaling (Marban, 2018). More specifically, transplanted stem or progenitor cells release extracellular vesicles (EVs) loaded with beneficial nucleic acids to resident cardiac cells (Ibrahim et al., 2014).

Small EVs may be formed in an endocytic manner from a parent, or source, cell. In a series of inward budding steps—first from a parent cell's plasma membrane and then multivesicular body—EVs may be formed as intraluminal vesicles that are ultimately released into the extracellular space to signal to neighboring cells. Importantly, during these biogenesis steps, EVs acquire specific proteins and nucleic acid cargo directly from the parent cell. We and others have shown that stem and progenitor cells are enriched in certain RNAs that promote processes such as cell proliferation, cell migration, and angiogenesis, that may drive therapeutic success of cell therapy preclinical models (Agarwal et al., 2017; Gray et al., 2015; Hoffman et al., 2022; Ibrahim et al., 2014; Sahoo et al., 2011; Yu et al., 2015). To better understand the mechanistic EV RNA signals driving reparative outcomes, we have previously built regression models to link EV RNA-sequencing data to *in vitro* and *in vivo* experimental outcomes (Agarwal et al., 2017; Gray et al., 2015; Saha et al., 2019; Trac et al., 2019b). However, these studies have been limited in sample size and have not been connected to clinical trial samples.

Despite growing evidence of stem/progenitor cell-derived EVs repairing the myocardium, there is a dearth of quantitative studies investigating the EV specific factors (mRNA, miRNA, proteins, etc.) contributing to repair. Indeed, cardiac cell therapy had suffered from variability in patient outcomes; some cell types and/or patients exhibit greater improvements than others. Therefore, to identify contributing factors of cell therapy variability and improve clinical outcomes, we need studies linking high dimensional datasets, like next generation sequencing, to clinical outcomes. Here, we construct an *in vitro* model of CPC-EV "repair" for the prediction of clinical outcomes from the CHILD trial. We collected and sequenced EVs from cultures of (1) CPCs previously isolated from CHD patients (n=29), and (2) CPCs from patients in the CHILD clinical trial (n=7). Then, we treated various cell types (cardiac endothelial cells, cardiac fibroblasts, and mesenchymal stromal cells) with CPC-EVs and measured their effects. We put together the EV sequencing data and in vitro outcomes data to form regression models, potentially predictive of clinical outcomes.

4.2 Methods

4.2.1 Isolation and Culture of c-kit+ Progenitor Cells (CPCs)

This study was approved by the Institutional Review Board at Children's Healthcare of Atlanta and Emory University. Human CPCs were isolated from the right atrial appendage tissue, routinely removed during surgical repair of congenital heart defects, via magnetic cell sorting (CD-177, BioLegend, San Diego, USA). Cells were collected from 16 neonatal (<1 month), 13 infant (<1 year), and 8 child (>1 year) patients. Cells were cultured and expanded in Ham's F-12 medium (Corning Cellgro®, Corning, NY, USA) with 10% fetal bovine serum, 1% penicillin-streptomycin, 1% L-glutamine, and 0.04% human fibroblast growth factor-β. The adherent CPCs

were expanded to reach confluency in 700cm^2. Patient characteristics for samples used in the study are listed in Table S1.

4.2.2 Extracellular Vesicle (EV) Collection

CPCs were grown to 90% confluency, washed with PBS, and quiesced with serum free medium for 24 hours. Conditioned media was collected and subjected to sequential centrifugation: 3000 g for 10 min to remove cells, 28,000 g for 30 min to remove cell debris, and 118,000 g for 1 hr 54 min to pellet EVs (Optima XPN-100 ultracentrifuge; Beckman Coulter SW 41 Ti rotor). EV protein content was analyzed by Micro BCA Protein Assay Kit (Thermo Scientific Pierce 23235), according to manufacturer's instructions. EV size and concentration was determined by NanoSight NS300. Samples were diluted 1:10 in PBS, and three, 60-second video images were captured per sample and analyzed by NanoSight NTA 3.4 software.

4.2.3 Tube Formation Assay

Rat cardiac endothelial cells (CECs) were cultured in endothelial cell growth medium (Endothelial Cell Growth Medium-2 BulletKit™, Lonza, Bend, OR). Before experimentation, CECs were washed with PBS and quiesced in endothelial bare media (FBS and growth factor free) with 1% penicillin-streptomycin overnight. Quiesced CECs were seeded at 10,000 cells/well onto µ-slide Angiogenesis slides (IBIDI) pre-coated with 10 µL/well Matrigel (Matrigel® Matrix, Corning) or Geltrex (Geltrex™ LDEV-Free hESC-qualified Reduced Growth Factor Basement Membrane Matrix, Gibco) with three technical replicates per group. CECs were treated with 20 µg/mL protein of EVs in 50 µL of media per well. After 20 hours, live cells were stained with calcein-AM (Thermo Fisher Scientific) and imaged with fluorescent microscopy (Olympus IX71). The ImageJ Angiogenesis Analyzer plug-in was used to quantify number of tubes and total tube length (Fiji, National Institutes of Health, Bethesda, MD, USA).

Quiesced CECs and CECs grown in full growth medium (with FBS and growth factors) without EV treatment served as a negative and positive controls, respectively. Tube formation experiments were performed five times (n=5) for each CPC-EV.

4.2.4 Mesenchymal Stromal Cell (MSC) Migration Assay

Bone marrow-derived MSCs were purchased (StemPro™ BM Mesenchymal Stem Cells, Gibco) and grown in a 1:1 mixture of Dulbecco's Modified Eagle Medium and Ham's F-12 media (Corning Cellgro®, Corning, NY, USA) with 10% fetal bovine serum, 1% penicillin-streptomycin, 1% L-glutamine, and 0.04% human fibroblast growth factor-β. Before experimentation, cells were washed in PBS and quiesced in serum free media overnight. The bottom of a 24 well plate was coated in 0.1% gelatin for 1 hour. Then, excess gelatin was aspirated and cells were seeded onto a Transwell Insert with 8 μm pore (Corning® Transwell® polycarbonate membrane cell culture inserts, Corning, NY, USA) and placed in the 24 well plate (two technical replicates per group). The basolateral compartment was treated with 20 μg/mL protein of EV in 300 μL of media. After 48 hours, cells that migrated through the porous membrane were detached and stained (CellTracker™ Orange CMRA Dye, Invitrogen). Fluorescence was detected by the Synergy 2 Microplate Reader (Biotek, Winooski, VT, USA) and fold-change was computed over the negative control (MSCs without EV treatment). MSC migration experiments were performed four times (n=4) for each CPC-EV.

4.2.5 Fibroblast TGF-β Stimulation Assay

Rat cardiac fibroblast cells (CFs) were cultured in a 1:1 mixture of Dulbecco's Modified Eagle Medium and Ham's F-12 media (Corning Cellgro®, Corning, NY, USA) with 10% fetal bovine serum, 1% penicillin-streptomycin, and 1% L-glutamine. At 90% confluency, cells were washed with PBS and quiesced in serum free media. CFs were treated with 20 μg/mL protein of EV in

500 µL of media per well (two technical replicates per group). After 12 hours, CFs were stimulated with 10 ng/mL of TGF-β (Invitrogen™ TGFB1 Recombinant Human Protein, Invitrogen) for 12 hours. Then, media was aspirated, and RNA lysate was collected for PCR analysis. CFs without EV treatment, with and without TGF-β treatment served as positive and negative controls, respectively. Fibrotic gene assays were performed three times (n=3) for each CPC-EV.

4.2.6 Endothelial Cell TNF-α Stimulation Assay

CECs were cultured as previously described in 24 well plates. At 90% confluency, cells were washed with PBS and quiesced in endothelial bare medium overnight. CECs were treated with 20 µg/mL protein of EV in 500 µL of media per well (two technical replicates per group). After 24 hours, CECs were stimulated with 20 ng/mL of rat TNF-α (Recombinant Rat TNF-alpha Protein, R&D Systems) for 4 hours. Then, media was aspirated, and RNA lysate was collected for PCR analysis. CECs without EV treatment, with and without TNF-α treatment served as positive and negative controls, respectively. Inflammatory gene assays were performed three times (n=3) for each CPC-EV.

4.2.7 Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

RNA from CECs and CFs was collected in lysis buffer from the Pure Link RNA Mini Kit (Life Technologies, Carlsbad, CA) with 1% 2-mercaptoethanol (Sigma-Aldrich). Total RNA was isolated with the kit, according to manufacturer's instructions. Next, cDNA was prepared, and RT-qPCR was performed on the StepOne System (Applied Biosystems, Foster City, CA) based on SYBR Green fluorescence detection of PCR products (Power SYBR Green PCR Master Mix, Applied Biosystems).

For CFs, RT-qPCR was used to evaluate transcript expression of connective tissue growth factor (*Ctgf*), collagen type 1 pro-α1 chain (*Col1a1*), collagen type 1 pro-α2 chain (*Col1a2*), collagen type 3 pro-α1 chain (*Col3a1*), and vimentin (*Vim*). For CECs, RT-qPCR was used to evaluate transcript expression of interleukin 1β (*Il-1β*), interleukin 1α (*Il-1a*), and interleukin 6 (*Il-6*). Relative mRNA levels were calculated using the 2−ΔΔCt method, as compared to Glyceraldehyde 3-phosphate dehydrogenase (*Gapdh*) housekeeping gene.

4.2.8 Next Generation Sequencing

In this study, we combined sequencing data from our previously published CPC-EVs (training data) with CHILD trial patient CPC-EVs (testing data). The previously published training data and information regarding RNA preparation and sequencing can be found under "E" EV samples in GSE202345 and GSE202347. Additionally, we sequenced CPC-EVs from patients 938 and 1097, which were not included in our previous publication. New CPC-EV sequencing data and information can be found in GSE203512. Briefly, we isolated RNA with the miRNeasy Mini Kit (Qiagen), according to manufacturer's instructions. We analyzed purified RNA (2100 Bioanalyzer and TapeStation Controller, Agilent Genomics) for size, quality, and quantity. We conducted RNA library preparation and sequencing at the Emory Yerkes Nonhuman Primate Genomics core (Illumina NovaSeq 6000).

Small RNA sequencing were aligned and hits were determined using the Qiagen GeneGlobe console with QIAseq miRNA Quantification tool. Default parameters were used: 3' adapters were trimmed using cutadapt, and reads with less than 16 base pair insert sequences or less than 10 base pair unique molecular index sequences were removed. Reads were aligned with Bowtie with GRCh38 and miRbase v21 references. Total RNA sequencing were aligned, and gene counts were determined with STAR in the Illumina BaseSpace app, RNA-Seq Alignment.

Reads were aligned with the hg19 reference genome. Biotypes were matched to alignment results using the Ensembl based annotation package (EnsDb.Hsapiens.v79). miRNAs were considered only from the small RNA sequencing results and were thus removed from the total RNA sequencing data set.

4.2.9 RNA Sequencing Data Analysis

Data analysis was completed in R. First, RNA counts for CPC-EVs were filtered: we removed RNAs with zero count entries in twenty or more samples and used edgeR package's 'filterByExpr' function with the default parameter settings. RNA counts were normalized with edgeR's weighted trimmed mean of M-values method (default parameters) and transformed into $\log_2$ counts per million (logCPM). Batch correction of logCPM values was implemented to account for sequencing performed at separate sites and times using limma's removeBatchEffect function.

4.2.10 WGCNA gene module detection

The WGCNA R package was used to construct co-expression networks for the filtered, normalized genes. The details of this algorithm are described by Langfelder and Horvath (Langfelder and Horvath, 2008). Briefly, the optimal soft-threshold power was graphically determined ($\beta = 9$) and the minimum module size was set to 50. Clusters, or modules, of RNAs were determined by first computing the adjacency matrix and then transforming it to form the topological overlap matrix (TOM). Then, the corresponding dissimilarity matrix, 1-TOM, and the cutreeDynamic function was used for hierarchical clustering and module detection. Highly correlated modules ($r > 0.85$) were merged to form the final co-expression modules. Thirty-three modules were determined. The dissimilarity of the module Eigengenes was computed with the moduleEigengenes function and the association between Eigengene values and experimental

outcomes were assessed by Spearman's correlation. Modules that correlated to multiple outcomes ($p<0.1$) were examined for biological significance with Metascape pathway analysis (Zhou et al., 2019).

4.2.11 Regression Models

RNA-sequencing and experimental outcomes data were scaled and centered with the mdatools package in R before use in regression models. The mdatools package was also used to construct four partial least squares regression (PLSR) models using the SIMPLS algorithm. First, 3-component models were constructed using all features (RNA) and leave-one-out cross validation. VIP scores were calculated for each model and RNAs with an average score >2 across all outcomes in the category were selected. VIP type was categorized by sequencing dataset: from total RNA-sequencing vs. small RNA-sequencing. Pathway analysis was performed using Metascape (Zhou et al., 2019). Then, 3-component final, reduced models were constructed (2-component for migration). Model performance of the cross-validated training set was assessed with root-mean-square error (RMSE) and R^2 measurements. Finally, predictions for *in vitro* outcomes were predicted from the reduced models for the CHILD CPC-EV RNA-seq data. Observed vs. predicted plots for each outcome of the CHILD testing set were generated with the following statistics: RMSE, R^2, nLV (# of components).

Random forest regression was performed using the ranger and caret R packages. Models were constructed from the training data for each experimental outcomes using 5-fold cross validation. Hyperparameters mtry (number of features to consider at each split) and splitting rule (extra trees vs. variance) were tuned and the combination with the lowest RMSE were selected. Variable importance was determined by Gini index. Reduced models were constructed with

features > 10 importance score. The top 100 most important features for each outcome/model were considered for pathway analysis.

4.3 Results

4.3.1 CHILD clinical CPC samples release small EVs in cell culture

To build a predictive CPC-EV model of cardiac outcomes with an adequate sample size, we cultured previously collected CPCs from patients with various congenital heart conditions (training dataset, n=29), as well as CPCs from HLHS patients enrolled in phase I of the CHILD clinical trial (testing dataset, n=7). Patients' age and heart conditions are listed in **Table** 4-1. We expanded CPCs in 2D culture and isolated their EVs from the conditioned media via differential ultracentrifugation (Figure 4-1 a). We assessed CPC-EV concentration and size with Nanosight particle tracking and determined that CPC-EVs were <150nm, characteristic of small EVs or exosomes (**Figure** 4-1 **b**). Next, we isolated CPC-EV RNA and performed total and small RNA sequencing. Initial bioanalyzer plots of CPC-EV RNA revealed enrichment of small RNAs ~22nt, the size of miRNA, and confirmed the absence of ribosomal 18S and 28S RNA peaks (**Figure** 4-1**c**). Sequencing results identified 1,067 miRNAs and 10,469 total RNAs after removing lowly expressed RNAs. CPC-EV sample logCPMs are displayed in **Figure** 4-1 **d**. After processing sequencing results, dimension reduction plots using principal component analysis (PCA) showed comparable miRNA and total RNA vesicle content across both training and testing datasets (**Figure** 4-1 **e**).

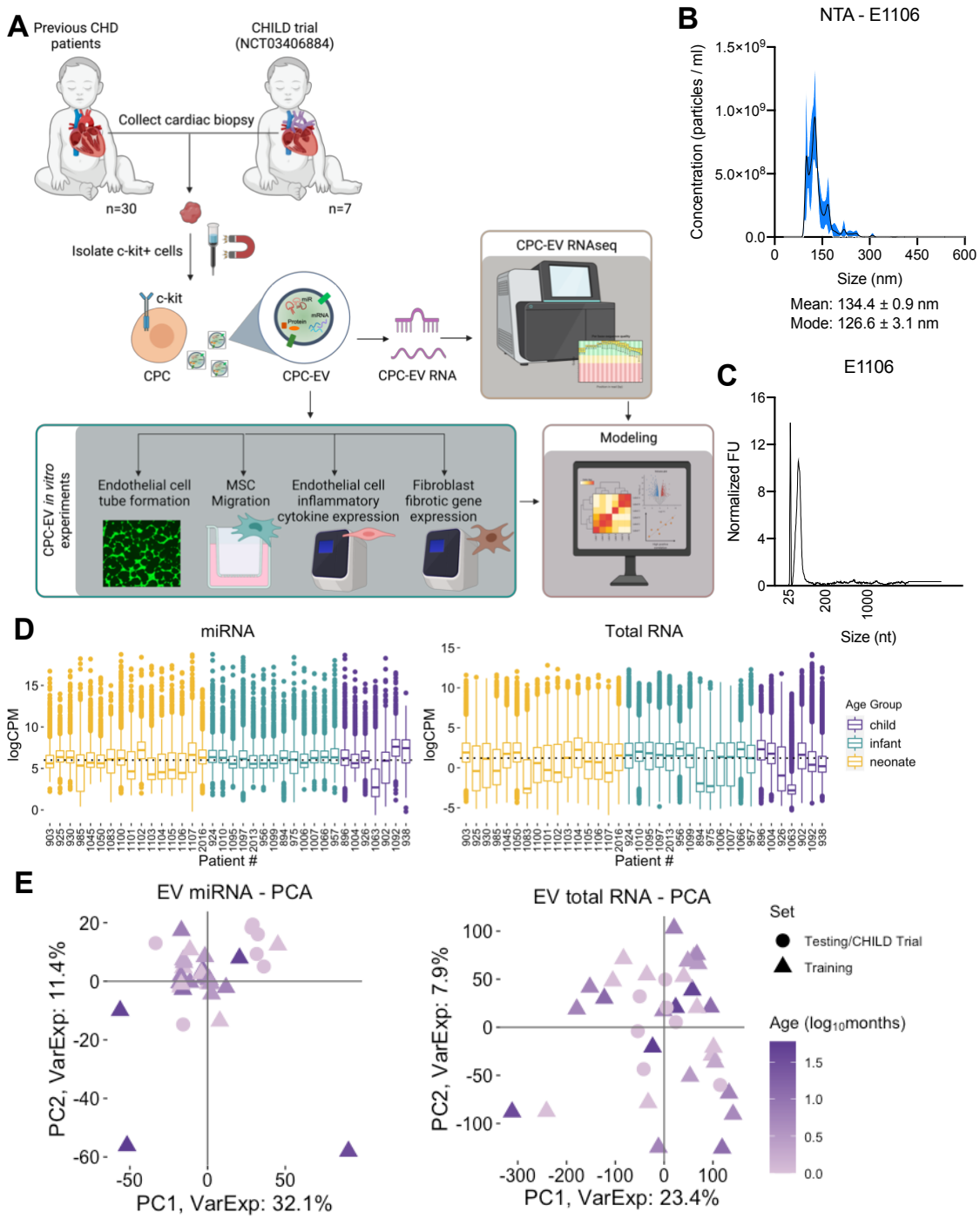


Figure 4-1. Characterization and RNA-sequencing of CPC-EVs.

A Schematic of experimental design: CPCs were previously collected from various CHD patients (n=30), as well as the lead-in patients for the CHILD trial (n=7). CPC-EVs were collected from conditioned media, sequenced, and used in four *in vitro* experiments: cardiac endothelial cell tube formation on matrigel, MSC migration in a boyden chamber, cardiac endothelial cell inflammatory gene expression after TNF-α treatment, and cardiac fibroblast fibrotic gene expression after TGF-β treatment.

B Nanoparticle tracking and **C** bioanalyzer plot of representative CPC-EVs from patient 1106 determine EV size/concentration and RNA composition, respectively.

D Boxplots show normalized counts of CPC-EV samples for miRNA and total RNA-sequencing.

E PCA plots of CPC-EV samples for miRNA and total RNA-sequencing, coded by dataset and patient age.

CPC: c-kit+ cardiac progenitor cell; CHD: congenital heart disease; EV: extracellular vesicle; PCA: principal component analysis; MSC: mesenchymal stem (or stromal) cell.

Table 4-1. Patient characteristics.

Patient #	Age	Age Group	Diagnosis
894	6 mo	infant	
896	12 mo	child	VSD
902	4 year	child	ASD
903	< 1 week	neonate	HLHS
924	2 mo	infant	TAPVR
925	< 1 week	neonate	TAPVR
926	14 Mo	child	VSD
930	< 1 week	neonate	TAPVR, UAVC
938	5 yr	child	Subaortic membrane resection, mitral valve ring
956	5 mo	infant	CAVC
957	7 mo	infant	PAPVR
975	6 mo	infant	ToF
985	2 Weeks	neonate	IAA
1004	12 mos	child	LSVC
1006	6 months	infant	CAVC
1007	6 months	infant	VSD
1010	9 wks	infant	ToF
1045	< 1 week	neonate	CoA, VSD
1048	6 yrs	child	ASD
1050	1 week	neonate	HLHS
1063	3 yrs	child	VSD
1066	6 mos	infant	TA, VSD, ASD
1083	<1 wk	neonate	IAA/VSD
1092	4 yrs	child	ASD
1095	3 mos	infant	DORV
1097	3 mos	infant	ToF
1099	5 mos	infant	CAVC
1100	<1 week	neonate	HLHS
1101		neonate	HLHS
1102		neonate	HLHS
1103		neonate	HLHS
1104		neonate	HLHS
1105		neonate	HLHS
1106		neonate	HLHS
1107		neonate	HLHS
2013	3 months	infant	VSD
2016	1-2 week	neonate	CoA/ASD/VSD

ASD: atrial septal defect; HLHS: hypoplastic left heart syndrome; CAVC: Complete Atrioventricular Canal defect; CoA: Coarctation of the aorta; DORV: Double outlet right ventricle; IAA: interrupted aortic arch; LSVC: persistent left superior vena cava; PAPVR: Partial anomalous pulmonary venous return; TA: Tricuspid atresia; TAPVR: total anomalous pulmonary venous return; ToF: Tetralogy of Fallot; UAVC: unicuspid aortic valve; VSD: ventricle septal defect;

4.3.2 CPC-EV treatment affects recipient cell processes

Multiple groups have reported the pleiotropic effects of transplanted stem or progenitor cell derived EVs (Agarwal et al., 2017; Gray et al., 2015; Ibrahim et al., 2014; Sahoo et al., 2011). To construct an *in vitro* model, predictive of clinical outcomes, we designed experiments to investigate primary mechanisms of EV-mediated cell therapy: modulation of inflammation, fibrosis, cell migration and angiogenesis. Specifically, we treated cardiac endothelial cells (CEC), cardiac fibroblasts (CF), and mesenchymal stem, or stromal, cells (MSCs) with patient-derived CPC-EVs. We measured MSC migration in a Boyden chamber system, CEC tube formation on Matrigel (number of tubes and total tube length), CEC inflammatory gene expression (*Il-1a, Il-1β, Il-6*) after TNF-α treatment, and CF fibrotic gene expression (*Col1a1, Col1a2, Col3a1, Vim, Ctgf*) after TGF-β treatment. Overall, CPC-EVs derived from different patients exerted different effects on recipient cells and *in vitro* outcome measurements clustered by assay type, as expected (**Figure** 4-2 **a**).

In the migration assay, CPC-EVs induced MSC migration as compared to the untreated control (**Figure** 4-2 **d**). Interestingly, CPC-EVs derived from neonate patients promoted MSC migration to a greater extent than CPC-EVs derived from older patients (**Figure** 4-2 **b**). Simple linear regression showed a decrease in MSC migration with age ($r = 0.33$, $m = -0.11$, $p = 0.059$, **Supplemental Figure** 4-1). Conversely, angiogenesis experiments showed the opposite trend: CPC-EVs derived from older patients induced greater CEC tube formation than CPC-EVs from neonate patients (**Figure** 4-2 **b and c**). Linear regression showed enhanced total tube length measurements in CPC-EVs from older patients ($r = 0.36$, $m = 0.16$, $p = 0.035$, **Supplemental Figure** 4-1). Furthermore, we did not observe age-dependent differences in the PCR-based fibrosis and inflammation experiments (**Figure** 4-2 **e and f, Supplemental Figure** 4-1). Overall,

CPC-EV treatment reduced *Il-6* expression after TNF-α treatment and reduced *Col1a2* and *Ctgf* expression after TGF-β treatment. Data for *Col1a1*, *Col3a1*, *Vim*, and Il-1β expression, as well as total number of measured tubes is provided in the supplement (**Supplemental Figure** 4-2). Notably, for all four assays, we observed differences in experimental outcomes among the CPC-EV groups. This variance provides a solid basis when constructing a predictive model and linking CPC-EV RNA-seq with experimental data.

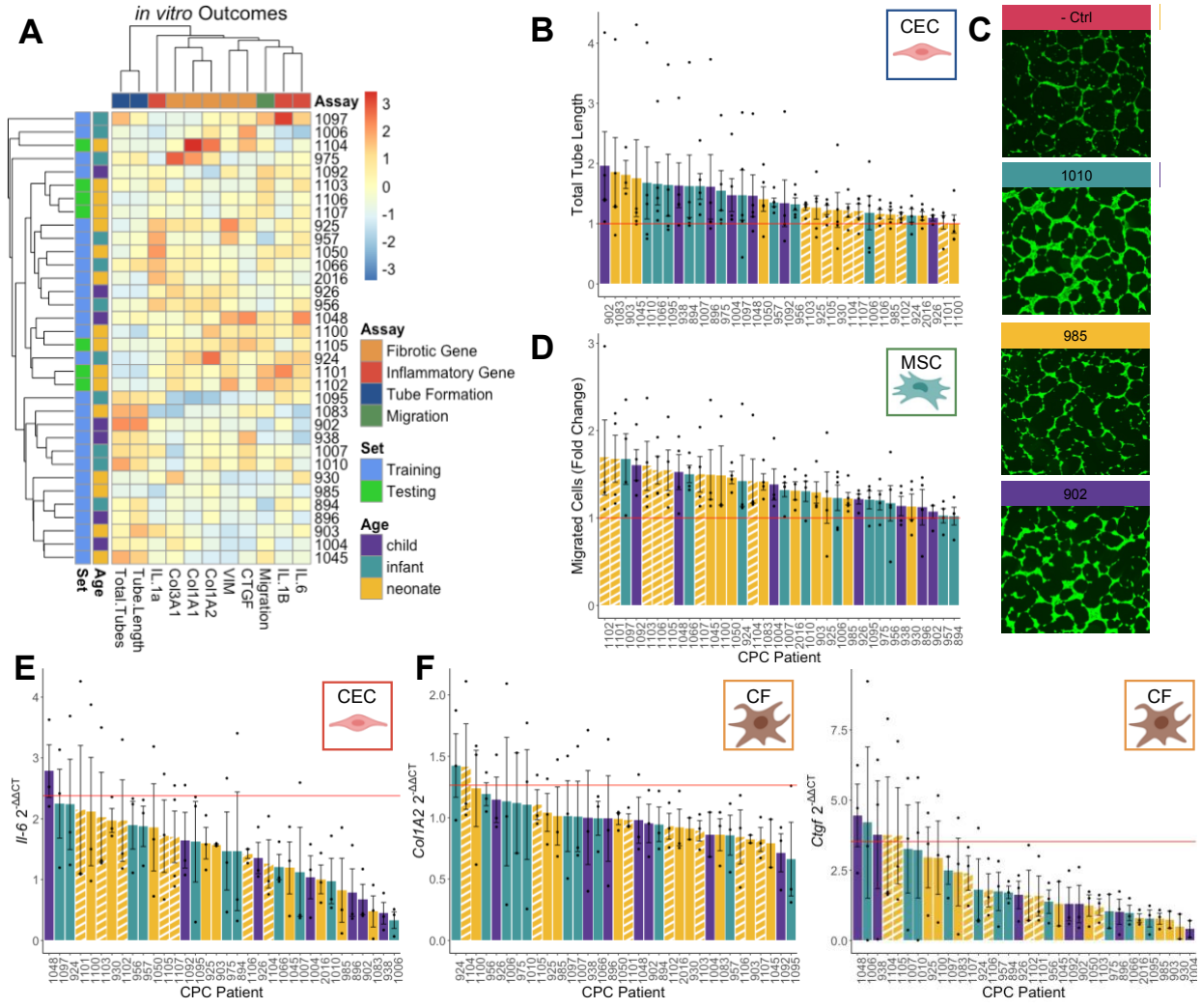


Figure 4-2. CPC-EV treatment affects cell processes in vitro.

A Heatmap of experimental outcomes (averaged) for CPC-EVs cluster by assay category. Patient IDs are listed in the rows and color-coded by age group and data set (training vs. testing). Outcomes are listed in the columns.

B Quantification of tube length, normalized to control, and **C** representative images of CEC tube formation.

D MSC migration in Boyden chamber system, normalized to control, after CPC-EV treatment.

E CEC *Il-6* expression after CPC-EV pretreatment and TNF-α stimulation.

F CF *Col1a2* (left) and *Ctgf* (right) expression after CPC-EV pretreatment and TGF-β stimulation.

CPC: c-kit+ cardiac progenitor cell; CEC: cardiac endothelial cell; CF: cardiac fibroblast; MSC: mesenchymal stem (or stromal) cell.

4.3.3 Weighted gene co-expression network analysis (WGCNA) identifies clusters of co-expressed CPC-EV RNAs which correlate to *in vitro* outcomes

To initially link CPC-EV RNAs with functional outcomes, we used the WGCNA unsupervised learning method. We used WGCNA to first identify clusters, or modules, of CPC-EV RNA and then determine the correlation of these modules to *in vitro* outcomes. For each method, we combined total RNA and miRNA data sets and centered and scaled the RNA features and outcomes.

First, we performed WGCNA and identified 33 modules of co-expressed RNAs (**Figure** 4-3 **a**). Then, we correlated these modules to each outcome and identified 8 modules of interest: M1, M2, M4, M5, M9, M10, M24, and M25. Interestingly, although the RNAs data sets were combined, centered, and scaled, the modules were primarily comprised of either RNA or miRNA (**Figure** 4-3 **b**). RNA modules of interest included M1 and M2 which have a negative correlation to fibrotic gene expression (*Vim*, *Ctgf* or *Col3a1*) and a positive correlation to both tube formation measurements ($p<0.1$). RNA modules M9 and M10 also have a positive correlation to tube length and total tubes ($p<0.05$) (**Figure** 4-3 **c**). Upon investigation of the RNAs belonging to these modules, we determined with pathway analysis that M1, M9, and M10 are enriched in tube morphogenesis; M1, M2, M9 are enriched in cell morphogenesis pathways; and M1 and M2 are enriched in RNAs involved in extracellular matrix organization, circulatory system processes, cell junction organization, and calcium signaling (**Figure** 4-3 **d**).

Additionally, we discovered miRNA modules M4, M5, M24, and M25 to correlate with *in vitro* outcomes. These modules had a positive correlation to expression of various fibrotic and

inflammatory genes. M5 and M25 had a negative correlation to tube formation measurements and MSC migration, respectively. M4 had a positive correlation to MSC migration. More concisely, these modules contained miRNAs which are positively correlated to 'poor' outcomes (fibrosis and inflammation) and negatively correlated to 'good' *in vitro* outcomes (migration and tube formation), opposite to the trend shown in the RNA modules (**Figure** 4-3 **c**). Notably, M4 included some of the most well-studied miRNAs: let-7a/b/d/e/f/g/i, miR-99/100, -23/24/25 families, as well as miR-19b, -20a, -21, -30, -125, -146a, -320a. To understand the biological significance of these modules, we determined the miRNAs' gene targets and performed pathway analyses. From miRNA modules' gene targets, we found enrichment of cell death, adhesion, migration, and differentiation pathways, as well as VEGFA-VEGFR2 signaling, tube morphogenesis, and immune system signaling (**Figure** 4-3 **d**).

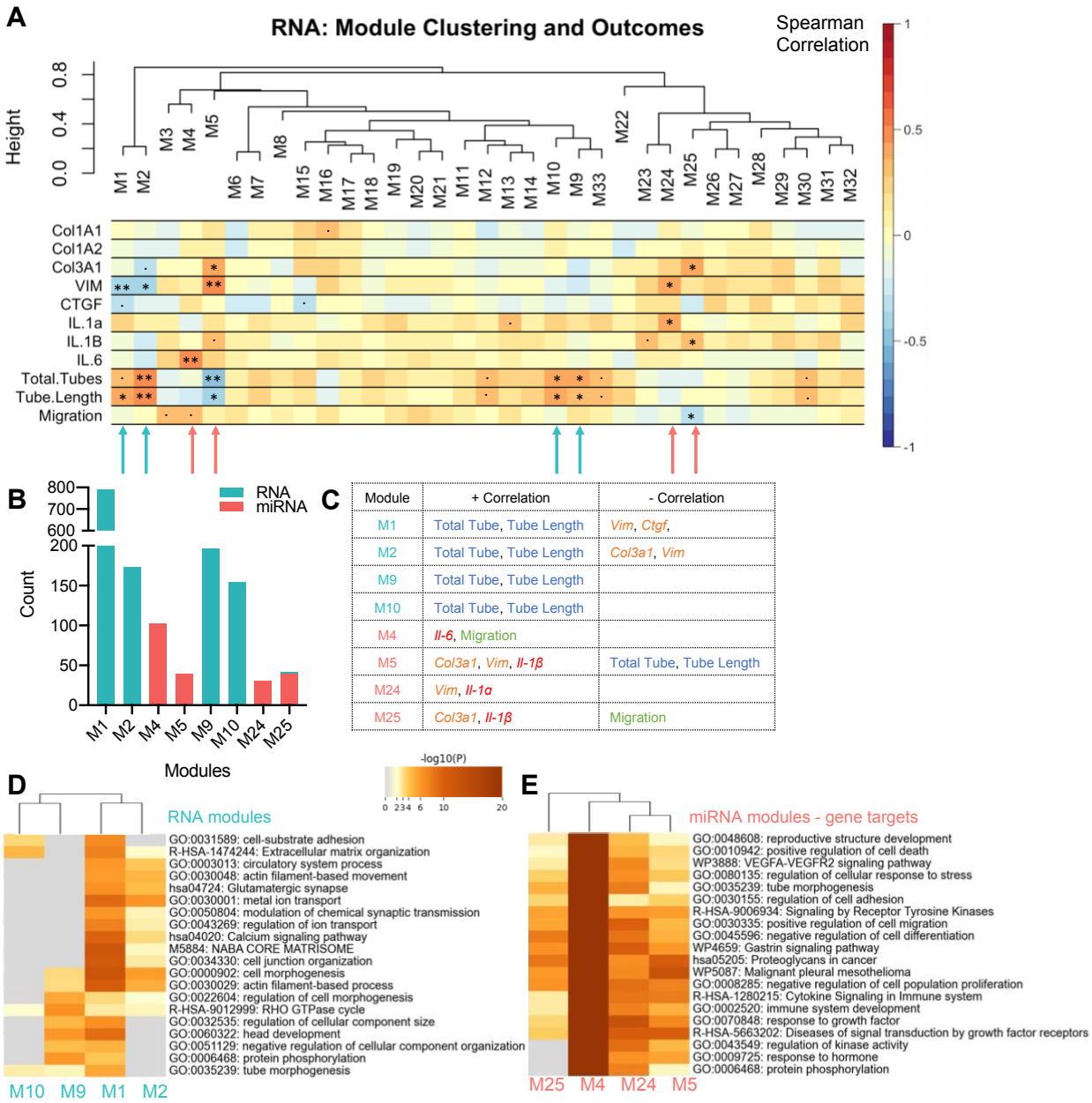


Figure 4-3. WGCNA identifies clusters of RNA which correlate to experimental outcomes.

A Thirty-three RNA modules were identified with WGCNA and then correlated to *in vitro* outcomes. Hierarchical clustering of the RNA modules is displayed in the top dendrogram. The bottom heatmap displays Spearman correlation of RNA modules to each outcome. .p<0.1, *p<0.05, **p<0.01.

B RNA composition of 8 modules which significantly correlate to experimental outcomes – blue: total RNA, pink: miRNA.

C Correlation, positive or negative, of modules to specific outcomes (p<0.1).

D Pathway analysis of RNAs in modules 1, 2, 9, and 10 show enrichment of extracellular matrix organization and cell morphogenesis pathways. Highlighted pathways deemed relevant by the authors.

E Pathway analysis of gene targets of miRNA modules 4, 5, 24, and 25 show enrichment of cytokine and VEGF signaling, as well as cell migration, differentiation, death pathways. Gene targets determined by miRTarBase with validation of at least three experiments.

WGCNA: weighted correlation network analysis

4.3.4 Partial least squares regression models predict CHILD CPC-EV *in vitro* outcomes

We aimed to construct an *in vitro* regression model that could be trained on CPC-EVs collected previously from various CHD patients and would be predictive of clinical samples, CPC-EVs from the CHILD trial. We first determined that *in vitro* outcomes were highly correlated within the outcome categories (**Figure** 4-4 **a**). Therefore, we used PLSR (an algorithm that handles collinear, multivariate outcomes) to construct four regression models: fibrosis, inflammation, angiogenesis, and migration. For each model we first constructed a full model based on all RNAs from the training CPC-EV set and leave-one-out cross validation. The full models did not capture variance (the first two components of each model explained 20-30% of RNA variance) and showed poor prediction performance within the cross-validated training set (**Figure** 4-4 **b**).

However, we then used feature selection to reduce the model and greatly improve performance. We computed the variable importance for the projection (VIP) scores for each RNA and selected RNAs with an average score across outcomes in a category > 2. Feature selection reduced the models to <300 RNAs; VIP count and distribution of RNA type are displayed in **Figure** 4-4 **c**. Noticeably, there was little overlap of VIPs across model categories (**Figure** 4-4 **d**). The reduced models displayed higher performance metrics and represented greater variance (>98% of RNA variance and 74-94% of outcome variance in all models) (**Figure** 4-4 **b**). Pathway analysis of the VIP RNAs showed enrichment of immune system processes, metabolic processes, and developmental processes (**Figure** 4-4 **e**). Finally, we plugged in the test data set and demonstrated that the reduced models predicted *in vitro* outcomes of the CHILD CPC-EV data set. Representative observed vs. predicted plots for total tubes length, *Col1a1*, migration, and *Il-1β* are shown (**Figure** 4-4 **f, Supplemental Figure** 4-3).

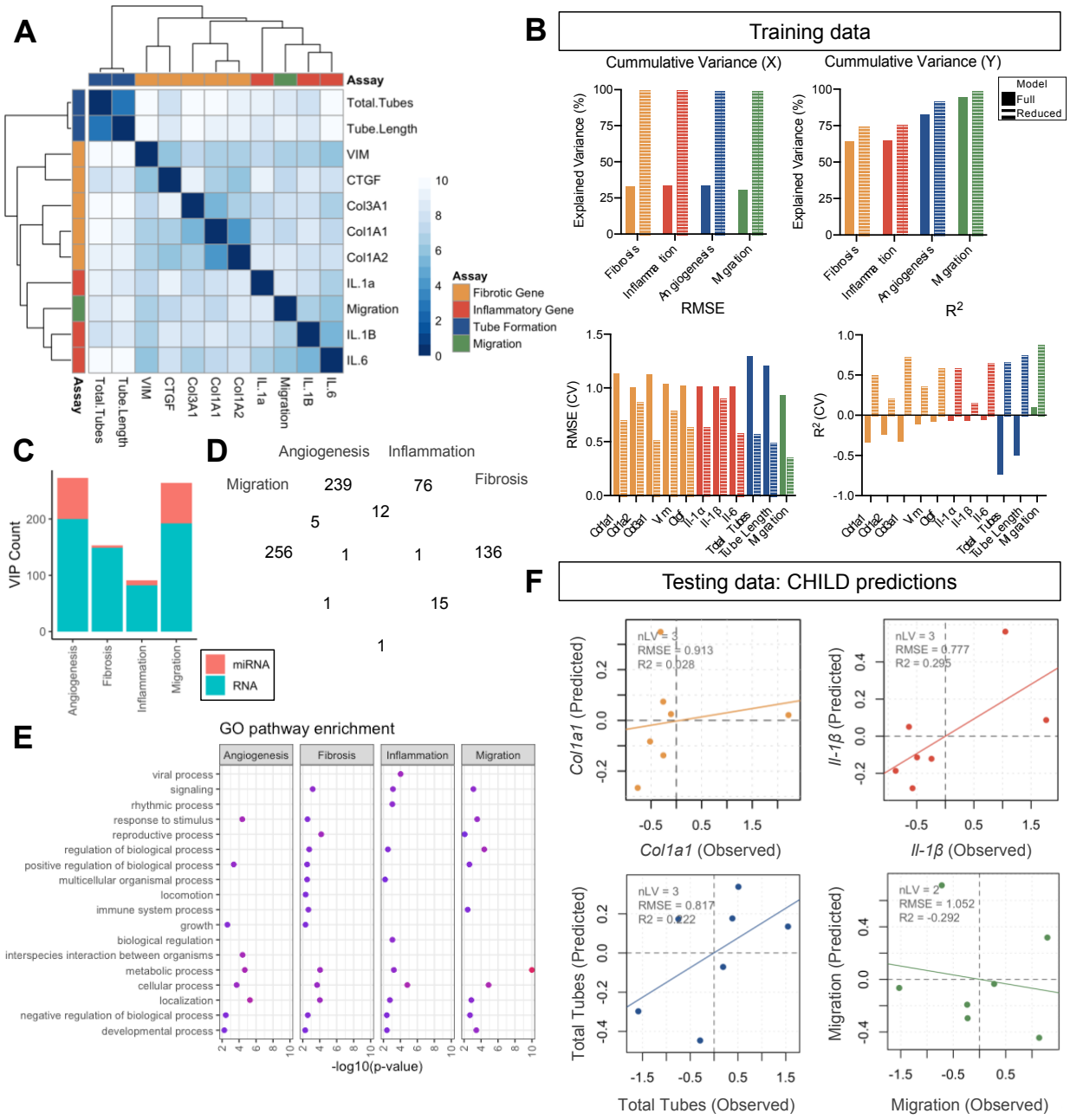


Figure 4-4. Partial least squares regression models predict CHILD in vitro outcomes.

A Experimental outcomes for combined samples cluster by category: angiogenesis, fibrosis, and inflammation. Four PLSR models for each outcome category were created and VIPs with average score >2 were identified.

B Variance explained, RMSE, and R^2, measurements from the four PLSR models, constructed from previously collected CPC-EVs (training set), before and after feature selection. Reduced models are based on VIPs with average score >2.

C composition of VIPs in each model (blue: total RNA, pink: miRNA) and **D** overlap of VIPs across the models**.**

E GO pathway analysis of VIP RNAs in each model.

F *in vitro* predictions for CHILD CPC-EVs. CHILD CPC-EV RNA-seq data were input into PLSR models trained on previously collected CPC-EV RNA-seq and experimental data. Prediction performance for this testing set was determined by observed vs. predicted plots, RMSE, and R^2 metrics.

CPC: c-kit+ cardiac progenitor cell; EV: extracellular vesicle; PLSR: partial least squares regression; RMSE: root mean square error.

4.4 Discussion

Outcomes for infants with HLHS have improved significantly with the emergence of the 3-stage palliation, including the Norwood, Bi-directional Glenn, and Fontan procedures. Surgical palliation supports single-ventricle physiology and redirect blood flow such that the RV maintains systemic circulation. Unfortunately, the RV is not physiologically adapted to support systemic circulation and HLHS patients may develop RV dysfunction, adverse RV modeling, and eventual RV failure. RV dysfunction has been shown to predict mortality: 18-month survival for patients with and without RV dysfunction are 35% and 70%, respectively (Altmann et al., 2000). Furthermore, mortality rates for HLHS patients remain the highest between stage I and II Norwood and Bi-directional Glenn surgeries. Therefore, therapeutic intervention during this stage, preventing RV dysfunction, is necessary to improve outcomes for HLHS patients.

Cell therapy has emerged as a promising avenue to promote cardiac repair and prevent adverse remodeling. Several cell types are under clinical investigation for use in treating HLHS, including CPCs (NCT03406884) cardiosphere-derived cells (NCT01273857, NCT01829750, NCT02781922), umbilical cord blood and bone marrow-derived mononuclear cells (NCT01883076, NCT02549625, NCT03779711, NCT04907526), and bone marrow-derived mesenchymal stem cells (NCT03525418, NCT04925024). In particular, our group is involved in the CHILD clinical trial investigating use of autologous CPCs during the stage II operation to boost the function of the RV and compensate for pressure and volume overload. Unfortunately, cardiac cell therapy suffers from too much variation in cell populations and patient outcomes. Previous work has investigated the genetics of patients receiving MSC injections for non-ischemic dilated cardiomyopathy and determined patients without genetic variants respond better

to the cell therapy than patients with pathological variants (Rieger et al., 2019). However, several groups have shown paracrine signaling plays a critical role in cell therapy function, but there have been a dearth of studies investigating potential paracrine (and EV) factor determinants for cell therapy responsiveness. Here, we investigate the variability of CPC-EVs from >30 CHD patients (including the lead-in patients for CHILD, n=7), and link these transcriptomic data to cardiac-relevant experimental outcomes. We aim to understand which *in vitro* experiments correlate with clinical outcomes, and subsequently, which CPC-EV RNA molecules may be driving these functional responses.

Initially, we confirmed that CPC-EVs derived from various patients differentially affect cell processes (**Figure** 4-2). We determined that CPC-EV migration and angiogenesis responses correlated with patient age. Here, we recapitulate our previous results demonstrating that CPC derived from younger patients exhibit greater paracrine pro-migratory behavior (Agarwal et al., 2016) (**Figure** 4-2 **d, Supplemental Figure** 4-1). Unexpectantly though, we determined that age positively correlated with angiogenesis. Previous results from our lab suggested that neonate-derived CPCs and CPC-EVs promoted angiogenesis to a greater extent than their child counterparts (Agarwal et al., 2017; Agarwal et al., 2016). However, these experiments were conducted using pooled samples, including the 3rd most angiogenic patient, #903. Given that the results presented here show large patient-to-patient variation, perhaps with repetitive cell passaging, specific patient lines may dominate pooled samples. Additionally, the CPC-EVs derived from CHILD samples (white-striped bars, **Figure** 4-2 **b**) are the least angiogenic and may be skewing these results. This observation may be due to technical variability from CPCs isolated at different locations: CHILD cells were initially isolated at University of Miami, whereas the other CPCs were isolated and expanded by our group. Overall, the purpose of this

study was to understand CPC variability which may be driving large variation in patient outcomes. To improve autologous cell therapy outcomes, it will be important to understand functional effects at the individual patient level, rather than exploring general effects from pooled samples.

A major issue with RNA sequencing experiments is the p>>n: there are far greater features, or RNAs, measured than samples. This issue makes conventional linear regression ill-suited to handle these problems. Fortunately, the multicollinearity, inherent to data sets this size, allows for the implementation of machine learning methods to reduce data set dimension and complexity and summarize the data. Here, we used WGCNA, an unsupervised learning approach, wherein clusters or modules of co-expressed EV-RNAs were determined, and then these modules were correlated to *in vitro* outcomes. WGCNA applied to our full data set (training and testing data) identified 33 modules of co-expressed RNAs (**Figure** 4-3 **a**). Eight modules were of particular interest, as they correlated to various outcomes. Interestingly, although the miRNA and total RNA data sets were centered and scaled, the 8 modules of interest were primarily comprised of either long RNAs (total RNA) or miRNAs (**Figure** 4-3 **b**). Additionally, the miRNA modules positively correlated with "poor" outcomes (fibrosis and inflammation) and negatively correlated with "good" outcomes (angiogenesis and MSC migration). This allowed us to directly translate the pathway analysis of these miRNA gene targets. Simply put, these miRNA modules target genes involved in cell migration, differentiation, adhesion, as well as VEGF signaling and immune processes. On the other hand, the RNA modules positively correlated with angiogenesis outcomes and negatively correlated with fibrosis responses. These RNAs were enriched in cell morphogenesis and adhesion processes.

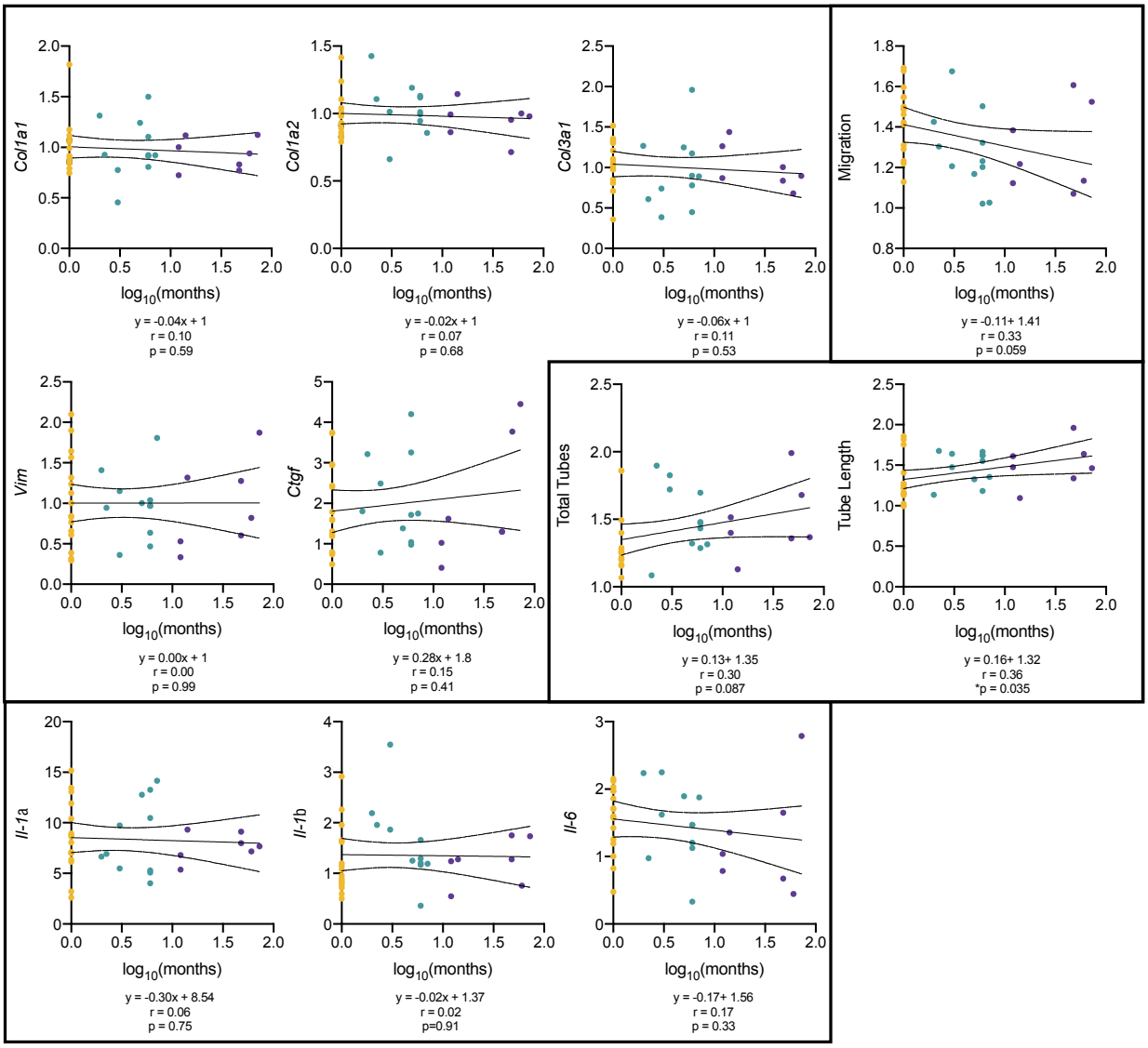
Notably, module 4 included well-studied miRNAs: let-7a/b/d/e/f/g/i, miR-99/100, -23/24/25 families, as well as miR-19b, -20a, -21, -30, -125, -146a, -320a. The let-7 family was the second miRNA discovered in C. elegans and it is widely studied in the context of cardiovascular disease. Finally, miR-19b and -20a belong to miR-17/92 cluster, known to regulate development and promote cell proliferation (Mogilyansky and Rigoutsos, 2013). Let-7 members drive cardiomyocyte maturation, controlling myocyte metabolism, cell size, and force contractility (Kuppusamy et al., 2015), and let-7i has been shown to attenuate angiotensin-induced cardiac inflammation and fibrosis (Wang et al., 2015). Also included in module 4, miR-146a may be cardioprotective and plays an anti-inflammatory role, regulating toll-like receptors and nuclear factor-κB signaling (Paterson and Kriegel, 2017). The inclusion of immunomodulatory miRNAs module 4 may explain its strong positive correlation ($p<0.01$) to *Il-6* expression in our inflammation experiments.

To directly link our CPC-EV RNA sequencing to experimental outcomes, we constructed partial least squares regression (PLSR) models of angiogenesis, migration, inflammation, and fibrosis responses. We constructed models using the training data—previously isolated CPCs from CHD patients—and measured the predictability in our test set—CHILD trial samples. PLSR is a supervised learning method that performs both dimension reduction and regression. Here, the SIMPLS PLSR algorithm computed components in RNAseq space that maximized the variance explained in the experimental outcomes space. Given that our outcomes were collinear within the experimental category (**Figure** 4-4 **a**), we sought to construct four PLSR models. Our original models described little variance (<35%) in the X, or RNAs, and produced poor prediction metrics for the cross-validation of our training set (**Figure** 4-4 **b**). Therefore, we computed VIP scores to determine the most important RNAs in the models and reduced our

regression models, from >11,000 RNAs to <300 RNAs. In doing so, we greatly improved upon our training data metrics: explained variance and RMSE/R^2 measurements. Interestingly, only a handful RNAs were deemed important across multiple models (**Figure** 4-4 **d**). This may indicate that different EV RNAs are driving different mechanisms. Further, pathway analysis revealed the enrichment of relevant GO parent pathways among the VIP RNAs. Importantly, PLSR has a hyperparameter to tune: the number of components in the model. Selecting the number of components is a matter of balancing performance with complexity: the best, most robust models do not over nor underfit. Here, we identified the number of components (2-3) to include by using cross-validation and RMSE measurements.

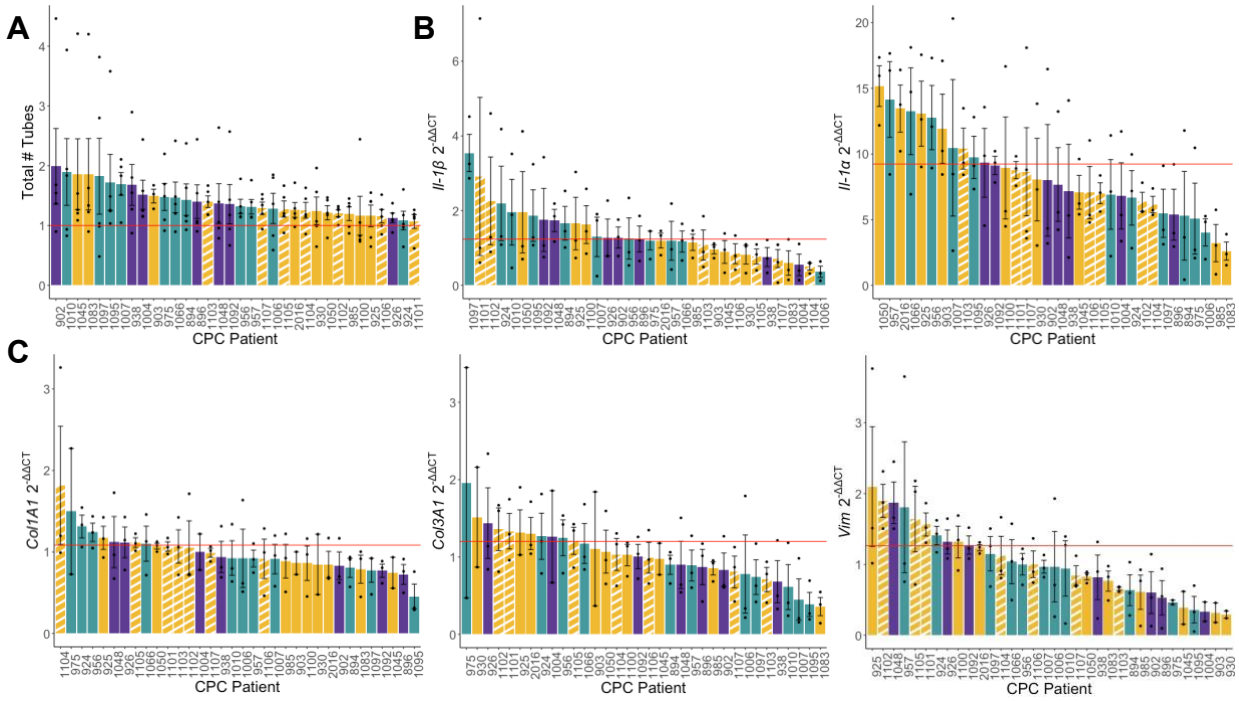
Overall, the study presented here investigates patient-to-patient variability of CPC-EVs and links CPC-EV RNA cargo to experimental outcomes. We identified differences in angiogenesis, migration, inflammation, and fibrosis responses from CPC-EV treated cells and determined the EV-RNAs which correlate and covary with responses using WGCNA and PLSR, respectively. Ultimately, we aim to connect these results to our phase I, patient-matched, results. We have ranked the *in vitro* performance of the CPC-EVs from patients included in the CHILD trial and when data is available, we will correlate these rankings with clinical rankings. These findings will help inform clinicians of patient outcomes, thereby potentially minimizing poor outcomes.

4.5 Supplemental Information



Supplemental Figure 4-1. Linear regression of patient age and experimental outcome.

Linear regression was performed for each experimental outcome. Patient age is quantified by $\log_{10}$(months). Correlation coefficients (r) and p-values are displayed. Regression models with $p<0.1$ are highlighted yellow.



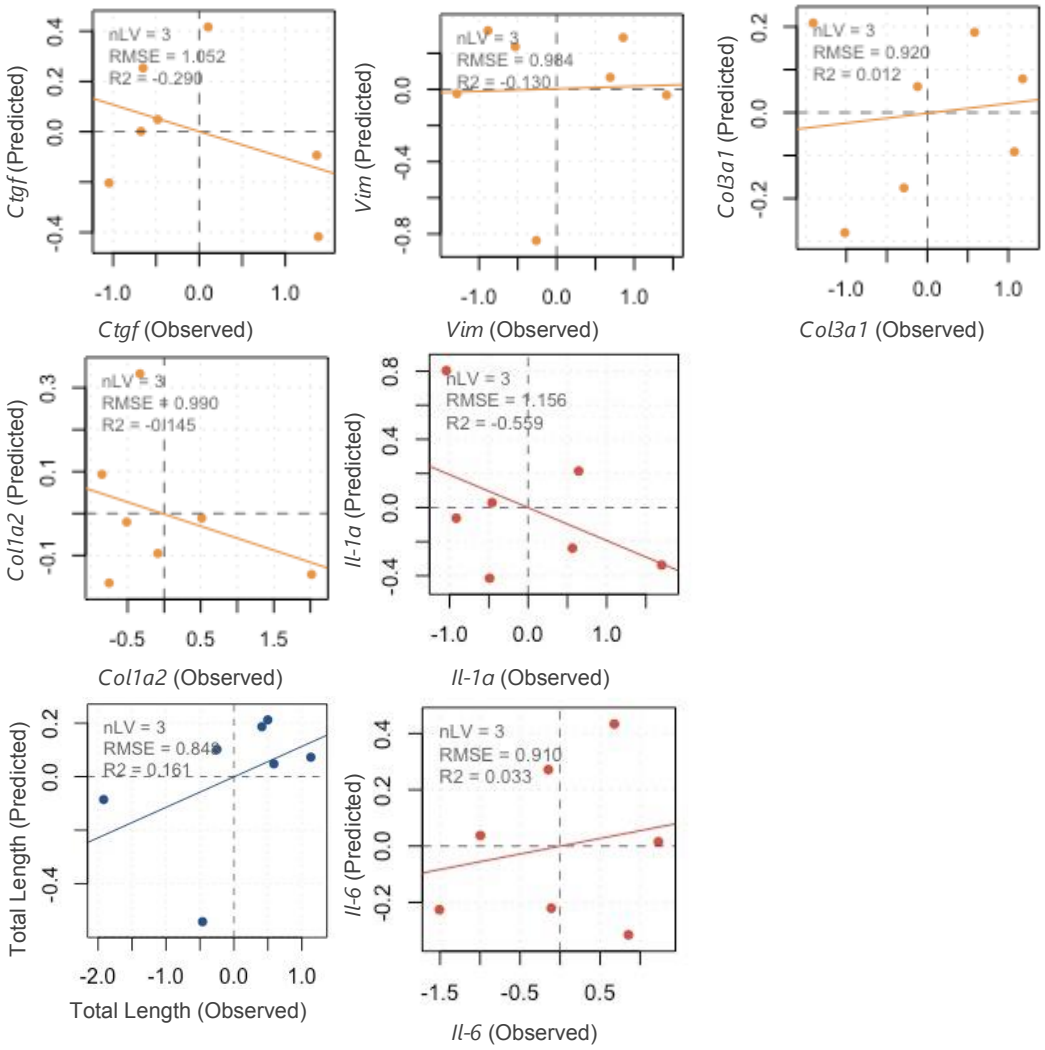
Supplemental Figure 4-2. Additional experimental outcomes.

A Quantification of total number of tubes of CECs grown on Matrigel, treated with CPC-EVs.

B *Il-1β* and *Il-1a* expression in CECs pre-treated with CPC-EVs and stimulated with TNF-α.

C *Col1a1*, *Col3a1*, and *Vim* expression in CFs pre-treated with CPC-EVs and stimulated with TGF-β.

CEC: cardiac endothelial cell; CPC: c-kit+ cardiac progenitor cell; EV: extracellular vesicle; CF: cardiac fibroblast.



Supplemental Figure 4-3. Observed vs. predicted CHILD plots.

In vitro predictions for CHILD CPC-EVs. Regression models built from previously collected CPC-EVs RNA-seq and experimental data. Model prediction performance determined by observed vs. predicted plots, RMSE, and R^2 metrics.

nLV: number of components; RMSE: root mean square error; CPC: c-kit+ cardiac progenitor cell; EV: extracellular vesicle.

Chapter 5: Discussion

5.1 Summary of results

This dissertation research explores the two main sources of patient variability of CPC-therapy: CPCs and CPC-EVs. Previous research of CPCs, and other stem/progenitor cell types, has shown that cells function in an age-dependent manner with cells losing therapeutic capacity with age (Agarwal et al., 2016; Efimenko et al., 2014; Fan et al., 2010). Chapter 2 investigated differences between neonate- and child-derived CPCs at the single cell level. In this work, we identified pro-fibrotic and pro-inflammatory cell clusters, enriched in child CPCs, which may be driving less reparative effects. Interestingly, in the pro-fibrotic cluster, we identified upregulated, well-studied lncRNAs which have recently been linked to adverse cardiac outcomes, like cardiac fibrosis. Additionally, in chapter 3 we investigated differential expression of competitive endogenous RNA networks between neonate- and child-derived CPCs from bulk RNAseq. We determined that child CPCs were enriched in non-coding RNAs overall, and the resulting differentially expressed, age-dependent RNA network was enriched in pathways related to blood vessel development, positive regulation of cell cycle, and regulation of Wnt signaling.

Next, in chapter 3, we explored bulk RNAseq differences between CPCs and patient-matched CPC-EVs using the differential expression for repeated measures method. Most obvious, we determined that CPC-EVs expressed more species of miRNAs and less long RNA (total RNAseq). As expected, CPCs were enriched in RNAs related to extracellular matrix organization and immune response, whereas CPC-EVs were enriched in RNAs related to cardiac development and cell signaling. Several groups have reported differences between parent cell and EV RNA content, as well as enrichment of pro-reparative cargo from stem or progenitor cell-derived EVs. Therefore, we used data mining to investigate differences between miRNAs

enriched in CPC-EVs vs. EVs derived from other cell types. We identified two classes of EV-enriched miRNAs: generic miRNAs that are enriched in other cell type-EVs (with some previously shown to regulate EV biogenesis), and potentially CPC-specific miRNAs which may be driving therapeutic success.

Finally, in chapter 4, we sought to link CPC-EV RNAseq to cardiac-relevant experimental outcomes, for the purposes of informing our CHILD clinical trial. In this section, we used regression models to identify CPC-EV RNA cargo molecules which correlate and covary with angiogenesis, migration, inflammatory, and fibrotic responses in recipient cells. We predicted *in vitro* responses for the CHILD clinical samples and determined patient rank for each outcome. We aim to correlate these ranks with clinical data, once made available, to understand (1) which *in vitro* responses best model clinical responses and (2) which EV RNA signals may be responsible for clinical improvements.

5.2 Limitations and future directions

The ultimate goals of this work are to gain mechanistic insight and build an informative clinical tool for CPC-therapy. Here, we investigated patient-derived CPCs and CPC-EVs from CHD patients, previously collected, and the first 7 lead-in patients of the CHILD trial. In this section, I will address the limitations of this dissertation work, as well as future directions.

First, the CPCs were grown in 2D culture. In order to collect enough EVs for sequencing and experimental outcomes, we expanded each patients' CPCs to ~30 million cells in multiple T175 flasks. We attempted to minimize transcriptomic drift by keeping cultures less than 10 or 15 passages. However, the 2D culture model for this work may not accurately reflect relevant *in vivo* conditions. In particular, we have shown that 3D culture of CPCs results in different EV cargo than 2D CPC culture (Trac et al., 2019a) (**Figure** 5-1). Future work should involve

growing CPCs in more biologically relevant models or isolating circulating CPC-EVs after cell transplantation. Notably, the latter was achieved by Kaushal group using an *in vivo* allogenic cell therapy model; circulating human CPC-EVs were isolated from plasma of rats via MHC mismatch (Saha et al., 2019). Recently, our group was granted approval to isolate EVs from the serum of patients in the phase II CHILD trial. Future work will involve using or modifying published deconvolution tools to identify sources of EV (Li et al., 2020; Murillo et al., 2019; Shi et al., 2020).

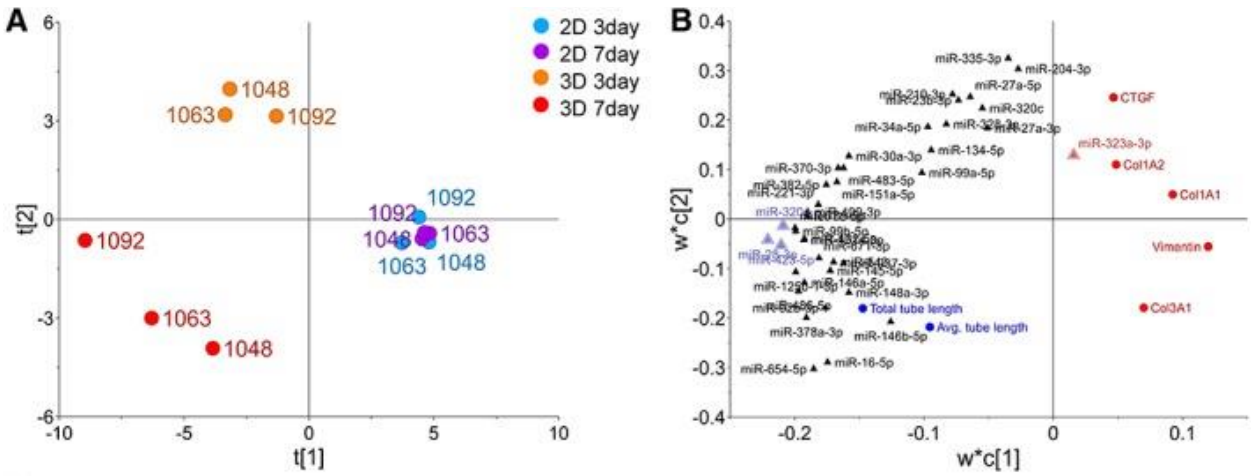


Figure 5-1. 2D vs. 3D CPC-EVs.

PLSR model constructed from CPC-EVs collected from CPCs cultured for 3 or 7 days, in 2D or 3D culture. MiRNAseq of CPC-EVs was performed and related to tube formation and fibrotic gene assays. **A** scores and **B** loadings plots of first two components from child-derived CPC-EV samples. Plots taken from Trac et al. (2019a). PLSR: partial least squares regression; CPC: c-kit+ cardiac progenitor cell

Second, the experimental section in chapter 4 was limited to inflammatory, fibrotic, angiogenic, and migratory outcomes. HLHS patients often suffer from adverse remodeling and eventual RV failure. Therefore, *in vitro* measurements of hypertrophy would likely improve our models and their correlation to clinical outcomes. Future work should consider investigating models of hypertrophy using induced pluripotent stem cell-derived cardiomyocytes or rat ventricular myocytes. An experimental approach for this may involve pre-treating cardiomyocytes with CPC-EVs, induce hypertrophy with angiotensin II or endothelin-1, and

measuring hypertrophy with wheat-germ agglutinin staining or expression of hypertrophy markers NPPA, NPPB, ACTA1 with PCR.

Third, our experiments were limited to one dose of CPC-EVs (20 μg/mL protein). This dose has been used previously by our group and elicits responses in angiogenesis and fibrosis assays (Trac et al., 2019a). However, a better approach would involve directly quantifying EVs. We have used NanoSight particle tracking analysis to quantify EV concentration, and this method for measuring EV dose is considered the gold standard (Thery et al., 2018). While groups have shown that EV protein correlates with EV concentration, future work will need to consider doses normalized by particle concentration (Saha et al., 2019). Further, the work presented here did not include dose-response experiments given the restraints of cell culture; this approach would require >10 T175 flasks of CPCs and potentially introduce greater transcriptomic drift. Future work investigating artificial vesicles, EVs derived from bio-fluids, or even cancer cell-derived EVs should consider dose-response experiments.

Fourth, it should be noted that in chapter 4, some of the *in vitro* outcomes for CHILD trial samples were predicted from our PLSR models (**Figure** 4-4), while others were not. Particularly, angiogenesis and inflammatory responses were better predicted by the PLSR models than the fibrotic or migration responses. This may be explained by a biological phenomenon, or some technical artifact causing the CHILD/testing and previously collected/training data sets to be incongruous. For example, poor predictions of fibrotic and migratory responses may be a consequence of a limited sample size (training or testing set) or a batch effect—CHILD CPCs were isolated at the University of Miami Clinical Research Cell Manufacturing Program. PCA plots of training and testing set did not show major differences in the EV cargo RNAseq data (**Figure** 4-1). However, there appear to be differences between the

sample sets in migration (high in CHILD samples) and angiogenesis (low in CHILD samples) outcomes (**Figure** 4-2**)**. Notably, the CHILD cells were further expanded in our lab for EV collection after initial isolation at University of Miami, unlike the training set which was isolated and expanded by our group. Additionally, this observation may be due to sample origin. The training set is comprised of CPCs derived from various ages and CHD conditions, whereas the CHILD set contains only neonate-derived samples from HLHS patient. Perhaps, CHD samples behave similarly in the context of angiogenesis and inflammatory responses.

Next, I will discuss the importance of experimentally validating our CPC-EV models. Not shown in the previous chapters, our group has used confocal microscopy and flow cytometry to confirm the uptake of CPC-EVs in recipient fibroblasts, endothelial cells previously (**Figure** 5-2). Additionally, our group has used IVIS imaging to show CPC-EV retention after intramyocardial injection (**Figure** 5-3). Analyzing EV RNA cargo is only half the story; to move forward with EV-based therapeutic strategies, we will need to validate EV retention and uptake by cardiac recipient cells. Not shown in this dissertation, Sruti Bheri and I are investigating CPC-EV and MSC-EV mechanisms of uptake in cardiac fibroblast and endolthelial cells. We (1) analyzed the EV membrane contents with lipidomics and proteomics, (2) quantified mechanisms of uptake (e.g. clathrin, lipid raft, macropinocytosis) in recipient cells with small molecule inhibitors and flow cytometry, and (3) aim to construct a regression model to determine which membrane molecules covary with mechanisms of uptake. This work will aid in efforts to understand vesicle uptake and engineer membranes for artificial vesicles.

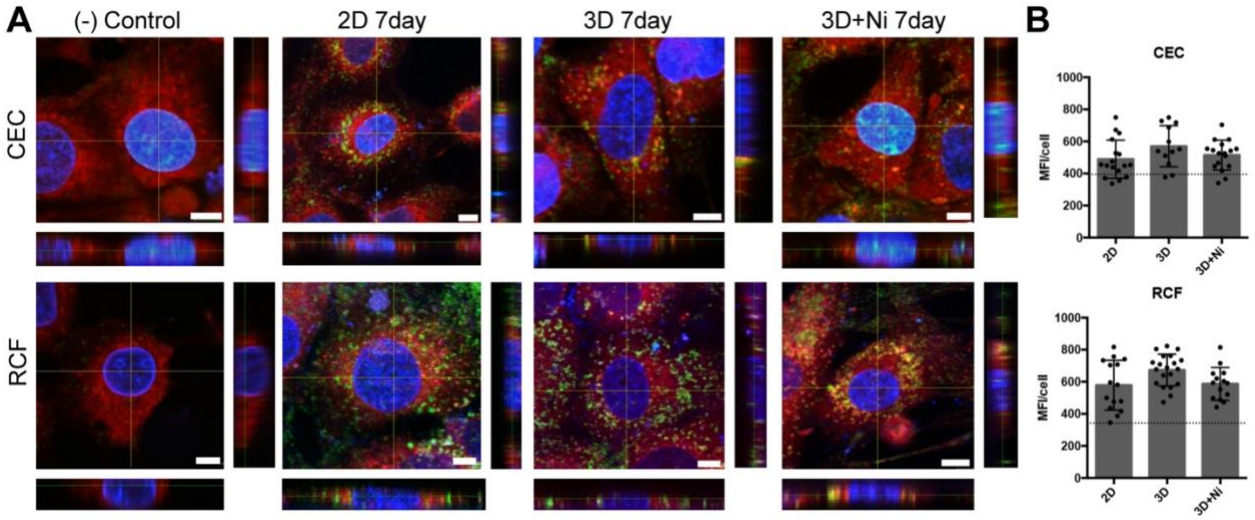


Figure 5-2. Uptake of CPC-EVs by cardiac fibroblasts endothelial cells.

CPC-EVs collected from 2D, 3D, and 3D + Notch1-shRNA (Ni) CPCs. CPC-EVs were stained with Calcein and dialyzed overnight to remove free dye. Cardiac fibroblasts (RCF) and endothelial cells (CECs) were treated with stained EVs (20 µg/mL protein) for 12 hours. Before imaging, cells were washed and stained with Hoescht and CMPTX. Dialyzed free calcein was used as a (-) control. **A** Confocal images from a central focal plane with orthogonal images on the bottom and right. **B** Quantification of EV uptake from individual cells. Dotted line represents (-) control. Adapted from Trac et al. (2019a).

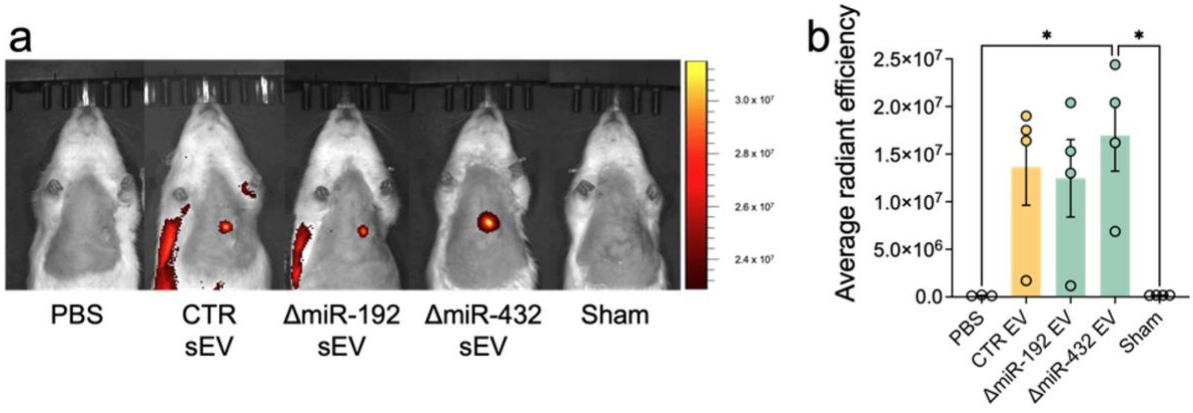


Figure 5-3. Retention of intramyocardial injection of CPC-EVs.

EVs derived from CPCs after miR-192 and miR-432 knockdown. **A** *In vivo* live imaging of CPC-EVs after Echo-guided intramyocardial injection into ischemic heart (ischemia-reperfusion injury). Fluorescent images of rats shown 7 days after injection of 5 µg EV per kg rat. **B** Quantification of average radiant efficiency. Figure adapted from manuscript currently under review: Park HJ, **Hoffman JR**, Brown ME, Bheri S, Davis ME.

The work presented in this dissertation identifies several EV RNAs of interest. However, it does not examine or validate these candidate RNAs. Our lab is currently working to validate the EV RNAs that our machine learning models deem important. In particular, work from Dr. Hyun-Ji Park is currently under review investigating the inhibition of miRNAs in EVs that two previous papers identified as deleterious. Agarwal et al. and Trac et al. demonstrated that CPC-

EV miR-192 and miR-432 covaried with reduced angiogenesis and enhanced fibrosis *in vitro*, as well as reduced ejection fraction in a rat RV failure model. To directly test the effects of these miRNAs, Park et al. delivered antagomiRs to parent cell CPCs and demonstrated knocked down expression of miR-192 and miR-432 in subsequent EVs. These modified CPC-EVs modulated immune response in an *in vivo* ischemia-reperfusion rat model (**Figure** 5-4). Further testing of candidate RNAs identified from the models presented in this dissertation need to be completed, either via parent cell or direct EV modification.

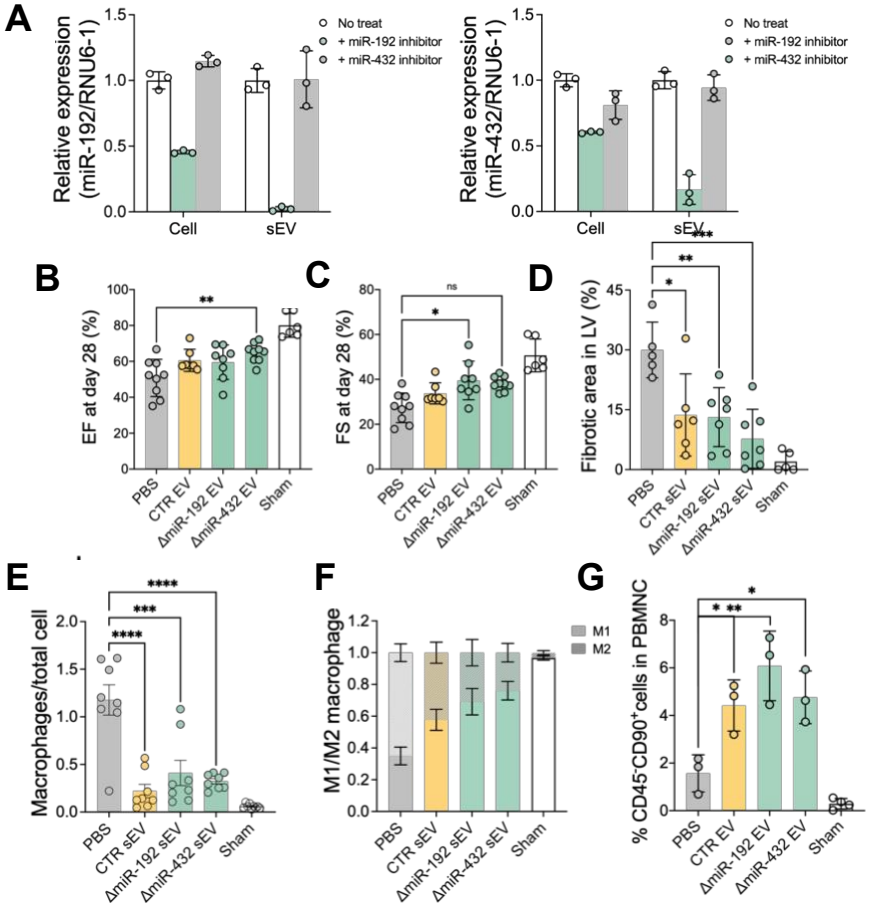


Figure 5-4. CPC-EV miR-192 and -432 knockdown.

Parent CPCs treated with miR-192 and -432 antagomiRs. **A** confirmation of miRNA knockdown in released CPC-EVs. **B** ejection fraction and **C** fractional shortening measurements 28 days after injury in a rat ischemia-reperfusion model. **D** fibrotic area of the left ventricle after CPC-EV treatment. Quantification of immunohistochemical staining of ischemic hearts show **E** total number of CD63+ macrophages and **F** ratio of iNOS-CD63+ cells (M2-type) to iNOS+CD63+ (M1-type). **G** Circulating CD45-CD90+ (MSC) populations in total PBMNCs from ischemia-reperfusion rats, quantified by flow cytometry. Adapted from manuscript currently under review: Park HJ, **Hoffman JR**, Brown ME, Bheri S, Davis ME.

Our group is also involved in the ELPIS trial, testing allogenic MSCs for treatment of HLHS. It will be interesting to compare important MSC-EV RNAs to CPC-EV RNAs. Perhaps there are consensus molecules driving cardiac cell therapy overall. There are studied miRNA families known to drive similar responses in recipient cells. For example, the miR-17/92 family are involved in development and upregulation of cell proliferation. A member of this family, miR-92a, is upregulated in high-potency CDCs and has been shown to potentiate the Wnt/β-catenin pathway (Ibrahim et al., 2019). If similar miRNAs are responsible for cardiac repair in various cell therapy models, these molecules may be leveraged for a cell-free therapeutic.

Current research is focused on EVs as a therapeutic candidate. They are non-immunogenic and have greater stability than the parent cells (Akers et al., 2016; Chen et al., 2016; Vandergriff et al., 2015). However, there are several drawbacks associated with EVs and future research could focus on scaling and creating a translatable artificial vesicle. The dependence of EV production on parent cells significantly limits the yield, as well as control over the cargo. As stated before, EV content is highly variable and depends on parent cell environment. Further, even though multiple groups have demonstrated that stem and progenitor cells yield reparative EVs, the cargo is limited and copy numbers for specific RNA molecules can be very low (Chevillet et al., 2014). Our group has demonstrated that the cargo of EVs may be optimized by using a modified thin-film hydration method to remove original cargo molecules and repackage vesicles with the RNA of one's choice (**Figure** 5-5). Finally, EVs have demonstrated higher tropism and reduced clearance, as compared to artificial vesicles (Sancho-Albero et al., 2019). Therefore, future research into artificial vesicles should focus on (1) improving vesicle yield, (2) enhancing the tropism/reduce clearance by tuning the lipid membrane, and (3) identifying and packaging pro-reparative nucleic acids.

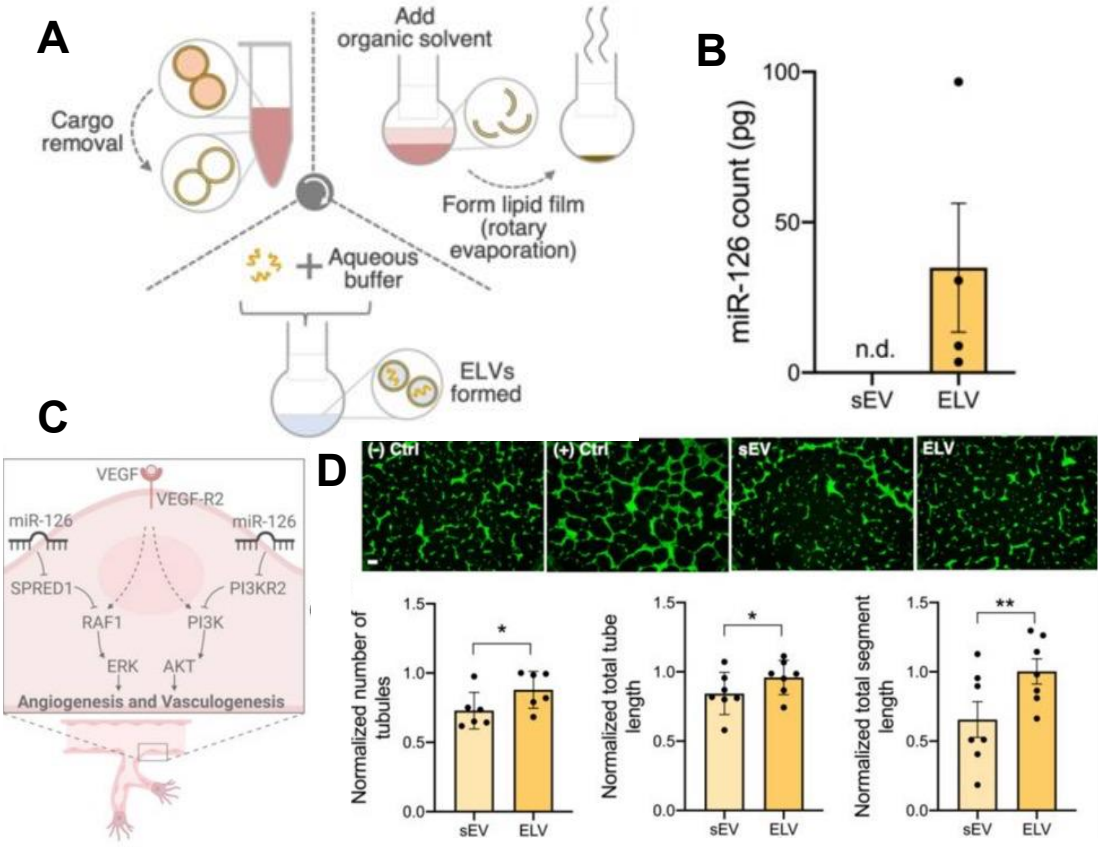


Figure 5-5. CPC-EV cargo optimization.

A CPC-EVs were engineered to (1) remove the cargo molecules, (2) form a lipid film, and (3) repackage miR-126 into new EV-like vesicles (ELVs). **B** ELVs were loaded with endothelial-specific miR-126. **C** Schematic of miR-126 mechanism of action for angiogenesis. **D** cardiac endothelial cells were treated with original EVs (sEV) or ELVs. ELVs induced angiogenesis. Adapted from Bheri et al. (2021).

Currently, EV studies are limited to bulk analyses. Higher resolution EV techniques are under development and may help understand heterogeneity at the single vesicle level. For example, there are size restrictions to conventional flow cytometers. However, groups are now investigating nanoparticle and extracellular vesicle flow cytometry and sorting (Pieragostino et al., 2019; Simonsen et al., 2019). I expect that groups are currently investigating single EV sequencing methods as well.

5.3 Conclusions

Overall, this work serves to better understand the variability underlying CPC-cell therapy. We investigated age-dependent differences in CPCs at the bulk and single-cell transcriptomic level,

we identified EV-specific miRNAs potentially responsible for reparative outcomes, and we linked EV RNA cargo to *in vitro* experimental outcomes to predict responses for clinical samples. I hope this dissertation work opens the door to future basic and clinical research to reduce patient outcome variability and improve reproducibility of cardiac cell therapy. By understanding the biological signals involved in repair we can improve patient outcomes at various levels: screening cells and EVs for potency, predicting patient improvements, or engineering better cells/vesicles.

References

Adamiak M, Cheng G, Bobis-Wozowicz S, Zhao L, Kedracka-Krok S, Samanta A, Karnas E, Xuan YT, Skupien-Rabian B, Chen X, Jankowska U, Girgis M, Sekula M, Davani A, Lasota S, Vincent RJ, Sarna M, Newell KL, Wang OL, Dudley N, Madeja Z, Dawn B and Zuba-Surma EK (2018) Induced Pluripotent Stem Cell (iPSC)-Derived Extracellular Vesicles Are Safer and More Effective for Cardiac Repair Than iPSCs. *Circ Res* **122**:296-309.

Agarwal U, George A, Bhutani S, Ghosh-Choudhary S, Maxwell JT, Brown ME, Mehta Y, Platt MO, Liang Y, Sahoo S and Davis ME (2017) Experimental, Systems, and Computational Approaches to Understanding the MicroRNA-Mediated Reparative Potential of Cardiac Progenitor Cell-Derived Exosomes From Pediatric Patients. *Circ Res* **120**:701-712.

Agarwal U, Smith AW, French KM, Boopathy AV, George A, Trac D, Brown ME, Shen M, Jiang R, Fernandez JD, Kogon BE, Kanter KR, Alsoufi B, Wagner MB, Platt MO and Davis ME (2016) Age-Dependent Effect of Pediatric Cardiac Progenitor Cells After Juvenile Heart Failure. *Stem Cells Transl Med* **5**:883-892.

Akers JC, Ramakrishnan V, Yang I, Hua W, Mao Y, Carter BS and Chen CC (2016) Optimizing preservation of extracellular vesicular miRNAs derived from clinical cerebrospinal fluid. *Cancer Biomark* **17**:125-132.

Al-Nedawi K, Meehan B, Micallef J, Lhotak V, May L, Guha A and Rak J (2008) Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat Cell Biol* **10**:619-624.

Altmann K, Printz BF, Solowiejczky DE, Gersony WM, Quaegebeur J and Apfel HD (2000) Two-dimensional echocardiographic assessment of right ventricular function as a predictor of outcome in hypoplastic left heart syndrome. *Am J Cardiol* **86**:964-968.

Apitz C, Webb GD and Redington AN (2009) Tetralogy of Fallot. *Lancet* **374**:1462-1471.

Bao L, Meng Q, Li Y, Deng S, Yu Z, Liu Z, Zhang L and Fan H (2017) C-Kit Positive Cardiac Stem Cells and Bone Marrow-Derived Mesenchymal Stem Cells Synergistically Enhance Angiogenesis and Improve Cardiac Function After Myocardial Infarction in a Paracrine Manner. *J Card Fail* **23**:403-415.

Barile L, Milano G and Vassalli G (2017a) Beneficial effects of exosomes secreted by cardiac-derived progenitor cells and other cell types in myocardial ischemia. *Stem Cell Investig* **4**:93.

Barile L, Moccetti T, Marban E and Vassalli G (2017b) Roles of exosomes in cardioprotection. *Eur Heart J* **38**:1372-1379.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S and Soboleva A (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**:D991-995.

Bausch-Fluck D, Hofmann A, Bock T, Frei AP, Cerciello F, Jacobs A, Moest H, Omasits U, Gundry RL, Yoon C, Schiess R, Schmidt A, Mirkowska P, Hartlova A, Van Eyk JE, Bourquin JP, Aebersold R, Boheler KR, Zandstra P and Wollscheid B (2015) A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**:e0121314.

Bergmann O, Bhardwaj RD, Bernard S, Zdunek S, Barnabe-Heider F, Walsh S, Zupicich J, Alkass K, Buchholz BA, Druid H, Jovinge S and Frisen J (2009) Evidence for cardiomyocyte renewal in humans. *Science* **324**:98-102.

Bheri S, Hoffman JR, Park HJ and Davis ME (2020) Biomimetic nanovesicle design for cardiac tissue repair. *Nanomedicine (Lond)* **15**:1873-1896.

Bheri S, Kassouf BP, Park HJ, Hoffman JR and Davis ME (2021) Engineering Cardiac Small Extracellular Vesicle-Derived Vehicles with Thin-Film Hydration for Customized microRNA Loading. *J Cardiovasc Dev Dis* **8**.

Bittle GJ, Morales D, Deatrick KB, Parchment N, Saha P, Mishra R, Sharma S, Pietris N, Vasilenko A, Bor C, Ambastha C, Gunasekaran M, Li D and Kaushal S (2018) Stem Cell Therapy for Hypoplastic Left Heart Syndrome: Mechanism, Clinical Application, and Future Directions. *Circ Res* **123**:288-300.

Bolli R, Mitrani RD, Hare JM, Pepine CJ, Perin EC, Willerson JT, Traverse JH, Henry TD, Yang PC, Murphy MP, March KL, Schulman IH, Ikram S, Lee DP, O'Brien C, Lima JA, Ostovaneh MR, Ambale-Venkatesh B, Lewis G, Khan A, Bacallao K, Valasaki K, Longsomboon B, Gee AP, Richman S, Taylor DA, Lai D, Sayre SL, Bettencourt J, Vojvodic RW, Cohen ML, Simpson L, Aguilar D, Loghin C, Moye L, Ebert RF, Davis BR, Simari RD and Cardiovascular Cell Therapy Research N (2021) 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.

Bolli R and Tang XL (2022) The sad plight of cell therapy for heart failure: causes and consequences. *J Cardiovasc Aging* **2**.

Bouzeghrane F, Reinhardt DP, Reudelhuber TL and Thibault G (2005) Enhanced expression of fibrillin-1, a constituent of the myocardial extracellular matrix in fibrosis. *Am J Physiol Heart Circ Physiol* **289**:H982-991.

Buratto E, Ye XT and Konstantinov IE (2016) Simple congenital heart disease: a complex challenge for public health. *J Thorac Dis* **8**:2994-2996.

Burkhart HM, Qureshi MY, Rossano JW, Cantero Peral S, O'Leary PW, Hathcock M, Kremers W, Nelson TJ and Wanek HCCP (2019) Autologous stem cell therapy for hypoplastic left heart syndrome: Safety and feasibility of intraoperative intramyocardial injections. *J Thorac Cardiovasc Surg* **158**:1614-1623.

Camussi G, Deregibus MC and Tetta C (2010) Paracrine/endocrine mechanism of stem cells on kidney repair: role of microvesicle-mediated transfer of genetic information. *Curr Opin Nephrol Hypertens* **19**:7-12.

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, Trapnell C and Shendure J (2019) The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**:496-502.

Chen H, Xia W and Hou M (2020) LncRNA-NEAT1 from the competing endogenous RNA network promotes cardioprotective efficacy of mesenchymal stem cell-derived exosomes induced by macrophage migration inhibitory factor via the miR-142-3p/FOXO1 signaling pathway. *Stem Cell Res Ther* **11**:31.

Chen J, Huang ZP, Seok HY, Ding J, Kataoka M, Zhang Z, Hu X, Wang G, Lin Z, Wang S, Pu WT, Liao R and Wang DZ (2013a) mir-17-92 cluster is required for and sufficient to induce cardiomyocyte proliferation in postnatal and adult hearts. *Circ Res* **112**:1557-1566.

Chen KH, Chen CH, Wallace CG, Yuen CM, Kao GS, Chen YL, Shao PL, Chen YL, Chai HT, Lin KC, Liu CF, Chang HW, Lee MS and Yip HK (2016) Intravenous administration of xenogenic adipose-derived mesenchymal stem cells (ADMSC) and ADMSC-derived

exosomes markedly reduced brain infarct volume and preserved neurological function in rat after acute ischemic stroke. *Oncotarget* **7**:74537-74556.

Chen L, Wang Y, Pan Y, Zhang L, Shen C, Qin G, Ashraf M, Weintraub N, Ma G and Tang Y (2013b) Cardiac progenitor-derived exosomes protect ischemic myocardium from acute ischemia/reperfusion injury. *Biochem Biophys Res Commun* **431**:566-571.

Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, Cheng HH, Arroyo JD, Meredith EK, Gallichotte EN, Pogosova-Agadjanyan EL, Morrissey C, Stirewalt DL, Hladik F, Yu EY, Higano CS and Tewari M (2014) Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proc Natl Acad Sci U S A* **111**:14888-14893.

Chung IM and Rajakumar G (2016) Genetics of Congenital Heart Defects: The NKX2-5 Gene, a Key Player. *Genes (Basel)* **7**.

Conigliaro A, Fontana S, Raimondo S and Alessandro R (2017) Exosomes: Nanocarriers of Biological Messages. *Adv Exp Med Biol* **998**:23-43.

Costa Verdera H, Gitz-Francois JJ, Schiffelers RM and Vader P (2017) Cellular uptake of extracellular vesicles is mediated by clathrin-independent endocytosis and macropinocytosis. *J Control Release* **266**:100-108.

De Jong S (1993) SIMPLS: an alternative approach squares regression to partial least. *Chemom Intell Lab Syst* **18**:2-263.

de Soysa TY, Ranade SS, Okawa S, Ravichandran S, Huang Y, Salunga HT, Schricker A, Del Sol A, Gifford CA and Srivastava D (2019) Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature* **572**:120-124.

Deregibus MC, Cantaluppi V, Calogero R, Lo Iacono M, Tetta C, Biancone L, Bruno S, Bussolati B and Camussi G (2007) Endothelial progenitor cell derived microvesicles activate an angiogenic program in endothelial cells by a horizontal transfer of mRNA. *Blood* **110**:2440-2448.

Ding Y, Wang Y, Zhang W, Jia Q, Wang X, Li Y, Lv S and Zhang J (2020) Roles of Biomarkers in Myocardial Fibrosis. *Aging Dis* **11**:1157-1174.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15-21.

Efimenko A, Dzhoyashvili N, Kalinina N, Kochegura T, Akchurin R, Tkachuk V and Parfyonova Y (2014) Adipose-derived mesenchymal stromal cells from aged patients with coronary artery disease keep mesenchymal stromal cell properties but exhibit characteristics of aging and have impaired angiogenic potential. *Stem Cells Transl Med* **3**:32-41.

Emanueli C, Shearn AI, Angelini GD and Sahoo S (2015) Exosomes and exosomal miRNAs in cardiovascular protection and repair. *Vascul Pharmacol* **71**:24-30.

Emmrich S, Rasche M, Schoning J, Reimer C, Keihani S, Maroz A, Xie Y, Li Z, Schambach A, Reinhardt D and Klusmann JH (2014) miR-99a/100~125b tricistrons regulate hematopoietic stem and progenitor cell homeostasis by shifting the balance between TGFbeta and Wnt signaling. *Genes Dev* **28**:858-874.

Everitt MD, Boyle GJ, Schechtman KB, Zheng J, Bullock EA, Kaza AK, Dipchand AI, Naftel DC, Kirklin JK, Canter CE and Pediatric Heart Transplant Study I (2012) Early survival after heart transplant in young infants is lowest after failed single-ventricle palliation: a multi-institutional study. *J Heart Lung Transplant* **31**:509-516.

Fan M, Chen W, Liu W, Du GQ, Jiang SL, Tian WC, Sun L, Li RK and Tian H (2010) The effect of age on the efficacy of human mesenchymal stem cell transplantation after a myocardial infarction. *Rejuvenation Res* **13**:429-438.

Feinstein JA, Benson DW, Dubin AM, Cohen MS, Maxey DM, Mahle WT, Pahl E, Villafane J, Bhatt AB, Peng LF, Johnson BA, Marsden AL, Daniels CJ, Rudd NA, Caldarone CA, Mussatto KA, Morales DL, Ivy DD, Gaynor JW, Tweddell JS, Deal BJ, Furck AK, Rosenthal GL, Ohye RG, Ghanayem NS, Cheatham JP, Tworetzky W and Martin GR (2012) Hypoplastic left heart syndrome: current considerations and expectations. *J Am Coll Cardiol* **59**:S1-42.

Frank S, Aguirre A, Hescheler J and Kurian L (2016) A lncRNA Perspective into (Re)Building the Heart. *Front Cell Dev Biol* **4**:128.

French KM and Davis ME (2014) Isolation and expansion of c-kit-positive cardiac progenitor cells by magnetic cell sorting. *Methods Mol Biol* **1181**:39-50.

Fuentes TI, Appleby N, Tsay E, Martinez JJ, Bailey L, Hasaniya N and Kearns-Jonker M (2013) Human neonatal cardiovascular progenitors: unlocking the secret to regenerative ability. *PLoS One* **8**:e77464.

Gan L, Xie D, Liu J, Bond Lau W, Christopher TA, Lopez B, Zhang L, Gao E, Koch W, Ma XL and Wang Y (2020) Small Extracellular Microvesicles Mediated Pathological Communications Between Dysfunctional Adipocytes and Cardiomyocytes as a Novel Mechanism Exacerbating Ischemia/Reperfusion Injury in Diabetic Mice. *Circulation* **141**:968-983.

Gastpar R, Gehrmann M, Bausero MA, Asea A, Gross C, Schroeder JA and Multhoff G (2005) Heat shock protein 70 surface-positive tumor exosomes stimulate migratory and cytolytic activity of natural killer cells. *Cancer Res* **65**:5238-5247.

Ge Z, Yin C, Li Y, Tian D, Xiang Y, Li Q, Tang Y and Zhang Y (2022) Long noncoding RNA NEAT1 promotes cardiac fibrosis in heart failure through increased recruitment of EZH2 to the Smad7 promoter region. *J Transl Med* **20**:7.

Gray WD, French KM, Ghosh-Choudhary S, Maxwell JT, Brown ME, Platt MO, Searles CD and Davis ME (2015) Identification of therapeutic covariant microRNA clusters in hypoxia-treated cardiac progenitor cell exosomes using systems biology. *Circ Res* **116**:255-263.

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

Haghverdi L, Lun ATL, Morgan MD and Marioni JC (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* **36**:421-427.

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P and Satija R (2021) Integrated analysis of multimodal single-cell data. *Cell* **184**:3573-3587 e3529.

Haraszti RA, Didiot MC, Sapp E, Leszyk J, Shaffer SA, Rockwell HE, Gao F, Narain NR, DiFiglia M, Kiebish MA, Aronin N and Khvorova A (2016) High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. *J Extracell Vesicles* **5**:32570.

Haraszti RA, Miller R, Dubuke ML, Rockwell HE, Coles AH, Sapp E, Didiot MC, Echeverria D, Stoppato M, Sere YY, Leszyk J, Alterman JF, Godinho B, Hassler MR, McDaniel J,

Narain NR, Wollacott R, Wang Y, Shaffer SA, Kiebish MA, DiFiglia M, Aronin N and Khvorova A (2019) Serum Deprivation of Mesenchymal Stem Cells Improves Exosome Activity and Alters Lipid and Protein Composition. *iScience* **16**:230-241.

Hare JM, Fishman JE, Gerstenblith G, DiFede Velazquez DL, Zambrano JP, Suncion VY, Tracy M, Ghersin E, Johnston PV, Brinker JA, Breton E, Davis-Sproul J, Schulman IH, Byrnes J, Mendizabal AM, Lowery MH, Rouy D, Altman P, Wong Po Foo C, Ruiz P, Amador A, Da Silva J, McNiece IK, Heldman AW, George R and Lardo A (2012) Comparison of allogeneic vs autologous bone marrow-derived mesenchymal stem cells delivered by transendocardial injection in patients with ischemic cardiomyopathy: the POSEIDON randomized trial. *JAMA* **308**:2369-2379.

Hare JM, Traverse JH, Henry TD, Dib N, Strumpf RK, Schulman SP, Gerstenblith G, DeMaria AN, Denktas AE, Gammon RS, Hermiller JB, Jr., Reisman MA, Schaer GL and Sherman W (2009) A randomized, double-blind, placebo-controlled, dose-escalation study of intravenous adult human mesenchymal stem cells (prochymal) after acute myocardial infarction. *J Am Coll Cardiol* **54**:2277-2286.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R and Gene Ontology C (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**:D258-261.

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ and Hammond SM (2005) A microRNA polycistron as a potential human oncogene. *Nature* **435**:828-833.

Hemler ME (2003) Tetraspanin proteins mediate cellular penetration, invasion, and fusion events and define a novel type of membrane microdomain. *Annu Rev Cell Dev Biol* **19**:397-422.

Hernandez I, Baio JM, Tsay E, Martinez AF, Fuentes TI, Bailey LL, Hasaniya NW and Kearns-Jonker M (2018) Short-term hypoxia improves early cardiac progenitor cell function in vitro. *Am J Stem Cells* **7**:1-17.

Hoffman GE and Roussos P (2021) Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**:192-201.

Hoffman GE and Schadt EE (2016) variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**:483.

Hoffman JR, Park HJ, Bheri S, Jayaraman AR and Davis ME (2022) Comparative computational RNA analysis of cardiac-derived progenitor cells and their extracellular vesicles. *Genomics* **114**:110349.

Hristov M, Erl W, Linder S and Weber PC (2004) Apoptotic bodies from endothelial cells enhance the number and initiate the differentiation of human endothelial progenitor cells in vitro. *Blood* **104**:2761-2766.

Hsu DT and Lamour JM (2015) Changing indications for pediatric heart transplantation: complex congenital heart disease. *Circulation* **131**:91-99.

Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH, Wu MC, Huang CY, Tsou AP and Huang HD (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* **39**:D163-169.

Huang S, Zhang L, Song J, Wang Z, Huang X, Guo Z, Chen F and Zhao X (2019) Long noncoding RNA MALAT1 mediates cardiac fibrosis in experimental postinfarct myocardium mice model. *J Cell Physiol* **234**:2997-3006.

Huijbregts RP, Topalof L and Bankaitis VA (2000) Lipid metabolism and regulation of membrane trafficking. *Traffic* **1**:195-202.

Ibrahim AG, Cheng K and Marban E (2014) Exosomes as critical agents of cardiac regeneration triggered by cell therapy. *Stem Cell Reports* **2**:606-619.

Ibrahim AGE, Li C, Rogers R, Fournier M, Li L, Vaturi SD, Antes T, Sanchez L, Akhmerov A, Moseley JJ, Tobin B, Rodriguez-Borlado L, Smith RR, Marban L and Marban E (2019) Augmenting canonical Wnt signalling in therapeutically inert cells converts them into therapeutically potent exosome factories. *Nat Biomed Eng* **3**:695-705.

Ikonen E (2001) Roles of lipid rafts in membrane transport. *Curr Opin Cell Biol* **13**:470-477.

Ishigami S, Ohtsuki S, Eitoku T, Ousaka D, Kondo M, Kurita Y, Hirai K, Fukushima Y, Baba K, Goto T, Horio N, Kobayashi J, Kuroko Y, Kotani Y, Arai S, Iwasaki T, Sato S, Kasahara S, Sano S and Oh H (2017) Intracoronary Cardiac Progenitor Cells in Single Ventricle Physiology: The PERSEUS (Cardiac Progenitor Cell Infusion to Treat Univentricular Heart Disease) Randomized Phase 2 Trial. *Circ Res* **120**:1162-1173.

Ishigami S, Ohtsuki S, Tarui S, Ousaka D, Eitoku T, Kondo M, Okuyama M, Kobayashi J, Baba K, Arai S, Kawabata T, Yoshizumi K, Tateishi A, Kuroko Y, Iwasaki T, Sato S, Kasahara S, Sano S and Oh H (2015) Intracoronary autologous cardiac progenitor cell transfer in patients with hypoplastic left heart syndrome: the TICAP prospective phase 1 controlled trial. *Circ Res* **116**:653-664.

Jackson CE, Scruggs BS, Schaffer JE and Hanson PI (2017) Effects of Inhibiting VPS4 Support a General Role for ESCRTs in Extracellular Vesicle Biogenesis. *Biophys J* **113**:1342-1352.

Jaiswal JK, Andrews NW and Simon SM (2002) Membrane proximal lysosomes are the major vesicles responsible for calcium-dependent exocytosis in nonsecretory cells. *J Cell Biol* **159**:625-635.

Jechlinger M, Sommer A, Moriggl R, Seither P, Kraut N, Capodiecci P, Donovan M, Cordon-Cardo C, Beug H and Grunert S (2006) Autocrine PDGFR signaling promotes mammary cancer metastasis. *J Clin Invest* **116**:1561-1570.

Jeggari A, Marks DS and Larsson E (2012) miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* **28**:2062-2063.

Jeppesen DK, Fenix AM, Franklin JL, Higginbotham JN, Zhang Q, Zimmerman LJ, Liebler DC, Ping J, Liu Q, Evans R, Fissell WH, Patton JG, Rome LH, Burnette DT and Coffey RJ (2019) Reassessment of Exosome Composition. *Cell* **177**:428-445 e418.

Johnstone RM, Adam M, Hammond JR, Orr L and Turbide C (1987) Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J Biol Chem* **262**:9412-9420.

Jumabay M, Moon JH, Yeerna H and Bostrom KI (2015) Effect of Diabetes Mellitus on Adipocyte-Derived Stem Cells in Rat. *J Cell Physiol* **230**:2821-2828.

Kahounova Z, Kurfurstova D, Bouchal J, Kharaishvili G, Navratil J, Remsik J, Simeckova S, Student V, Kozubik A and Soucek K (2018) 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. *Cytometry A* **93**:941-951.

Kanlikilicer P, Rashed MH, Bayraktar R, Mitra R, Ivan C, Aslan B, Zhang X, Filant J, Silva AM, Rodriguez-Aguayo C, Bayraktar E, Pichler M, Ozpolat B, Calin GA, Sood AK and Lopez-Berestein G (2016) Ubiquitous Release of Exosomal Tumor Suppressor miR-6126 from Ovarian Cancer Cells. *Cancer Res* **76**:7194-7207.

Karantalis V, Schulman IH, Balkan W and Hare JM (2015) Allogeneic cell therapy: a new paradigm in therapeutics. *Circ Res* **116**:12-15.

Kaushal S, Hare JM, Shah AM, Pietris NP, Bettencourt JL, Piller LB, Khan A, Snyder A, Boyd RM, Abdullah M, Mishra R, Sharma S, Slesnick TC, Si MS, Chai PJ, Davis BR, Lai D, Davis ME and Mahle WT (2022) Autologous Cardiac Stem Cell Injection in Patients with Hypoplastic Left Heart Syndrome (CHILD Study). *Pediatr Cardiol*.

Kaushal S, Wehman B, Pietris N, Naughton C, Bentzen SM, Bigham G, Mishra R, Sharma S, Vricella L, Everett AD, Deatrick KB, Huang S, Mehta H, Ravekes WA, Hibino N, Difede DL, Khan A and Hare JM (2017) Study design and rationale for ELPIS: A phase I/IIb randomized pilot study of allogeneic human mesenchymal stem cell injection in patients with hypoplastic left heart syndrome. *Am Heart J* **192**:48-56.

Kesidou D, da Costa Martins PA, de Windt LJ, Brittan M, Beqqali A and Baker AH (2020) Extracellular Vesicle miRNAs in the Promotion of Cardiac Neovascularisation. *Front Physiol* **11**:579892.

Khairy P, Ionescu-Ittu R, Mackie AS, Abrahamowicz M, Pilote L and Marelli AJ (2010) Changing mortality in congenital heart disease. *J Am Coll Cardiol* **56**:1149-1157.

Khong SML, Lee M, Kosaric N, Khong DM, Dong Y, Hopfner U, Aitzetmuller MM, Duscher D, Schafer R and Gurtner GC (2019) Single-Cell Transcriptomics of Human Mesenchymal Stem Cells Reveal Age-Related Cellular Subpopulation Depletion and Impaired Regenerative Function. *Stem Cells* **37**:240-246.

Kim H, Han JW, Lee JY, Choi YJ, Sohn YD, Song M and Yoon YS (2015) Diabetic Mesenchymal Stem Cells Are Ineffective for Improving Limb Ischemia Due to Their Impaired Angiogenic Capability. *Cell Transplant* **24**:1571-1584.

Kishore R and Khan M (2016) More Than Tiny Sacks: Stem Cell Exosomes as Cell-Free Modality for Cardiac Repair. *Circ Res* **118**:330-343.

Kochanek KD, Xu J and Arias E (2020) Mortality in the United States, 2019. *NCHS Data Brief, no 395 Hyattsville, MD National Center for Health Statistics*.

Kolde R (2012) Pheatmap: pretty heatmaps, in *R package*.

Konstandin MH, Toko H, Gastelum GM, Quijada P, De La Torre A, Quintana M, Collins B, Din S, Avitabile D, Volkers M, Gude N, Fassler R and Sussman MA (2013) Fibronectin is essential for reparative cardiac progenitor cell response after myocardial infarction. *Circ Res* **113**:115-125.

Kuppusamy KT, Jones DC, Sperber H, Madan A, Fischer KA, Rodriguez ML, Pabon L, Zhu WZ, Tulloch NL, Yang X, Sniadecki NJ, Laflamme MA, Ruzzo WL, Murry CE and Ruohola-Baker H (2015) Let-7 family of microRNA is required for maturation and adult-like metabolism in stem cell-derived cardiomyocytes. *Proc Natl Acad Sci U S A* **112**:E2785-2794.

Lamichhane TN, Raiker RS and Jay SM (2015) Exogenous DNA Loading into Extracellular Vesicles via Electroporation is Size-Dependent and Enables Limited Gene Delivery. *Mol Pharm* **12**:3650-3657.

Lamour JM, Kanter KR, Naftel DC, Chrisant MR, Morrow WR, Clemson BS, Kirklin JK, Cardiac Transplant Registry D and Pediatric Heart Transplant S (2009) The effect of age, diagnosis, and previous surgery in children and adults undergoing heart transplantation for congenital heart disease. *J Am Coll Cardiol* **54**:160-165.

Lang JK, Young RF, Ashraf H and Canty JM, Jr. (2016) Inhibiting Extracellular Vesicle Release from Human Cardiosphere Derived Cells with Lentiviral Knockdown of nSMase2 Differentially Effects Proliferation and Apoptosis in Cardiomyocytes, Fibroblasts and Endothelial Cells In Vitro. *PLoS One* **11**:e0165926.

Langfelder P and Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559.

Lee C, Mitsialis SA, Aslam M, Vitali SH, Vergadi E, Konstantinou G, Sdrimas K, Fernandez-Gonzalez A and Kourembanas S (2012) Exosomes mediate the cytoprotective action of mesenchymal stromal cells on hypoxia-induced pulmonary hypertension. *Circulation* **126**:2601-2611.

Leek JT, Johnson WE, Parker HS, Jaffe AE and Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**:882-883.

Levine JH, Simonds EF, Bendall SC, Davis KL, Amir el AD, Tadmor MD, Litvin O, Fienberg HG, Jager A, Zunder ER, Finck R, Gedman AL, Radtke I, Downing JR, Pe'er D and Nolan GP (2015) Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**:184-197.

Li J, Zhao Y, Lu Y, Ritchie W, Grau G, Vadas MA and Gamble JR (2016a) The Poly-cistronic miR-23-27-24 Complexes Target Endothelial Cell Junctions: Differential Functional and Molecular Effects of miR-23a and miR-23b. *Mol Ther Nucleic Acids* **5**:e354.

Li S, Geng J, Xu X, Huang X, Leng D, Jiang D, Liang J, Wang C, Jiang D and Dai H (2016b) miR-130b-3p Modulates Epithelial-Mesenchymal Crosstalk in Lung Fibrosis by Targeting IGF-1. *PLoS One* **11**:e0150418.

Li SP, Lin ZX, Jiang XY and Yu XY (2018) Exosomal cargo-loading and synthetic exosome-mimics as potential therapeutic tools. *Acta Pharmacol Sin* **39**:542-551.

Li T, Ding ZL, Zheng YL and Wang W (2017) MiR-484 promotes non-small-cell lung cancer (NSCLC) progression through inhibiting Apaf-1 associated with the suppression of apoptosis. *Biomed Pharmacother* **96**:153-164.

Li Y, He X, Li Q, Lai H, Zhang H, Hu Z, Li Y and Huang S (2020) EV-origin: Enumerating the tissue-cellular origin of circulating extracellular vesicles using exLR profile. *Comput Struct Biotechnol J* **18**:2851-2859.

Liu Y, Li Q, Hosen MR, Zietzer A, Flender A, Levermann P, Schmitz T, Fruhwald D, Goody P, Nickenig G, Werner N and Jansen F (2019) Atherosclerotic Conditions Promote the Packaging of Functional MicroRNA-92a-3p Into Endothelial Microvesicles. *Circ Res* **124**:575-587.

Llorente A, Skotland T, Sylvanne T, Kauhanen D, Rog T, Orlowski A, Vattulainen I, Ekroos K and Sandvig K (2013) Molecular lipidomics of exosomes released by PC-3 prostate cancer cells. *Biochim Biophys Acta* **1831**:1302-1309.

Lu Y, Thomson JM, Wong HY, Hammond SM and Hogan BL (2007) Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Dev Biol* **310**:442-453.

Luther KM, Haar L, McGuinness M, Wang Y, Lynch Iv TL, Phan A, Song Y, Shen Z, Gardner G, Kuffel G, Ren X, Zilliox MJ and Jones WK (2018) Exosomal miR-21a-5p mediates cardioprotection by mesenchymal stem cells. *J Mol Cell Cardiol* **119**:125-137.

Ma Z, Li Y, Zhang Y, Chen J, Tan T and Fan Y (2020) A lncRNA-miRNA-mRNA network for human primed, naive and extended pluripotent stem cells. *PLoS One* **15**:e0234628.

Mai CT, Isenburg JL, Canfield MA, Meyer RE, Correa A, Alverson CJ, Lupo PJ, Riehle-Colarusso T, Cho SJ, Aggarwal D, Kirby RS and National Birth Defects Prevention N (2019) National population-based estimates for major birth defects, 2010-2014. *Birth Defects Res* **111**:1420-1435.

Malliaras K and Marban E (2011) Cardiac cell therapy: where we've been, where we are, and where we should be headed. *Br Med Bull* **98**:161-185.

Marban E (2018) A mechanistic roadmap for the clinical application of cardiac cell therapies. *Nat Biomed Eng* **2**:353-361.

McKenzie AJ, Hoshino D, Hong NH, Cha DJ, Franklin JL, Coffey RJ, Patton JG and Weaver AM (2016) KRAS-MEK Signaling Controls Ago2 Sorting into Exosomes. *Cell Rep* **15**:978-987.

Melo SA, Sugimoto H, O'Connell JT, Kato N, Villanueva A, Vidal A, Qiu L, Vitkin E, Perelman LT, Melo CA, Lucci A, Ivan C, Calin GA and Kalluri R (2014) Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. *Cancer Cell* **26**:707-721.

Miao Y, Tian L, Martin M, Paige SL, Galdos FX, Li J, Klein A, Zhang H, Ma N, Wei Y, Stewart M, Lee S, Moonen JR, Zhang B, Grossfeld P, Mital S, Chitayat D, Wu JC, Rabinovitch M, Nelson TJ, Nie S, Wu SM and Gu M (2020) Intrinsic Endocardial Defects Contribute to Hypoplastic Left Heart Syndrome. *Cell Stem Cell* **27**:574-589 e578.

Mishra R, Vijayan K, Colletti EJ, Harrington DA, Matthiesen TS, Simpson D, Goh SK, Walker BL, Almeida-Porada G, Wang D, Backer CL, Dudley SC, Jr., Wold LE and Kaushal S (2011) Characterization and functionality of cardiac progenitor cells in congenital heart patients. *Circulation* **123**:364-373.

Mittelbrunn M, Gutierrez-Vazquez C, Villarroya-Beltri C, Gonzalez S, Sanchez-Cabo F, Gonzalez MA, Bernad A and Sanchez-Madrid F (2011) Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nat Commun* **2**:282.

Mogilyansky E and Rigoutsos I (2013) The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ* **20**:1603-1614.

Mollova M, Bersell K, Walsh S, Savla J, Das LT, Park SY, Silberstein LE, Dos Remedios CG, Graham D, Colan S and Kuhn B (2013) Cardiomyocyte proliferation contributes to heart growth in young humans. *Proc Natl Acad Sci U S A* **110**:1446-1451.

Mullen M, Zhang A, Lui GK, Romfh AW, Rhee JW and Wu JC (2021) Race and Genetics in Congenital Heart Disease: Application of iPSCs, Omics, and Machine Learning Technologies. *Front Cardiovasc Med* **8**:635280.

Muntean I, Toganel R and Benedek T (2017) Genetics of Congenital Heart Disease: Past and Present. *Biochem Genet* **55**:105-123.

Murillo OD, Thistlethwaite W, Rozowsky J, Subramanian SL, Lucero R, Shah N, Jackson AR, Srinivasan S, Chung A, Laurent CD, Kitchen RR, Galeev T, Warrell J, Diao JA, Welsh JA, Hanspers K, Riutta A, Burgstaller-Muehlbacher S, Shah RV, Yeri A, Jenkins LM, Ahsen ME, Cordon-Cardo C, Dogra N, Gifford SM, Smith JT, Stolovitzky G, Tewari AK, Wunsch BH, Yadav KK, Danielson KM, Filant J, Moeller C, Nejad P, Paul A, Simonson B, Wong DK, Zhang X, Balaj L, Gandhi R, Sood AK, Alexander RP, Wang L, Wu C, Wong DTW, Galas DJ, Van Keuren-Jensen K, Patel T, Jones JC, Das S, Cheung KH, Pico AR, Su AI, Raffai RL, Laurent LC, Roth ME, Gerstein MB and Milosavljevic A (2019) exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell* **177**:463-477 e415.

Nguyen PK, Neofytou E, Rhee JW and Wu JC (2016) Potential Strategies to Address the Major Clinical Barriers Facing Stem Cell Regenerative Therapy for Cardiovascular Disease: A Review. *JAMA Cardiol* **1**:953-962.

Ong SG, Huber BC, Lee WH, Kodo K, Ebert AD, Ma Y, Nguyen PK, Diecke S, Chen WY and Wu JC (2015) Microfluidic Single-Cell Analysis of Transplanted Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes After Acute Myocardial Infarction. *Circulation* **132**:762-771.

Ono M, Kosaka N, Tominaga N, Yoshioka Y, Takeshita F, Takahashi RU, Yoshida M, Tsuda H, Tamura K and Ochiya T (2014) Exosomes from bone marrow mesenchymal stem cells contain a microRNA that promotes dormancy in metastatic breast cancer cells. *Sci Signal* **7**:ra63.

Oommen S, Yamada S, Cantero Peral S, Campbell KA, Bruinsma ES, Terzic A and Nelson TJ (2015) Human umbilical cord blood-derived mononuclear cells improve murine ventricular function upon intramyocardial delivery in right ventricular chronic pressure overload. *Stem Cell Res Ther* **6**:50.

Oster ME, Strickland MJ and Mahle WT (2011) Racial and ethnic disparities in post-operative mortality following congenital heart surgery. *J Pediatr* **159**:222-226.

Pan BT, Teng K, Wu C, Adam M and Johnstone RM (1985) Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. *J Cell Biol* **101**:942-948.

Paterson MR and Kriegel AJ (2017) MiR-146a/b: a family with shared seeds and different roots. *Physiol Genomics* **49**:243-252.

Pegtel DM, Cosmopoulos K, Thorley-Lawson DA, van Eijndhoven MA, Hopmans ES, Lindenberg JL, de Gruijl TD, Wurdinger T and Middeldorp JM (2010) Functional delivery of viral miRNAs via exosomes. *Proc Natl Acad Sci U S A* **107**:6328-6333.

Pieragostino D, Lanuti P, Cicalini I, Cufaro MC, Ciccocioppo F, Ronci M, Simeone P, Onofrj M, van der Pol E, Fontana A, Marchisio M and Del Boccio P (2019) Proteomics characterization of extracellular vesicles sorted by flow cytometry reveals a disease-specific molecular cross-talk from cerebrospinal fluid and tears in multiple sclerosis. *J Proteomics* **204**:103403.

Pierpont ME, Basson CT, Benson DW, Jr., Gelb BD, Giglia TM, Goldmuntz E, McGee G, Sable CA, Srivastava D, Webb CL and American Heart Association Congenital Cardiac Defects Committee CoCDitY (2007) Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* **115**:3015-3038.

Qi X, Zhang DH, Wu N, Xiao JH, Wang X and Ma W (2015) ceRNA in cancer: possible functions and clinical implications. *J Med Genet* **52**:710-718.

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA and Trapnell C (2017) Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**:979-982.

Quevedo HC, Hatzistergos KE, Oskouei BN, Feigenbaum GS, Rodriguez JE, Valdes D, Pattany PM, Zambrano JP, Hu Q, McNiece I, Heldman AW and Hare JM (2009) Allogeneic mesenchymal stem cells restore cardiac function in chronic ischemic cardiomyopathy via trilineage differentiating capacity. *Proc Natl Acad Sci U S A* **106**:14022-14027.

Qureshi MY, Cabalka AK, Khan SP, Hagler DJ, Haile DT, Cannon BC, Olson TM, Cantero-Peral S, Dietz AB, Radel DJ, Taggart NW, Kelle AM, Rodriguez V, Dearani JA, O'Leary PW and Wanek Program Regenerative Medicine P (2017) Cell-Based Therapy for Myocardial Dysfunction After Fontan Operation in Hypoplastic Left Heart Syndrome. *Mayo Clin Proc Innov Qual Outcomes* **1**:185-191.

Rieger AC, Myerburg RJ, Florea V, Tompkins BA, Natsumeda M, Premer C, Khan A, Schulman IH, Vidro-Casiano M, DiFede DL, Heldman AW, Mitrani R and Hare JM (2019) Genetic determinants of responsiveness to mesenchymal stem cell injections in non-ischemic dilated cardiomyopathy. *EBioMedicine* **48**:377-385.

Robinson MD, McCarthy DJ and Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139-140.

Rosen MR, Myerburg RJ, Francis DP, Cole GD and Marban E (2014) Translating stem cell research to cardiac disease therapies: pitfalls and prospects for improvement. *J Am Coll Cardiol* **64**:922-937.

Saha P, Sharma S, Korutla L, Datla SR, Shoja-Taheri F, Mishra R, Bigham GE, Sarkar M, Morales D, Bittle G, Gunasekaran M, Ambastha C, Arfat MY, Li D, Habertheuer A, Hu R, Platt MO, Yang P, Davis ME, Vallabhajosyula P and Kaushal S (2019) Circulating exosomes derived from transplanted progenitor cells aid the functional recovery of ischemic myocardium. *Sci Transl Med* **11**.

Sahoo S, Klychko E, Thorne T, Misener S, Schultz KM, Millay M, Ito A, Liu T, Kamide C, Agrawal H, Perlman H, Qin G, Kishore R and Losordo DW (2011) Exosomes from human CD34(+) stem cells mediate their proangiogenic paracrine activity. *Circ Res* **109**:724-728.

Sahoo S and Losordo DW (2014) Exosomes and cardiac repair after myocardial infarction. *Circ Res* **114**:333-344.

Salmena L, Poliseno L, Tay Y, Kats L and Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**:353-358.

Sancho-Albero M, Navascues N, Mendoza G, Sebastian V, Arruebo M, Martin-Duque P and Santamaria J (2019) Exosome origin determines cell targeting and the transfer of therapeutic nanoparticles towards target cells. *J Nanobiotechnology* **17**:16.

Saraf A, Book WM, Nelson TJ and Xu C (2019) Hypoplastic left heart syndrome: From bedside to bench and back. *J Mol Cell Cardiol* **135**:109-118.

Saraste A and Pulkki K (2000) Morphologic and biochemical hallmarks of apoptosis. *Cardiovasc Res* **45**:528-537.

Satija R, Farrell JA, Gennert D, Schier AF and Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**:495-502.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**:2498-2504.

Sharma S, Mishra R, Bigham GE, Wehman B, Khan MM, Xu H, Saha P, Goo YA, Datla SR, Chen L, Tulapurkar ME, Taylor BS, Yang P, Karathanasis S, Goodlett DR and Kaushal S (2017) A Deep Proteome Analysis Identifies the Complete Secretome as the Functional Unit of Human Cardiac Progenitor Cells. *Circ Res* **120**:816-834.

Shi A, Kasumova GG, Michaud WA, Cintolo-Gonzalez J, Diaz-Martinez M, Ohmura J, Mehta A, Chien I, Frederick DT, Cohen S, Plana D, Johnson D, Flaherty KT, Sullivan RJ, Kellis M and Boland GM (2020) Plasma-derived extracellular vesicle analysis and deconvolution enable prediction and tracking of melanoma checkpoint blockade outcome. *Sci Adv* **6**.

Shoja-Taheri F, George A, Agarwal U, Platt MO, Gibson G and Davis ME (2019) Using Statistical Modeling to Understand and Predict Pediatric Stem Cell Function. *Circ Genom Precis Med* **12**:e002403.

Siffel C, Riehle-Colarusso T, Oster ME and Correa A (2015) Survival of Children With Hypoplastic Left Heart Syndrome. *Pediatrics* **136**:e864-870.

Simonsen JB, Larsen JB, Hempel C, Eng N, Fossum A and Andresen TL (2019) Unique Calibrators Derived from Fluorescence-Activated Nanoparticle Sorting for Flow Cytometric Size Estimation of Artificial Vesicles: Possibilities and Limitations. *Cytometry A* **95**:917-924.

Simpson DL, Mishra R, Sharma S, Goh SK, Deshmukh S and Kaushal S (2012) A strong regenerative ability of cardiac stem cells derived from neonatal hearts. *Circulation* **126**:S46-53.

Stolzing A, Jones E, McGonagle D and Scutt A (2008) Age-related changes in human bone marrow-derived mesenchymal stem cells: consequences for cell therapies. *Mech Ageing Dev* **129**:163-173.

Stoorvogel W, Strous GJ, Geuze HJ, Oorschot V and Schwartz AL (1991) Late endosomes derive from early endosomes by maturation. *Cell* **65**:417-427.

Svensson KJ, Christianson HC, Wittrup A, Bourseau-Guilmain E, Lindqvist E, Svensson LM, Morgelin M and Belting M (2013) Exosome uptake depends on ERK1/2-heat shock protein 27 signaling and lipid Raft-mediated endocytosis negatively regulated by caveolin-1. *J Biol Chem* **288**:17713-17724.

Tadokoro H, Umezu T, Ohyashiki K, Hirano T and Ohyashiki JH (2013) Exosomes derived from hypoxic leukemia cells enhance tube formation in endothelial cells. *J Biol Chem* **288**:34343-34351.

Tang XL, Rokosh DG, Guo Y and Bolli R (2010) Cardiac progenitor cells and bone marrow-derived very small embryonic-like stem cells for cardiac repair after myocardial infarction. *Circ J* **74**:390-404.

Tang YL, Zhu W, Cheng M, Chen L, Zhang J, Sun T, Kishore R, Phillips MI, Losordo DW and Qin G (2009) Hypoxic preconditioning enhances the benefit of cardiac progenitor cell therapy for treatment of myocardial infarction by inducing CXCR4 expression. *Circ Res* **104**:1209-1216.

Temoche-Diaz MM, Shurtleff MJ, Nottingham RM, Yao J, Fadadu RP, Lambowitz AM and Schekman R (2019) Distinct mechanisms of microRNA sorting into cancer cell-derived extracellular vesicle subtypes. *Elife* **8**.

Teng Y, Ren Y, Hu X, Mu J, Samykutty A, Zhuang X, Deng Z, Kumar A, Zhang L, Merchant ML, Yan J, Miller DM and Zhang HG (2017) MVP-mediated exosomal sorting of miR-193a promotes colon cancer progression. *Nat Commun* **8**:14448.

Thery C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, Antoniou A, Arab T, Archer F, Atkin-Smith GK, Ayre DC, Bach JM, Bachurski D, Baharvand H, Balaj L, Baldacchino S, Bauer NN, Baxter AA, Bebawy M, Beckham C, Bedina Zavec A, Benmoussa A, Berardi AC, Bergese P, Bielska E, Blenkiron C, Bobis-Wozowicz S, Boilard E, Boireau W, Bongiovanni A, Borras FE, Bosch S, Boulanger CM, Breakefield X, Breglio AM, Brennan MA, Brigstock DR, Brisson A, Broekman ML, Bromberg JF, Bryl-Gorecka P, Buch S, Buck AH, Burger D, Busatto S, Buschmann D, Bussolati B, Buzas EI, Byrd JB, Camussi G, Carter DR, Caruso S, Chamley LW, Chang YT, Chen C, Chen S, Cheng L, Chin AR, Clayton A, Clerici SP, Cocks A, Cocucci E, Coffey RJ, Cordeiro-da-Silva A, Couch Y, Coumans FA, Coyle B, Crescitelli R, Criado MF, D'Souza-Schorey C, Das S, Datta Chaudhuri A, de Candia P, De Santana EF, De Wever O, Del Portillo HA, Demaret T, Deville S, Devitt A, Dhondt B, Di Vizio D, Dieterich LC, Dolo V, Dominguez Rubio AP, Dominici M, Dourado MR, Driedonks TA, Duarte FV, Duncan HM, Eichenberger RM, Ekstrom K, El Andaloussi S, Elie-Caille C, Erdbrugger U, Falcon-Perez JM, Fatima F, Fish JE, Flores-Bellver M, Forsonits A, Frelet-Barrand A, Fricke F, Fuhrmann G, Gabrielsson S, Gamez-Valero A, Gardiner C, Gartner K, Gaudin R, Gho YS, Giebel B, Gilbert C, Gimona M, Giusti I, Goberdhan DC, Gorgens A, Gorski SM, Greening DW, Gross JC, Gualerzi A, Gupta GN, Gustafson D, Handberg A, Haraszti RA, Harrison P, Hegyesi H, Hendrix A, Hill AF, Hochberg FH, Hoffmann KF, Holder B, Holthofer H, Hosseinkhani B, Hu G, Huang Y, Huber V, Hunt S, Ibrahim AG, Ikezu T, Inal JM, Isin M, Ivanova A, Jackson HK, Jacobsen S, Jay SM, Jayachandran M, Jenster G, Jiang L, Johnson SM, Jones JC, Jong A, Jovanovic-Talisman T, Jung S, Kalluri R, Kano SI, Kaur S, Kawamura Y, Keller ET, Khamari D, Khomyakova E, Khvorova A, Kierulf P, Kim KP, Kislinger T, Klingeborn M, Klinke DJ, 2nd, Kornek M, Kosanovic MM, Kovacs AF, Kramer-Albers EM, Krasemann S, Krause M, Kurochkin IV, Kusuma GD, Kuypers S, Laitinen S, Langevin SM, Languino LR, Lannigan J, Lasser C, Laurent LC, Lavieu G, Lazaro-Ibanez E, Le Lay S, Lee MS, Lee YXF, Lemos DS, Lenassi M, Leszczynska A, Li IT, Liao K, Libregts SF, Ligeti E, Lim R, Lim SK, Line A, Linnemannstons K, Llorente A, Lombard CA, Lorenowicz MJ, Lorincz AM, Lotvall J, Lovett J, Lowry MC, Loyer X, Lu Q, Lukomska B, Lunavat TR, Maas SL, Malhi H, Marcilla A, Mariani J, Mariscal J, Martens-Uzunova ES, Martin-Jaular L, Martinez MC, Martins VR, Mathieu M, Mathivanan S, Maugeri M, McGinnis LK, McVey MJ, Meckes DG, Jr., Meehan KL, Mertens I, Minciacchi VR, Moller A, Moller Jorgensen M, Morales-Kastresana A, Morhayim J, Mullier F, Muraca M, Musante L, Mussack V, Muth DC, Myburgh KH, Najrana T, Nawaz M, Nazarenko I, Nejsum P, Neri C, Neri T, Nieuwland R, Nimrichter L, Nolan JP, Nolte-'t Hoen EN, Noren Hooten N, O'Driscoll L, O'Grady T, O'Loghlen A, Ochiya T, Olivier M, Ortiz A, Ortiz LA, Osteikoetxea X, Ostergaard O, Ostrowski M, Park J, Pegtel DM, Peinado H, Perut F, Pfaffl MW, Phinney DG, Pieters BC, Pink RC, Pisetsky DS, Pogge von Strandmann E, Polakovicova I, Poon IK, Powell BH, Prada I, Pulliam L, Quesenberry P, Radeghieri A, Raffai RL, Raimondo S, Rak J, Ramirez MI, Raposo G, Rayyan MS, Regev-Rudzki N, Ricklefs FL, Robbins PD, Roberts DD, Rodrigues SC, Rohde E, Rome S, Rouschop KM, Rughetti A, Russell AE, Saa P, Sahoo S, Salas-Huenuleo E, Sanchez C, Saugstad JA, Saul MJ, Schiffelers

RM, Schneider R, Schoyen TH, Scott A, Shahaj E, Sharma S, Shatnyeva O, Shekari F, Shelke GV, Shetty AK, Shiba K, Siljander PR, Silva AM, Skowronek A, Snyder OL, 2nd, Soares RP, Sodar BW, Soekmadji C, Sotillo J, Stahl PD, Stoorvogel W, Stott SL, Strasser EF, Swift S, Tahara H, Tewari M, Timms K, Tiwari S, Tixeira R, Tkach M, Toh WS, Tomasini R, Torrecilhas AC, Tosar JP, Toxavidis V, Urbanelli L, Vader P, van Balkom BW, van der Grein SG, Van Deun J, van Herwijnen MJ, Van Keuren-Jensen K, van Niel G, van Royen ME, van Wijnen AJ, Vasconcelos MH, Vechetti IJ, Jr., Veit TD, Vella LJ, Velot E, Verweij FJ, Vestad B, Vinas JL, Visnovitz T, Vukman KV, Wahlgren J, Watson DC, Wauben MH, Weaver A, Webber JP, Weber V, Wehman AM, Weiss DJ, Welsh JA, Wendt S, Wheelock AM, Wiener Z, Witte L, Wolfram J, Xagorari A, Xander P, Xu J, Yan X, Yanez-Mo M, Yin H, Yuana Y, Zappulli V, Zarubova J, Zekas V, Zhang JY, Zhao Z, Zheng L, Zheutlin AR, Zickler AM, Zimmermann P, Zivkovic AM, Zocco D and Zuba-Surma EK (2018) Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles* **7**:1535750.

Tian T, Zhu YL, Hu FH, Wang YY, Huang NP and Xiao ZD (2013) Dynamics of exosome internalization and trafficking. *J Cell Physiol* **228**:1487-1495.

Traag VA, Waltman L and van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**:5233.

Trac D, Hoffman JR, Bheri S, Maxwell JT, Platt MO and Davis ME (2019a) Predicting Functional Responses of Progenitor Cell Exosome Potential with Computational Modeling. *Stem Cells Transl Med* **8**:1212-1221.

Trac D, Maxwell JT, Brown ME, Xu C and Davis ME (2019b) Aggregation of Child Cardiac Progenitor Cells Into Spheres Activates Notch Signaling and Improves Treatment of Right Ventricular Heart Failure. *Circ Res* **124**:526-538.

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS and Rinn JL (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381-386.

Umezu T, Ohyashiki K, Kuroda M and Ohyashiki JH (2013) Leukemia cell to endothelial cell communication via exosomal miRNAs. *Oncogene* **32**:2747-2755.

Umezu T, Tadokoro H, Azuma K, Yoshizawa S, Ohyashiki K and Ohyashiki JH (2014) Exosomal miR-135b shed from hypoxic multiple myeloma cells enhances angiogenesis by targeting factor-inhibiting HIF-1. *Blood* **124**:3748-3757.

Vagnozzi RJ, Maillet M, Sargent MA, Khalil H, Johansen AKZ, Schwanekamp JA, York AJ, Huang V, Nahrendorf M, Sadayappan S and Molkentin JD (2020) An acute immune response underlies the benefit of cardiac stem cell therapy. *Nature* **577**:405-409.

Vandergriff AC, de Andrade JB, Tang J, Hensley MT, Piedrahita JA, Caranasos TG and Cheng K (2015) Intravenous Cardiac Stem Cell-Derived Exosomes Ameliorate Cardiac Dysfunction in Doxorubicin Induced Dilated Cardiomyopathy. *Stem Cells Int* **2015**:960926.

Varga ZV, Zvara A, Farago N, Kocsis GF, Pipicz M, Gaspar R, Bencsik P, Gorbe A, Csonka C, Puskas LG, Thum T, Csont T and Ferdinandy P (2014) MicroRNAs associated with ischemia-reperfusion injury and cardioprotection by ischemic pre- and postconditioning: protectomiRs. *Am J Physiol Heart Circ Physiol* **307**:H216-227.

Vasa M, Fichtlscherer S, Aicher A, Adler K, Urbich C, Martin H, Zeiher AM and Dimmeler S (2001) Number and migratory activity of circulating endothelial progenitor cells inversely correlate with risk factors for coronary artery disease. *Circ Res* **89**:E1-7.

Vignard V, Labbe M, Marec N, Andre-Gregoire G, Jouand N, Fonteneau JF, Labarriere N and Fradin D (2020) MicroRNAs in Tumor Exosomes Drive Immune Escape in Melanoma. *Cancer Immunol Res* **8**:255-267.

Vinas JL, Spence M, Gutsol A, Knoll W, Burger D, Zimpelmann J, Allan DS and Burns KD (2018) Receptor-Ligand Interaction Mediates Targeting of Endothelial Colony Forming Cell-derived Exosomes to the Kidney after Ischemic Injury. *Sci Rep* **8**:16320.

Wang X, Wang HX, Li YL, Zhang CC, Zhou CY, Wang L, Xia YL, Du J and Li HH (2015) MicroRNA Let-7i negatively regulates cardiac inflammation and fibrosis. *Hypertension* **66**:776-785.

Wehman B and Kaushal S (2015) The emergence of stem cell therapy for patients with congenital heart disease. *Circ Res* **116**:566-569.

Wehman B, Pietris N, Bigham G, Siddiqui O, Mishra R, Li T, Aiello E, Jack G, Wang W, Murthi S, Sharma S and Kaushal S (2017) Cardiac Progenitor Cells Enhance Neonatal Right Ventricular Function After Pulmonary Artery Banding. *Ann Thorac Surg* **104**:2045-2053.

Williams JB, Jauch EC, Lindsell CJ and Campos B (2007) Endothelial microparticle levels are similar in acute ischemic stroke and stroke mimics due to activation and not apoptosis/necrosis. *Acad Emerg Med* **14**:685-690.

Willms E, Johansson HJ, Mager I, Lee Y, Blomberg KE, Sadik M, Alaarg A, Smith CI, Lehtio J, El Andaloussi S, Wood MJ and Vader P (2016) Cells release subpopulations of exosomes with distinct molecular and biological properties. *Sci Rep* **6**:22519.

Wolf FA, Angerer P and Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**:15.

Wolf P (1967) The nature and significance of platelet products in human plasma. *Br J Haematol* **13**:269-288.

Wolock SL, Lopez R and Klein AM (2019) Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**:281-291 e289.

Xin L, Lin X, Zhou F, Li C, Wang X, Yu H, Pan Y, Fei H, Ma L and Zhang S (2020) A scaffold laden with mesenchymal stem cell-derived exosomes for promoting endometrium regeneration and fertility restoration through macrophage immunomodulation. *Acta Biomater* **113**:252-266.

Yamada NO, Heishima K, Akao Y and Senda T (2019) Extracellular Vesicles Containing MicroRNA-92a-3p Facilitate Partial Endothelial-Mesenchymal Transition and Angiogenesis in Endothelial Cells. *Int J Mol Sci* **20**.

Yan F, Yao Y, Chen L, Li Y, Sheng Z and Ma G (2012) Hypoxic preconditioning improves survival of cardiac progenitor cells: role of stromal cell derived factor-1alpha-CXCR4 axis. *PLoS One* **7**:e37948.

Yin Z, Ma T, Huang B, Lin L, Zhou Y, Yan J, Zou Y and Chen S (2019) Macrophage-derived exosomal microRNA-501-3p promotes progression of pancreatic ductal adenocarcinoma through the TGFBR3-mediated TGF-beta signaling pathway. *J Exp Clin Cancer Res* **38**:310.

Yokoi A, Matsuzaki J, Yamamoto Y, Yoneoka Y, Takahashi K, Shimizu H, Uehara T, Ishikawa M, Ikeda SI, Sonoda T, Kawauchi J, Takizawa S, Aoki Y, Niida S, Sakamoto H, Kato K,

Kato T and Ochiya T (2018) Integrated extracellular microRNA profiling for ovarian cancer screening. *Nat Commun* **9**:4319.

Yu B, Kim HW, Gong M, Wang J, Millard RW, Wang Y, Ashraf M and Xu M (2015) Exosomes secreted from GATA-4 overexpressing mesenchymal stem cells serve as a reservoir of anti-apoptotic microRNAs for cardioprotection. *Int J Cardiol* **182**:349-360.

Zhang X, Wang X, Zhu H, Zhu C, Wang Y, Pu WT, Jegga AG and Fan GC (2010) Synergistic effects of the GATA-4-mediated miR-144/451 cluster in protection against simulated ischemia/reperfusion-induced cardiomyocyte death. *J Mol Cell Cardiol* **49**:841-850.

Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ and Bielas JH (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**:14049.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C and Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**:1523.

Zhu B, Zhang L, Liang C, Liu B, Pan X, Wang Y, Zhang Y, Zhang Y, Xie W, Yan B, Liu F, Yip HK, Yu XY and Li Y (2019) Stem Cell-Derived Exosomes Prevent Aging-Induced Cardiac Dysfunction through a Novel Exosome/lncRNA MALAT1/NF-kappaB/TNF-alpha Signaling Pathway. *Oxid Med Cell Longev* **2019**:9739258.

Zou L, Ma X, Lin S, Wu B, Chen Y and Peng C (2019) Long noncoding RNA-MEG3 contributes to myocardial ischemia-reperfusion injury through suppression of miR-7-5p expression. *Biosci Rep* **39**.