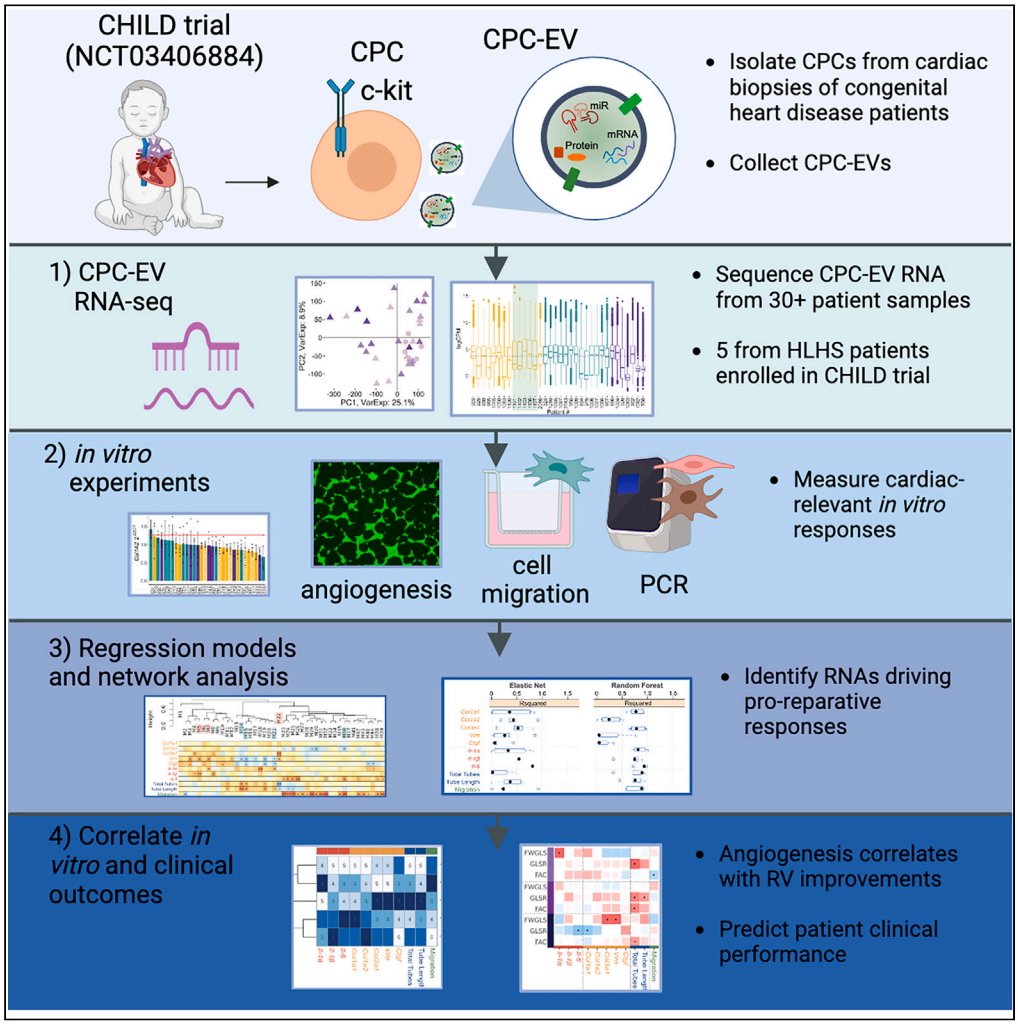
iScience

CellPress
OPEN ACCESS

Article

Statistical modeling of extracellular vesicle cargo to predict clinical trial outcomes for hypoplastic left heart syndrome



Jessica R. Hoffman, Hyun-Ji Park, Sruti Bheri, ..., Timothy C. Slesnick, William T. Mahle, Michael E. Davis

michael.davis@bme.gatech.edu

Highlights

Statistical models link CPC-EV cargo to cardiac-relevant *in vitro* outcomes

CPC-EV-induced angiogenesis correlates with clinical improvements in cardiac performance

CPC-EV miRNA content may be used as a biomarker for clinical performance

Hoffman et al., iScience *26*, 107980
October 20, 2023
https://doi.org/10.1016/j.isci.2023.107980

iScience

CellPress
OPEN ACCESS

Article

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] Manu O. Platt,[1] Joshua M. Hare,[3] Sunjay Kaushal,[4] Judith L. Bettencourt,[5] Dejian Lai,[5] Timothy C. Slesnick,[6,7] William T. Mahle,[6,7] and Michael E. Davis[1,2,7,8,*]

SUMMARY

Cardiac-derived c-kit+ progenitor cells (CPCs) are under investigation in the CHILD phase I clinical trial (NCT03406884) for the treatment of hypoplastic left heart syndrome (HLHS). The therapeutic efficacy of CPCs can be attributed to the release of extracellular vesicles (EVs). To understand sources of cell therapy variability we took a machine learning approach: combining bulk CPC-derived EV (CPC-EV) RNA sequencing and cardiac-relevant *in vitro* experiments to build a predictive model. We isolated CPCs from cardiac biopsies of patients with congenital heart disease (n = 29) and the lead-in patients with HLHS in the CHILD trial (n = 5). We sequenced CPC-EVs, and measured EV inflammatory, fibrotic, angiogenic, and migratory responses. Overall, CPC-EV RNAs involved in pro-reparative outcomes had a significant fit to cardiac development and signaling pathways. Using a model trained on previously collected CPC-EVs, we predicted *in vitro* outcomes for the CHILD clinical samples. Finally, CPC-EV angiogenic performance correlated to clinical improvements in right ventricle performance.

INTRODUCTION

Congenital heart disease (CHD) affects nearly 1% of births in the United States, with conditions ranging in severity.[1] Hypoplastic left heart syndrome (HLHS) is one of the most complex forms of CHD and is characterized by an underdeveloped left ventricle.[2,3] HLHS is palliated by a series of three surgeries which redirects blood flow 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.[4] In particular, patients with HLHS 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.[2] 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).[5]

CPCs have shown promising results in pre-clinical and early-phase clinical trials, including the CONCERT-HF trial for use in adult patients with heart failure.[6–9] CPCs are isolated from the heart tissue—mainly the right atrial appendage that is routinely removed during cannulation—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.[8,10–13] 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.[14] More specifically, transplanted stem or progenitor cells release extracellular vesicles (EVs) loaded with beneficial or reparative nucleic acid and protein cargo to resident cardiac cells.[15]

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 from a 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

[1]Wallace H. Coulter Department of Biomedical Engineering, Emory University School of Medicine & Georgia Institute of Technology, Atlanta, GA 30322, USA
[2]Molecular & Systems Pharmacology Graduate Training Program, Laney Graduate School, Emory University, Atlanta, GA 30322, USA
[3]Miller School of Medicine, University of Miami, Miami, FL 33136, USA
[4]Departments of Surgery and Pediatrics, Ann and Robert H. Lurie Children's Hospital of Chicago, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA
[5]Coordinating Center for Clinical Trials, Department of Biostatistics and Data Science, University of Texas Health Science Center School of Public Health, Houston, TX 77030, USA
[6]Division of Pediatric Cardiology, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA
[7]Children's Heart Research & Outcomes (HeRO) Center, Children's Healthcare of Atlanta & Emory University, Atlanta, GA 30322, USA
[8]Lead contact
*Correspondence: michael.davis@bme.gatech.edu
https://doi.org/10.1016/j.isci.2023.107980

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.[13,15–19] 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.[7,9,13,16] Here, we aim to build upon this work, expanding our overall sample size and collecting results for 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 cargo molecules (mRNA, miRNA, proteins, etc.) contributing to repair. Indeed, cardiac cell therapy has suffered from variability in patient outcomes; some cell types and/or patients exhibit greater improvements than others.[20] Therefore, to identify contributing factors of cell therapy variability and improve clinical outcomes, we need studies linking high dimensional EV RNA-sequencing datasets 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 patients with CHD (n = 29), and (2) CPCs from five patients enrolled in the phase I open-label group of the CHILD clinical trial (#1101, 1102, 1103, 1106, 1107). Then, we treated various cell types—cardiac endothelial cells (CECs), cardiac fibroblasts (CFs), and mesenchymal stromal cells (MSCs)—with CPC-EVs and measured CEC tube formation, MSC migration, CF fibrotic gene expression, and CEC inflammatory gene expression. We linked the EV sequencing data and *in vitro* outcomes data to generate regression models, potentially predictive of CHILD clinical trial outcomes.

RESULTS

Clinical CPC samples release small extracellular vesicles 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 select patients with HLHS enrolled in phase I of the CHILD clinical trial (testing dataset, n = 5). Patients' age and heart conditions are listed in Table 1. We expanded CPCs in 2D culture and isolated their EVs from the conditioned media via differential ultracentrifugation (Figure 1A). We assessed CPC-EV size with transmission electron microscopy (Figure 1B) and Nanosight particle tracking (Figure 1C), and determined that CPC-EVs were <150nm, characteristic of small EVs or exosomes. Next, we isolated CPC-EV RNA and performed total and small RNA sequencing. Initial bioanalyzer plots of CPC-EV RNA revealed the enrichment of small RNAs ~22nt, the size of miRNA, and confirmed the absence of ribosomal 18S and 28S RNA peaks (Figure 1D). Sequencing detected 1,067 miRNAs and 10,469 total RNAs after removing lowly expressed RNAs. CPC-EV sample TMM-normalized logCPMs are displayed in Figure 1E. 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 1F).

CPC-EV treatment affects recipient cell processes

Multiple studies have reported the pleiotropic effects of transplanted stem or progenitor cell-derived EVs.[13,15,16,18] 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-1α, Il-1β, Il-6*) after TNF-α treatment, and CF fibrotic gene expression (*Col1a1, Col1a2, Col3a1, Vim, Ctgf*) after TGF-β treatment. CPC-EVs derived from different patients exerted different effects on recipient cells. We observed two major sample clusters with a mix of CPC-EVs derived from various age groups (clusters A and B, Figure 2A; Figure S1). Cluster A included the testing set samples and showed greater migration, fibrotic and inflammatory gene expression, and lower tube formation values than cluster B. As expected, we observed *in vitro* outcomes largely clustered within assay type: fibrotic genes, tube formation measurements, and two of the three inflammatory genes (Figure 2A).

In the angiogenesis experiments, CPC-EVs induced CEC tube formation, as compared to the untreated control (Figures 2B and 2C). Interestingly, CPC-EVs derived from older patients induced greater CEC tube formation than CPC-EVs from neonate patients (Figure 2B). Linear regression showed enhanced total tube length measurements in CPC-EVs from older patients ($r = 0.34$, $p = 0.059$, Figure S2). In the migration assay, we observed the opposite trend: CPC-EVs derived from neonate patients promoted MSC migration to a greater extent than CPC-EVs derived from older patients (Figure 2D). Linear regression showed a modest decrease in MSC migration with age ($r = -0.30$, $p = 0.092$, Figure S2). We did not observe age-dependent differences in the PCR-based fibrosis and inflammation experiments with linear regression (Figures 2E, 2F, and S2). Interestingly, simplifying these analyses by binning CPC-EV responses into 3 age groups— neonate (<2 weeks), infant (2 weeks - 1 year), child (>1 year)— did not reveal a strong interaction between age group and experimental outcome (Figure S3). Furthermore, grouping CPC-EV samples by sex did not reveal differences between males and females across the various experimental outcomes (Figure S4).

Overall, CPC-EV treatment reduced *Il-6* expression in TNF-α treated CECs and reduced *Col1a2* and *Ctgf* expression in TGF-β treated CFs. Data for *Col1a1*, *Col3a1*, *Vim*, and Il-1β expression, as well as total number of measured tubes are provided in the supplement (Figure S5). Notably, for all four assays, we observed differences in experimental outcomes among the CPC-EVs. This variance among donor -derived CPC-EVs provides a solid foundation for constructing regression models that link CPC-EV RNA-seq with experimental data.

Table 1. Patient characteristics

Patient	Age	Age Group	Data Set	Diagnosis
894	6 months	infant	Training	Not reported
896	12 months	child	Training	Ventricle septal defect
902	4 years	child	Training	Atrial septal defect
903	<1 week	neonate	Training	Hypoplastic left heart syndrome
924	2 months	infant	Training	Total anomalous pulmonary venous return
925	<1 week	neonate	Training	Total anomalous pulmonary venous return
926	14 months	child	Training	Ventricle septal defect
930	<1 week	neonate	Training	Total anomalous pulmonary venous return
938	5 years	child	Training	Subaortic stenosis
956	5 months	infant	Training	Complete atrioventricular canal defect
957	7 months	infant	Training	Partial anomalous pulmonary venous return
975	6 months	infant	Training	Tetralogy of Fallot
985	2 weeks	neonate	Training	Interrupted aortic arch
1004	12 months	child	Training	Anomalous superior vena cava to left atrium
1006	6 months	infant	Training	Complete atrioventricular canal defect
1007	6 months	infant	Training	Ventricle septal defect
1010	9 weeks	infant	Training	Tetralogy of Fallot
1045	<1 week	neonate	Training	Coarctation of the aorta, ventricle septal defect
1048	6 years	child	Training	Atrial septal defect
1050	1 week	neonate	Training	Hypoplastic left heart syndrome
1063	3 years	child	Training	Ventricle septal defect
1066	6 months	infant	Training	Tricuspid atresia, ventricle septal defect, atrial septal defect
1083	<1 week	neonate	Training	Interrupted aortic arch, ventricle septal defect
1092	4 years	child	Training	Atrial septal defect
1095	3 months	infant	Training	Double outlet right ventricle
1097	3 months	infant	Training	Tetralogy of Fallot
1099	5 months	infant	Training	Complete atrioventricular canal defect
1100	<1 week	neonate	Training	Hypoplastic left heart syndrome
1101		neonate	Testing/CHILD	Hypoplastic left heart syndrome
1102		neonate	Testing/CHILD	Hypoplastic left heart syndrome
1103		neonate	Testing/CHILD	Hypoplastic left heart syndrome
1106		neonate	Testing/CHILD	Hypoplastic left heart syndrome
1107		neonate	Testing/CHILD	Hypoplastic left heart syndrome
2013	3 months	infant	Training	Ventricle septal defect
2016	1-2 weeks	neonate	Training	Coarctation of the aorta, atrial septal defect, ventricle septal defect

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

To link CPC-EV miRNA and RNA to functional *in vitro* outcomes (Figure 2), we initially used the WGCNA unsupervised learning method.[21] We used WGCNA to first identify clusters, or modules, of co-expressed CPC-EV RNAs, and then determined the Spearman's correlation of these modules to *in vitro* outcomes. We combined total RNA and miRNA datasets from all EV samples and standardized (mean centered and scaled) the RNA features and outcomes (Figure S6).

First, we performed WGCNA and identified 44 modules of co-expressed RNAs (Figure 3A; Table S1). Then, we correlated these modules to each continuous *in vitro* outcome and identified 7 modules of interest: M5, M7, M14, M15, M21, M22 and M36. Interestingly, although the RNA datasets were combined, centered, and scaled, the modules were primarily comprised of either RNA or miRNA (Figure 3B). RNA modules of interest included M14, M15, and M21 which had a negative correlation with fibrotic gene expression (*Vim*,

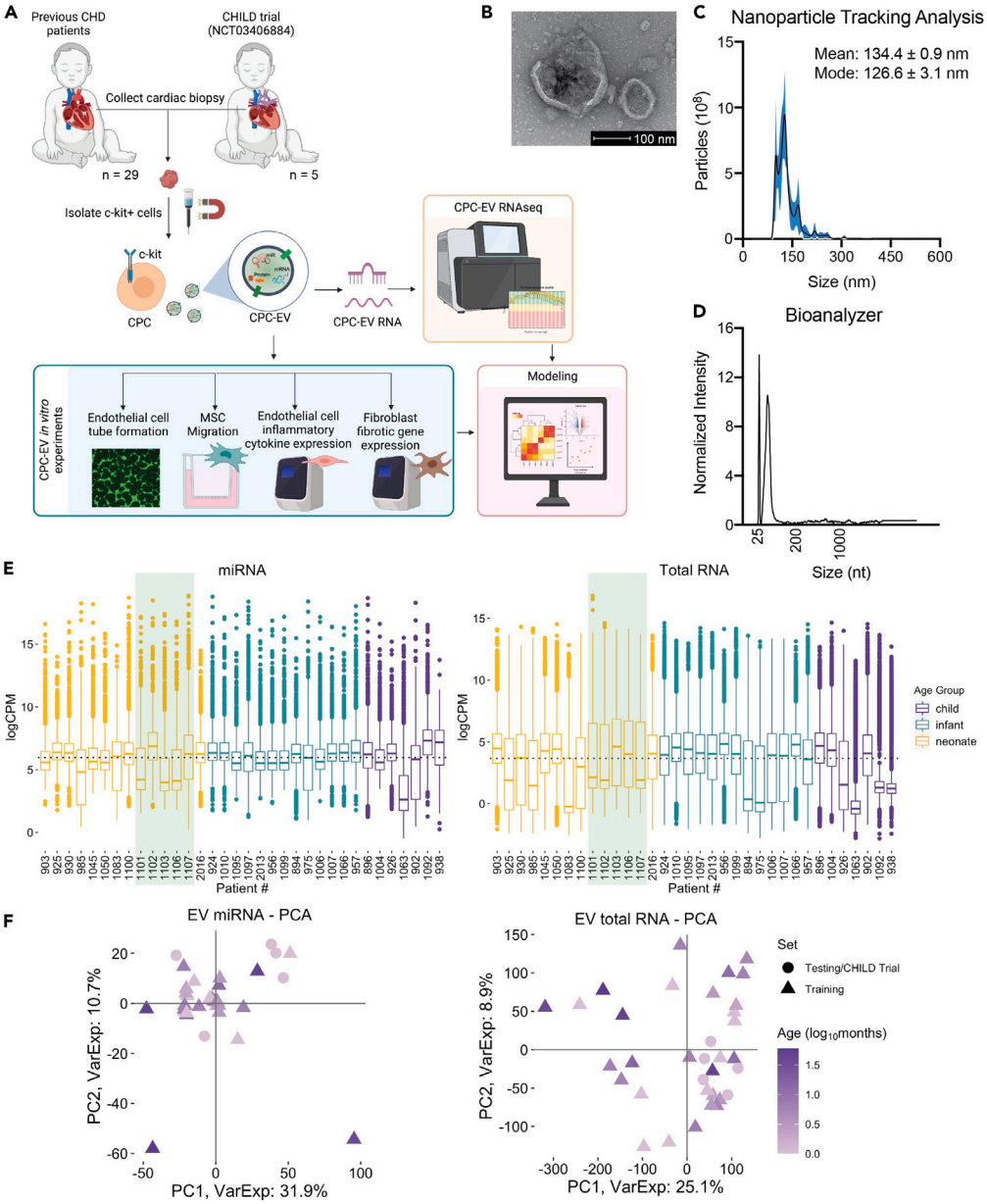


Figure 1. Characterization and RNA-sequencing of CPC-EVs

(A) Schematic of experimental design: CPCs were previously collected from various patients with CHD, as well as five patients with HLHS from the phase I open-label group in the CHILD trial. 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) Transmission electron microscopy image of EVs derived from patient 1101 CPCs.

(C and D) Nanoparticle tracking (mean ± SD) and (D) bioanalyzer plot of representative CPC-EVs from patient 1106 determine EV size/concentration and RNA composition, respectively.

(E) Boxplots show TMM-normalized counts of CPC-EV samples for miRNA and total RNA-sequencing. Middle line and box length represent the median and interquartile range, respectively. Outside "whisker" lines represent minimum and maximum logCPM values in each sample. Dots represent outlier logCPM values (>1.5 times the interquartile range). Samples from the CHILD trial shaded in green.

(F) PCA plots of CPC-EV samples for miRNA and total RNA-sequencing, coded by dataset and patient age. CHD: congenital heart disease; HLHS: hypoplastic left heart syndrome; MSC: mesenchymal stromal cell; PCA: principal component analysis; CPM: counts per million.

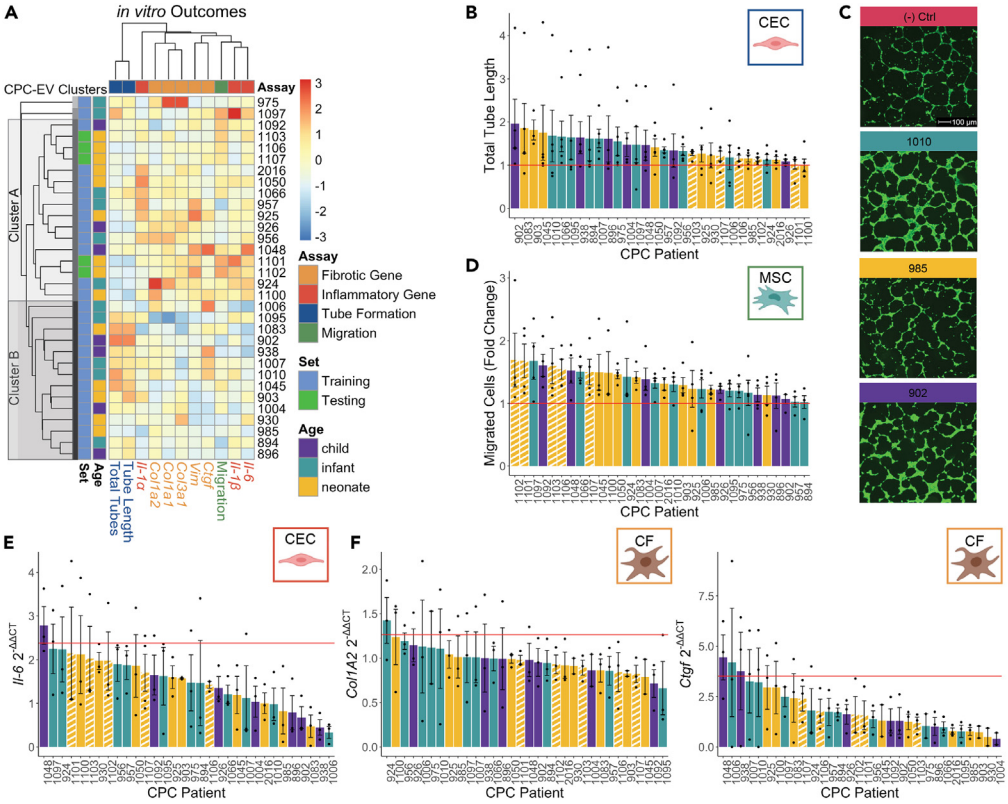


Figure 2. CPC-EV treatment affects cell processes *in vitro*

(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 dataset (training vs. testing). Two major CPC-EV sample clusters are highlighted in gray. Outcomes are listed in the columns and color-coded by assay category.

(B) Quantification of tube length, normalized to control (red line), and (C) representative images of CEC tube formation on Matrigel after 20 h with CPC-EV treatment. Yellow, white-striped bars represent samples from five neonates with HLHS, enrolled in the phase I open-label group in the CHILD trial.

(D) MSC migration in Boyden chamber system, normalized to control (red line), 48 h after CPC-EV treatment.

(E) CEC *Il-6* expression after 24-h CPC-EV pretreatment and 4-h TNF-α stimulation. Relative *Il-6* expression calculated using 2−ΔΔCt method, compared to the *Gapdh* housekeeping gene. Red line indicates control: CECs stimulated with TNF-α, without CPC-EV treatment.

(F) CF *Col1a2* (left) and *Ctgf* (right) expression after CPC-EV pretreatment and TGF-β stimulation. Relative *Col1a2 and Ctgf* expression calculated using 2−ΔΔCt method, compared to *Gapdh* housekeeping gene. Red lines indicate controls: CFs stimulated with TGF-β, without CPC-EV treatment. Data are represented as mean ± SEM (B, D–F). See also Figures S1–S5. CEC: cardiac endothelial cell; CF: cardiac fibroblast.

Ctgf or *Col1a2*) and a positive correlation with tube formation measurements ($p < 0.1$). RNA module M36 had a positive correlation with migration ($p < 0.01$) (Figure 3C). Upon the investigation of the RNAs belonging to these modules, we determined with pathway analysis that M14 was enriched in extracellular matrix organization and potassium ion homeostasis pathways; M14, M15, and M21 were enriched in membrane depolarization and muscle organ development pathways; M14 and M15 were enriched in macrophage, myeloid cell, and B cell-related immune responses; and M36 was enriched in RNAs involved in mRNA processing, and IL-12 and general cytokine signaling pathways (Figure 3D).

Additionally, we discovered that miRNA modules M5, M7, and M22 correlate with *in vitro* outcomes. These modules had a positive correlation with the expression of various fibrotic and inflammatory genes. M22 had a positive correlation with tube formation measurements. More concisely, these modules contained miRNAs which are positively correlated with "poor" outcomes (fibrosis and inflammation) and negatively correlated with 'good' *in vitro* outcomes (tube formation), opposite to the trend shown in the RNA modules (Figure 3C). 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 the enrichment of cell cycle, apoptosis, migration, and differentiation pathways, as well as Notch signaling, and positive regulation of fibroblast proliferation (Figure 3D).

Partial least-squares regression models predict CPC-EV *in vitro* outcomes for the test dataset

Next, we aimed to construct multivariate *in vitro* regression models that could be trained on CPC-EVs collected previously from various patients with CHD and would be predictive of our test dataset: CPC-EVs from the five phase I open-label CHILD trial patients. Partial least-squares regression (PLSR) is a supervised learning method that performs both dimension reduction and regression. Here, the SIMPLS

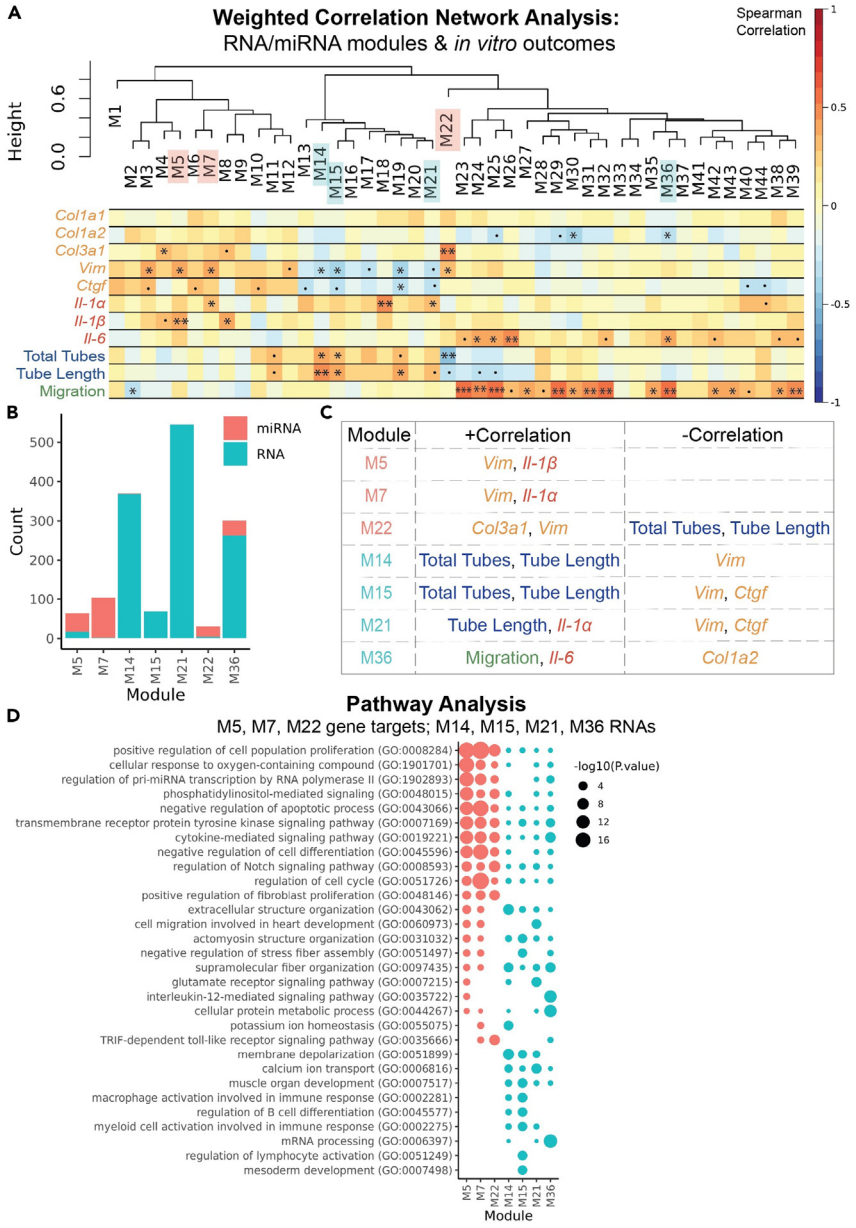


Figure 3. WGCNA identifies clusters of RNA which correlate to experimental outcomes

(A) Forty-four miRNA/RNA modules were identified with WGCNA and then correlated to *in vitro* outcomes. RNA data from all EV samples were used. Hierarchical clustering of the RNA modules is displayed in the top dendrogram. The bottom heatmap displays the Spearman's correlation of RNA modules to each outcome. ·p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001.

(B) RNA composition of 7 modules which significantly correlate to experimental outcomes – blue: total RNA, pink: miRNA.

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

(D) Pathway analysis of RNAs in modules 14, 15, 21, and 36 (blue) show the enrichment of supramolecular fiber organization, calcium ion transport, and immune response pathways. Pathway analysis of gene targets of miRNA modules 5, 7, and 22 show the enrichment of cytokine and Notch signaling, regulation of fibroblast proliferation, as well as cell migration, differentiation, death pathways (pink). Gene targets determined by miRTarBase with the validation of at least three experiments. See also Table S3. M: module.

PLSR algorithm computed components in RNA-seq ("X") space that maximized variance explained in the experimental outcomes ("Y") space. Given that our multivariate outcomes were correlated within each experimental category—after mean centering and scaling the data (Figure 4A)—we sought to construct four PLSR models (fibrosis, inflammation, angiogenesis, and migration). For each model we first constructed three-component full models based on all RNAs from the training CPC-EV set and leave-one-out cross-validation. The full models did not

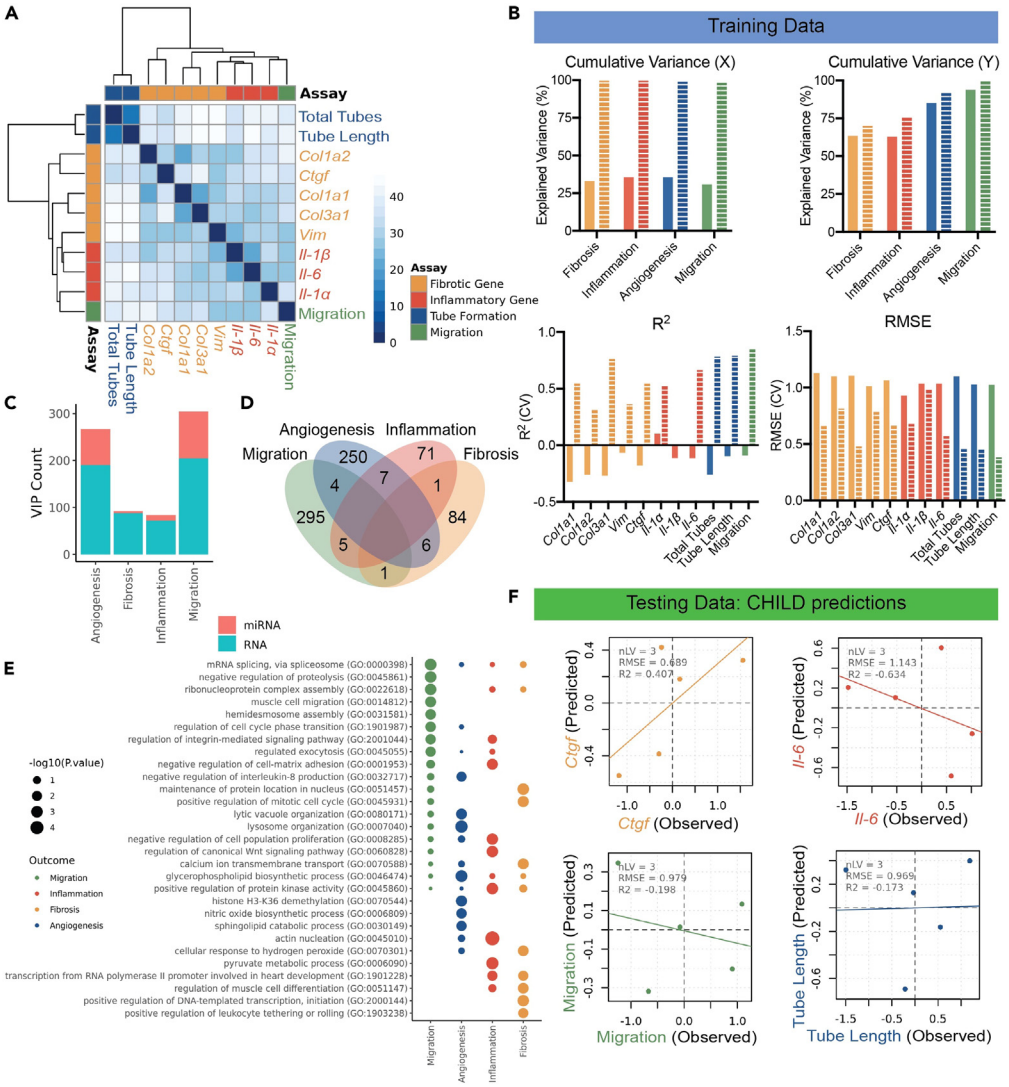


Figure 4. Partial least-squares regression models predict *in vitro* outcomes from test dataset

(A) Experimental outcomes (continuous variables) 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, R^2, and RMSE measurements from the four PLSR models, constructed from previously collected CPC-EVs (training set), before and after feature selection.

(C) composition of VIPs and (D) overlap of VIPs across the models.

(E) GO biological processes pathway analysis of VIP RNAs in each model.

(F) *in vitro* predictions for CHILD CPC-EVs. See also Figure S7. PLSR: partial least-squares regression; VIP: variable importance in the projection; RMSE: root-mean-square error.

capture a meaningful amount of variance (the models explained just 30–36% of the RNA variance) and showed poor prediction performance (high root-mean-square error (RMSE) and low R^2) within the cross-validated training set (Figure 4B).

However, we then used feature selection to reduce the model from >11,000 RNAs to <300 RNAs, and this 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. The reduced three-component models displayed higher performance metrics and captured greater variance, >98% of RNA variance and 70–99% of outcome variance in all models (Figure 4B). VIP count and distribution of RNA type are displayed in Figure 4C. Noticeably, there was little overlap of VIPs across model categories (Figure 4D). Pathway analysis of the VIP RNAs showed the enrichment of mRNA splicing, cell cycle, exocytosis, cell adhesion, IL-8 signaling, and Wnt signaling processes (Figure 4E). Finally, we validated the models on the test dataset and demonstrated that the reduced models predicted fibrotic *in vitro* outcomes of the CHILD CPC-EV dataset. Representative observed vs. predicted plots for *Ctgf*, *Il-6*, tube length, and migration are shown (Figure 4F; Figure S7).

Table 2. Tuned hyperparameters and testing set performance metrics for lasso, ridge, and elastic net regression models

Model	Parameter	*Col1a1*	*Col1a2*	*Col3a1*	*Vim*	*Ctgf*	*Il-1α*	*Il-1β*	*Il-6*	Total Tubes	Tube Length	Migration
Lasso	λ	0.66	0.30	0.51	0.76	0.76	0.76	0.34	0.76	0.01	0.26	0.34
	RMSE	0.89	1.03	0.83	0.89	0.89	0.88	0.82	0.89	0.94	1.04	1.01
	R^2	NA	0.27	0.18	NA	NA	0.27	0.22	NA	0.06	0.00	0.29
Ridge	λ	500	336.7	226.2	500	9.4	9.4	500	500	500	384.5	6.3
	RMSE	0.88	0.97	0.88	0.82	0.68	0.92	0.97	0.97	0.89	0.96	1.03
	R^2	0.77	0.77	0.10	0.18	0.70	0.01	0.37	0.41	0.01	0.03	0.41
Elastic Net	λ	0.84	0.14	0.65	0.11	0.03	1.03	0.78	0.97	0.05	0.59	0.35
	α	0.80	0.60	0.80	0.10	0.90	0.10	0.90	0.80	0.30	0.60	1.00
	RMSE	0.89	1.26	0.85	0.80	0.67	0.88	0.89	0.89	0.88	0.93	1.02
	R^2	NA	0.49	0.12	0.25	0.49	0.03	NA	NA	0.07	0.01	0.33

Random forest and regularized regression models select important CPC-EV RNA signals

To robustly identify CPC-EV RNA signals which covary with functional responses, we constructed additional regularized regression (lasso, ridge, and elastic net) and random forest regression models for each of the eleven outcomes with the training dataset. Regularized regression models invoke a penalty to minimize the weight of RNA features, a hyperparameter which controls the learning process. We tuned the shrinkage hyperparameter, λ, for each model using 5-fold cross-validation and selected the optimal model with the highest R^2 measurements. Briefly, lasso (L1) and ridge (L2) regularization add a penalty equal to the absolute value or the square of the magnitude of the coefficients, respectively. Simply put, RNA coefficients can only be reduced in ridge regression but may be fully eliminated in lasso regression.

The four model types produced similar error (mean absolute error (MAE) and RMSE) and R^2 metrics for the training set, based on repeated 5-fold cross-validation (Figure S8). Tuned hyperparameters for ridge and lasso models are listed in Table 2. Given the size of our dataset (>10,000 RNAs, 26 samples) we expected that the optimal models would impose high penalties, or result in high λ values. Indeed, five of our eleven optimal ridge models invoked the maximum λ value, 500. For ridge regression models, we set the maximum λ value to 500 to prevent shrinking feature weights too much, given that our goal for creating these models was to identify important RNA features for specific outcomes. We also determined that the *Vim* and *Il-6* lasso and ridge models had the highest shrinkage parameters. Further, we used elastic net regression, a combination of lasso and ridge methods. The α hyperparameter denotes the ratio of L1/lasso and L2/ridge penalties ($0 \leq \alpha \leq 1$ with 0 = ridge and 1 = lasso). Here, our *Vim*, *Il-1α*, and total tubes models favor the L2/ridge penalty ($\alpha \leq 0.3$) and all other models favor the L1/lasso penalty ($\alpha \geq 0.6$). Notably, *Col1a1* and *Il-6* models had the highest hyperparameter values, lasso $\lambda > 0.66$, ridge $\lambda = 500$, and $\alpha = 0.8$, indicating many RNA features have poor predictive power.

Next, with the optimal regularized regression models, we generated predictions for the CHILD CPC-EV testing dataset and observed considerable differences in the predictive performance of these models. Root-mean-square error (RMSE) and R^2 measurements for *in vitro* CHILD predictions are listed in Table 2. Generally, ridge models had the highest R^2 values, compared to their respective lasso and elastic net models. In datasets with a very large number of features, a considerable drawback to lasso regression—and subsequently, elastic net models with a high α value–is that models will select at most n features. For our problem, each lasso model would include at most 26 features (number of CPC-EV samples in the training set), picking one random feature per collinear set. In this way, ridge may produce the best models by shrinking, but not eliminating important and predictive RNA features.

Finally, we used the ensemble learning method, random forest regression, to model each outcome and predict CHILD *in vitro* outcomes. We tuned the splitting rule parameter and importantly, the number of features considered at each split (mtry). A good rule for random forest regression models is to use an mtry estimate of 1/3 of the number of features. Here, we determined that optimal *Vim* and migration models had an mtry parameter near this estimate, and that *Col1a1*, *Il-6*, *Il-a*, and *Il-1b* full models were the most predictive of the clinical test set ($R^2 > 0.48$, Table 3). However, overall, we observed that most full random forest models resulted in low R^2 values. Thus, we then computed feature importance scores and tested whether reducing our random forest models to the most informative features would result in better predictions, as demonstrated in our PLSR models. We constructed reduced models based on features with importance scores >10. Reducing random forest models produced mixed results: *Col1a2*, *Col3a1*, Il-1a, tube length, and migration models improved; *Col1a1, Vim, Ctgf, Il-1β, Il-6,* and total tubes models worsened. Finally, we compiled important features from all models, within each outcome category: retained RNAs in lasso and elastic net and the top RNAs in ridge (by coefficients) and random forest (by importance score, Figure S9A). Pathway analysis of RNAs deemed important to multiple regression model revealed: migration models fit cell cycle, endothelial cell differentiation, and cardiac epithelial to mesenchymal transition pathways; angiogenesis models fit histone exchange and microtubulue organization processes; inflammation models fit alpha-beta T cell activation and response to DNA damage processes; fibrosis models fit cytokine production regulation and leukocyte tethering pathways (Figure S9B).

Table 3. Tuned hyperparameters and testing set performance metrics for full and reduced random forest regression models

Model	Parameter	*Col1a1*	*Col1a2*	*Col3a1*	*Vim*	*Ctgf*	*Il-1α*	*Il-1β*	*Il-6*	Total Tubes	Tube Length	Migration
Full	mtry	4000	10	5250	6500	11250	8000	11250	11500	750	9500	1250
	split rule	V	V	V	V	ET	ET	V	ET	ET	ET	V
	RMSE	0.81	0.90	0.91	0.81	0.84	0.77	0.99	1.06	0.89	0.94	0.92
	R^2	0.49	0.00	0.04	0.24	0.25	0.51	0.64	0.97	0.01	0.00	0.08
Reduced	mtry	2	2	2	2	2	22	14	8	62	32	2
	split rule	V	V	V	V	V	ET	V	ET	ET	V	V
	# features	42	954	54	71	41	71	64	12	198	39	121
	RMSE	0.78	0.92	0.92	0.86	0.92	0.72	1.09	1.36	0.94	1.14	0.95
	R^2	0.40	0.29	0.17	0.09	0.04	0.62	0.50	0.91	0.00	0.17	0.13

V, variance; ET, extratrees.

CHILD clinical ranking predictions

To make clinical predictions for the five patients enrolled in the phase I open-label group of the CHILD clinical trial, we ranked CPC-EV *in vitro* outcomes from "best" to "worst" (Figure 5). CPC-EVs derived from patient #1101 performed "the worst" with the highest average score: 4.2. Patients #1103 and #1107 had the lowest average score, 2.5. CPC-EVs derived from patient #1103 showed the greatest pro-angiogenic potential, but higher inflammatory responses. Further, patient #1102 had the highest pro-migratory response and lowest *Il-1α* response.

Next, these *in vitro* rankings were correlated to clinical rankings in right ventricle (RV) improvements (FAC: fractional area change, GLSR: global longitudinal strain rate, FWGLS: free wall longitudinal strain). Correlations were made for improvements from baseline to six months, baseline to twelve months, and six months to twelve months. Across these timepoints, angiogenesis measurements correlated to improvements in RV GLSR and FAC. Interestingly, we observed a negative correlation between six-month to twelve-month improvements in GLSR and anti-inflammatory and anti-fibrotic responses (*Il-6* and *Col1a1)*. Nevertheless, confident validation of the model will require complete clinical data, which remain blinded and will become available next year. The clinical trial will collect patient right ventricle ejection fraction, end-diastolic volume, end-systolic volume, and tricuspid regurgitation at baseline, six months, and twelve months after treatment.[5] Correlating these data will allow us to determine the best, or most informative, predictors of clinical outcomes.

DISCUSSION

Outcomes for infants with HLHS have improved significantly with the emergence of the 3-stage palliation, including the Norwood, Bidirectional Glenn, and Fontan procedures. Surgical palliation supports single-ventricle physiology and redirects blood flow such that the RV maintains systemic circulation. Unfortunately, the RV is not physiologically adapted to support systemic circulation and HLHS patients often 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.[22] Furthermore, mortality rates for patients with HLHS remain the highest between stage I and II Norwood and Bidirectional Glenn surgeries.[23,24] Therefore, therapeutic intervention during this stage, preventing RV dysfunction, is necessary to improve outcomes for patients with HLHS.

Cell therapy has emerged as a promising avenue to promote cardiac repair and prevent adverse remodeling in younger patients. 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, and NCT04907526), and bone marrow-derived mesenchymal stem cells (NCT03525418 and NCT04925024). In particular, our group is involved in the CHILD clinical trial investigating use of autologous CPCs during the stage II Glenn operation to potentially support 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 that patients without genetic variants in a panel of 11 ACMG-classified CHD genes respond better to the cell therapy than patients with pathological variants.[25] However, there have been a dearth of studies investigating potential paracrine, including EV, factor determinants for cell therapy responsiveness. Here, we investigate the variability of CPC-EVs from >30 patients with CHD (including patients enrolled in the CHILD trial, n = 5), 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 mechanistically driving these functional responses.

Initially, we confirmed that CPC-EVs derived from various patients differentially affect cell processes (Figure 2). We determined that CPC-EV migration and angiogenesis responses correlated with patient age. Here, we recapitulate our previous results demonstrating that CPCs derived from younger patients exhibit greater paracrine pro-migratory behavior[8] (Figure 2D; Figure S2). 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.[8,13] However, these experiments were conducted using pooled

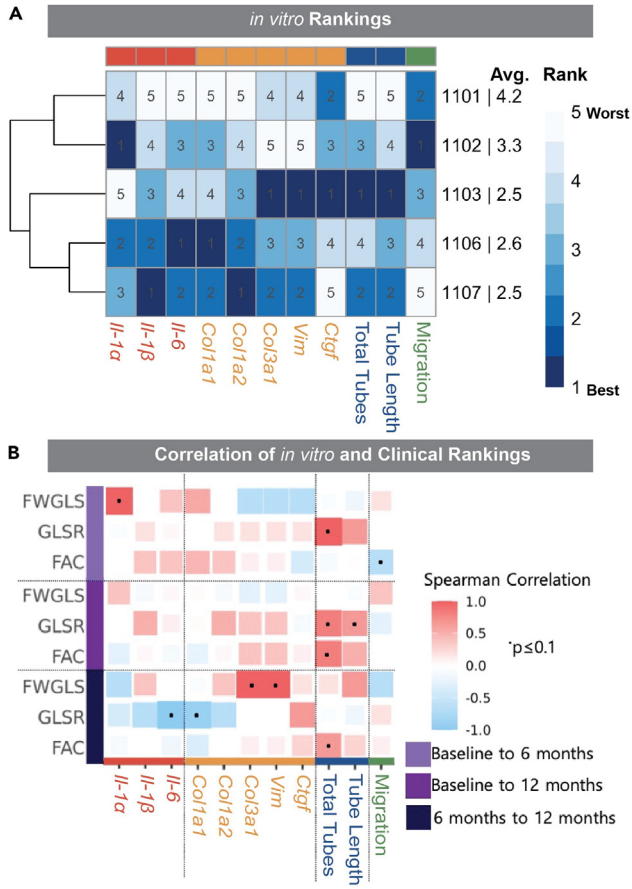


Figure 5. Correlation of *in vitro* CPC-EV performance and clinical CHILD patient outcomes

(A) *in vitro* outcomes were ranked from best (#1, anti-inflammatory, anti-fibrotic, pro-angiogenic, pro-migratory) to worst (#5) for the five patients with CHILD (#1101–1107).

(B) *in vitro* rankings were correlated (Spearman's) to their patient-matched clinical RV improvements associated with the trial. $\cdot p \leq 0.1$. FWGLS, free wall longitudinal strain; GLSR, global longitudinal strain rate; FAC, fractional area change.

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 2B) 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 the University of Miami, whereas the other CPCs were isolated and expanded by our group. However, the purpose of this study was to understand CPC-EV 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 that 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 datasets this size, allows for the implementation of machine learning methods to reduce dataset dimension and complexity and summarize the data. Initially, we used WGNCA, an unsupervised learning approach, to identify modules of co-expressed EV-RNAs correlate these modules to *in vitro* outcomes. WGCNA identified 3 miRNA enriched modules that had a negative correlation to inflammatory and fibrotic *in vitro* responses. The gene targets of these miRNAs were enriched in biological processes related to cell proliferation, apoptosis, cytokine signaling, Notch signaling, and fibroblast proliferation. Of note, module 5 contained well-studied miRNAs -139-5p (targets CXCR4, MCL1, and NOTCH1) and -149-5p (targets IL6, MYD88, and FGFR1). Module 7 included well-studied miR-302d-3p (targets CXCL8, CDK2, CCND2, and TGFBR2) and −135b-5p (targets SMAD5 and TGFBR1). Given that EV miRNAs are particularly potent, it may be expected that clusters of miRNAs that target inflammatory and fibrotic pathways correlate to their corresponding *in vitro* outcomes.

Furthermore, to directly link our CPC-EV RNA sequencing to experimental outcomes, we constructed univariate regularized regression and random forest regression models, as well as multivariate partial least-squares regression (PLSR) models of angiogenesis, migration,

inflammation, and fibrosis responses. We constructed models using the training data—previously isolated CPCs from patients with CHD—and measured the *in vitro* predictability in our test set— five patient-derived CPC-EVs in the phase I open-label group from the CHILD trial. As to be expected, in all cases, our original, full models described little variance of the RNAs, and produced poor prediction metrics for the cross-validation of our training set. 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 4D; Figure S9). In the case of comparing across outcomes, this may indicate that different EV RNAs are driving different mechanisms.

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 various regression models, respectively. We also determined that *in vitro* angiogenic potential of EVs correlates to clinical improvements in RV performance of patients with HLHS. Ultimately, we aim to connect these results to our phase I, patient-matched, results. Here, we have ranked the *in vitro* performance of the CPC-EVs from patients included in the CHILD trial (Figure 5) and made predictions as to which patient may perform better. These predictions will be validated when the trial is complete and unblinded. Given the importance of EV release in cell-therapy efficacy, our results and prediction tools will help to inform clinicians of patient outcomes, thereby potentially minimizing poor outcomes.[26]

Limitations of the study

In this article, we aimed to connect progenitor cell-derived EV cargo and their pro-reparative *in vitro* effect to clinical outcomes. Our study is primarily limited by clinical sample size. First, our training and testing datasets consisted of different patient populations; we trained our models with CPC-EV samples derived from cardiac biopsies of patients with a range of diagnoses and tested our models with CPC-EVs derived from patients with HLHS in the CHILD trial. Given a limited availability of HLHS cardiac biopsies—only three of the training set CPC-EVs were derived from patients with HLHS—we felt it was best to use the largest training dataset possible, and utilized samples collected from patients of varying congenital heart disease conditions. Second, we are constrained by our phase I trial sample size. Here, we were able to include data from five patients who completed the trial (four males and 1 female), with visits six months and twelve months post-surgery. Third, we are limited to clinical rankings in RV function (fractional area change, global longitudinal strain rate, and free wall longitudinal strain). We will need to wait until the trial is unblinded in early 2025 to validate and refine the model predictions.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICPANT DETAILS
  - Isolation and culture of c-kit+ progenitor cells (CPCs)
  - Extracellular vesicle (EV) collection
  - Tube formation assay
  - Mesenchymal stromal cell (MSC) migration assay
  - Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)
  - Fibroblast TGF-β stimulation assay
  - Endothelial cell TNF-α stimulation assay
  - Next generation sequencing
  - RNA sequencing data analysis
  - Identification of miRNA gene targets and pathway analysis
  - WGCNA gene module detection
  - Regression models
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107980.

ACKNOWLEDGMENTS

The authors acknowledge funding by the National Institutes of Health grants: R01HL145644 (M.E.D.), F31HL154725 (J.R.H.), T32GM008602 (J.R.H.), as well as the American Heart Association Postdoctoral Fellowship 837187 (H.J.P.), and Additional Ventures. This work was also funded by a grant from the Marcus Foundation (M.E.D. and J.M.H).

AUTHOR CONTRIBUTIONS

Conceptualization: JRH, MED; Methodology: JRH, HJP, SB, TCS; Software: JRH; Formal Analysis: JRH, TCS; Investigation: JRH, HJP, SB; Resources: JMH, SK, JLB, DL, TCS, WTM; Data Curation: JLB, DL; Writing—original draft: JRH; Writing—review & editing: JRH, HJP, SB, MED, MOP, DL, JLB; Visualization: JRH; Project Administration: JLB, DL, JMH, SK, TCS, WTM; Supervision: MED.

DECLARATION OF INTERESTS

Joshua M. Hare reports having a patent for cardiac cell-based therapy and holds equity in Vestion Inc. and maintains a professional relationship with Vestion Inc. as a consultant and member of the Board of Directors and Scientific Advisory Board. Vestion Inc. did not play a role in the design, conduct, or funding of the study. Dr. Joshua Hare is the Chief Scientific Officer, a compensated consultant and board member for Longeveron Inc. and holds equity in Longeveron. Dr. Hare is also the co-inventor of intellectual property licensed to Longeveron. The University of Miami is an equity owner in Longeveron Inc., which has licensed intellectual property from the University of Miami.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: March 16, 2023
Revised: August 24, 2023
Accepted: September 15, 2023
Published: September 21, 2023

REFERENCES

1. Hoffman, J.I.E., and Kaplan, S. (2002). The incidence of congenital heart disease. J. Am. Coll. Cardiol. *39*, 1890–1900. https://doi.org/10.1016/s0735-1097(02)01886-7.
2. Bittle, G.J., Morales, D., Deatrick, K.B., Parchment, N., Saha, P., Mishra, R., Sharma, S., Pietris, N., Vasilenko, A., Bor, C., et al. (2018). Stem Cell Therapy for Hypoplastic Left Heart Syndrome: Mechanism, Clinical Application, and Future Directions. Circ. Res. *123*, 288–300. https://doi.org/10.1161/CIRCRESAHA.117.311206.
3. Feinstein, J.A., Benson, D.W., Dubin, A.M., Cohen, M.S., Maxey, D.M., Mahle, W.T., Pahl, E., Villafañe, J., Bhatt, A.B., Peng, L.F., et al. (2012). Hypoplastic left heart syndrome: current considerations and expectations. J. Am. Coll. Cardiol. *59*, S1–S42. https://doi.org/10.1016/j.jacc.2011.09.022.
4. Saraf, A., Book, W.M., Nelson, T.J., and Xu, C. (2019). Hypoplastic left heart syndrome: From bedside to bench and back. J. Mol. Cell. Cardiol. *135*, 109–118. https://doi.org/10.1016/j.yjmcc.2019.08.005.
5. Kaushal, S., Hare, J.M., Shah, A.M., Pietris, N.P., Bettencourt, J.L., Piller, L.B., Khan, A., Snyder, A., Boyd, R.M., Abdullah, M., et al. (2022). Autologous Cardiac Stem Cell Injection in Patients with Hypoplastic Left Heart Syndrome (CHILD Study). Pediatr. Cardiol. *43*, 1481–1493. https://doi.org/10.1007/s00246-022-02872-6.
6. Bolli, R., Mitrani, R.D., Hare, J.M., Pepine, C.J., Perin, E.C., Willerson, J.T., Traverse, J.H., Henry, T.D., Yang, P.C., Murphy, M.P., et al. (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. https://doi.org/10.1002/ejhf.2178.
7. Saha, P., Sharma, S., Korutla, L., Datla, S.R., Shoja-Taheri, F., Mishra, R., Bigham, G.E., Sarkar, M., Morales, D., Bittle, G., et al. (2019). Circulating exosomes derived from transplanted progenitor cells aid the functional recovery of ischemic myocardium. Sci. Transl. Med. *11*, eaau1168. https://doi.org/10.1126/scitranslmed.aau1168.
8. Agarwal, U., Smith, A.W., French, K.M., Boopathy, A.V., George, A., Trac, D., Brown, M.E., Shen, M., Jiang, R., Fernandez, J.D., et al. (2016). Age-Dependent Effect of Pediatric Cardiac Progenitor Cells After Juvenile Heart Failure. Stem Cells Transl. Med. *5*, 883–892. https://doi.org/10.5966/sctm.2015-0241.
9. Trac, D., Maxwell, J.T., Brown, M.E., Xu, C., and Davis, M.E. (2019). Aggregation of Child Cardiac Progenitor Cells Into Spheres Activates Notch Signaling and Improves Treatment of Right Ventricular Heart Failure. Circ. Res. *124*, 526–538. https://doi.org/10.1161/CIRCRESAHA.118.313845.
10. Sharma, S., Mishra, R., Bigham, G.E., Wehman, B., Khan, M.M., Xu, H., Saha, P., Goo, Y.A., Datla, S.R., Chen, L., et al. (2017). A Deep Proteome Analysis Identifies the Complete Secretome as the Functional Unit of Human Cardiac Progenitor Cells. Circ. Res. *120*, 816–834. https://doi.org/10.1161/CIRCRESAHA.116.309782.
11. Simpson, D.L., Mishra, R., Sharma, S., Goh, S.K., Deshmukh, S., and Kaushal, S. (2012). A strong regenerative ability of cardiac stem cells derived from neonatal hearts. Circulation *126*, S46–S53. https://doi.org/10.1161/CIRCULATIONAHA.111.084699.
12. Mishra, R., Vijayan, K., Colletti, E.J., Harrington, D.A., Matthiesen, T.S., Simpson, D., Goh, S.K., Walker, B.L., Almeida-Porada, G., Wang, D., et al. (2011). Characterization and functionality of cardiac progenitor cells in congenital heart patients. Circulation *123*, 364–373. https://doi.org/10.1161/CIRCULATIONAHA.110.971622.
13. Agarwal, U., George, A., Bhutani, S., Ghosh-Choudhary, S., Maxwell, J.T., Brown, M.E., Mehta, Y., Platt, M.O., Liang, Y., Sahoo, S., and Davis, M.E. (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. https://doi.org/10.1161/CIRCRESAHA.116.309935.
14. Marbán, E. (2018). A mechanistic roadmap for the clinical application of cardiac cell therapies. Nat. Biomed. Eng. *2*, 353–361. https://doi.org/10.1038/s41551-018-0216-z.
15. Ibrahim, A.G.E., Cheng, K., and Marbán, E. (2014). Exosomes as critical agents of cardiac regeneration triggered by cell therapy. Stem Cell Rep. *2*, 606–619. https://doi.org/10.1016/j.stemcr.2014.04.006.
16. Gray, W.D., French, K.M., Ghosh-Choudhary, S., Maxwell, J.T., Brown, M.E., Platt, M.O., Searles, C.D., and Davis, M.E. (2015). Identification of therapeutic covariant microRNA clusters in hypoxia-treated cardiac progenitor cell exosomes using systems biology. Circ. Res. *116*, 255–263. https://doi.org/10.1161/CIRCRESAHA.116.304360.
17. Hoffman, J.R., Park, H.J., Bheri, S., Jayaraman, A.R., and Davis, M.E. (2022). Comparative computational RNA analysis of cardiac-derived progenitor cells and their extracellular vesicles. Genomics *114*, 110349.

https://doi.org/10.1016/j.ygeno.2022.110349.

18. Sahoo, S., Klychko, E., Thorne, T., Misener, S., Schultz, K.M., Millay, M., Ito, A., Liu, T., Kamide, C., Agrawal, H., et al. (2011). Exosomes from human CD34(+) stem cells mediate their proangiogenic paracrine activity. Circ. Res. *109*, 724–728. https://doi.org/10.1161/CIRCRESAHA.111.253286.
19. Yu, B., Kim, H.W., Gong, M., Wang, J., Millard, R.W., 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. https://doi.org/10.1016/j.ijcard.2014.12.043.
20. Banerjee, M.N., Bolli, R., and Hare, J.M. (2018). Clinical Studies of Cell Therapy in Cardiovascular Medicine: Recent Developments and Future Directions. Circ. Res. *123*, 266–287. https://doi.org/10.1161/CIRCRESAHA.118.311217.
21. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.
22. Altmann, K., Printz, B.F., Solowiejczky, D.E., Gersony, W.M., Quaegebeur, J., and Apfel, H.D. (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. https://doi.org/10.1016/s0002-9149(00)01131-0.
23. Carlo, W.F., Carberry, K.E., Heinle, J.S., Morales, D.L., McKenzie, E.D., Fraser, C.D., Jr., and Nelson, D.P. (2011). Interstage attrition between bidirectional Glenn and Fontan palliation in children with hypoplastic left heart syndrome. J. Thorac. Cardiovasc. Surg. *142*, 511–516. https://doi.org/10.1016/j.jtcvs.2011.01.030.
24. McGuirk, S.P., Griselli, M., Stumper, O.F., Rumball, E.M., Miller, P., Dhillon, R., de Giovanni, J.V., Wright, J.G., Barron, D.J., and Brawn, W.J. (2006). Staged surgical management of hypoplastic left heart syndrome: a single institution 12 year experience. Heart *92*, 364–370. https://doi.org/10.1136/hrt.2005.068684.
25. Rieger, A.C., Myerburg, R.J., Florea, V., Tompkins, B.A., Natsumeda, M., Premer, C., Khan, A., Schulman, I.H., Vidro-Casiano, M., DiFede, D.L., et al. (2019). Genetic determinants of responsiveness to mesenchymal stem cell injections in non-ischemic dilated cardiomyopathy. EBioMedicine *48*, 377–385. https://doi.org/10.1016/j.ebiom.2019.09.043.
26. Kaushal, S., Hare, J.M., Hoffman, J.R., Boyd, R.M., Ramdas, K.N., Pietris, N., Kutty, S., Tweddell, J.S., Husain, S.A., Menon, S.C., et al. (2023). Intramyocardial cell-based therapy with Lomecel-B during bidirectional cavopulmonary anastomosis for hypoplastic left heart syndrome: the ELPIS phase I trial. Eur. Heart J. Open. *3*, oead002. https://doi.org/10.1093/ehjopen/oead002.
27. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
28. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.
29. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523. https://doi.org/10.1038/s41467-019-09234-6.
30. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. *33*, 1–22.
31. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. J. Stat. Softw. *28*, 1–26.
32. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. *14*, 128. https://doi.org/10.1186/1471-2105-14-128.
33. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.
34. Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M., et al. (2011). miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. *39*, D163–D169. https://doi.org/10.1093/nar/gkq1107.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD117	Santa Cruz	sc-5535
GATA-4	Santa Cruz	sc-9053
CD34	eBioscience	11-0341-81
CD117 MicroBead, human	Miltenyi Biotec	130-091-332
Biological Samples		
Cardiac-derived c-kit+ progenitor cells, human	Children's Healthcare of Atlanta and University of Miami Clinical Research Cell Manufacturing Program	Table 1
Chemicals, Peptides, and Recombinant Proteins		
Matrigel Matrix	Corning	356237
Geltrex	Gibco	A1413302
Recombinant TGFB1, human	Invitrogen	PHG9204
Recombinant TNF-alpha, rat	R&D Systems	510-RT
Recombinant basic fibroblast growth factor , human	STEMCELL Technologies	78003
Fetal Bovine Serum	Cytiva	SH30071.03
Penicillin-Streptonmycin	Thermo Fisher Scientific	15140122
L-glutamine	MP Biomedicals	091680149
Calcein, AM	Invitrogen	C3099
Endothelial Cell Growth Medium-2 BulletKit	Lonza	CC-3162
Ham's F-12 medium	Corning	10-080-CV
Dulbecco's Modified Eagle Medium and Ham's F-12 medium	Corning	10-090-CM
Critical Commercial Assays		
Micro BCA Protein Assay Kit	Thermo Scientific Pierce	23235
CellTracker Orange CMRA Dye	Invitrogen	C34551
Pure Link RNA Mini Kit	Invitrogen	12183018A
miRNeasy Mini Kit	Qiagen	217084
SYBR Green PCR Master Mix	Applied Biosystems	4309155
Deposited Data		
CPC-EV bulk RNA sequencing	GEO	GSE202345, GSE202347, GSE203512
Experimental Models: Cell Lines		
Mesenchymal stromal cells, human	StemPro/Gibco	A15652
Cardiac endothelial cells and fibroblasts, rat	Emory University	N/A
Oligonucleotides		
Primer sequences	Integrated DNA technologies	Table S2
Software and Algorithms		
Nanoparticle tracking analysis software	NanoSight NS300	NanoSight NTA 3.4
ImageJ	NIH	ImageJ.org
StepOne Real-Time PCR System	Applied Biosystems	4376357

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
2100 Bioanalyzer and TapeStation Controller	Agilent Genomics	N/A
QIAseq miRNA Quantification tool	Qiagen GeneGlobe	geneglobe.qiagen.com/
STAR aligner	Dobin et al.[27]	github.com/alexdobin/STAR
miRTarBase	Hsu et al.[28]	mirtarbase.cuhk.edu.cn
Metascape	Zhou et al.[29]	metascape.org
edgeR, R package	Robinson et al.[28]	bioconductor.org/packages/release/bioc/html/edgeR.html
WGCNA, R package	Langfelder and Horvath[21]	horvath.genetics.ucla.edu/
mdatools, R package	Sergey Kucheryavskiy	github.com/svkucheryavski/mdatools
glmnet, R package	Friedman, Hastie, and Tibshirani[30]	glmnet.stanford.edu/
caret, R package	Kuhn[31]	topepo.github.io/caret/
enrichR, R package	Chen et al.[32]	maayanlab.cloud/Enrichr/
Other		
Transwell, 6.5 mm with 8.0 μm pore	Corning	3422
Ibidi μ-Slide Angiogenesis	Ibidi	81506

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Michael Davis (michael.davis@bme.gatech.edu).

Materials availability

This study did not generate new unique resources.

Data and code availability

- RNA-sequencing data have been deposited (GEO: GSE203512) and is publicly accessible.
- Code was generated using a variety of published algorithms (R packages referenced) and is available from the lead contact upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICPANT DETAILS

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 (IRB00005500 and IRB00096402). Human CPCs were isolated from the right atrial appendage tissue, routinely removed during surgical repair of congenital heart defects, via magnetic cell sorting (CD177 human, Miltenyi Biotec, MD, USA). CPCs from the CHILD trial (#1101, 1102, 1103, 1106, 1107; 4 males, 1 female) were initially isolated at the University of Miami Clinical Research Cell Manufacturing Program, as described,[5] and shipped to Emory for expansion. In total, CPCs were collected from 14 neonatal (<2 weeks), 13 infant (2 weeks - 1 year), and 7 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. CPCs were characterized with flow cytometry (Figure S10). Patient characteristics for samples used in the study are listed in Table 1.

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. Morphology and size of EVs were initially characterized by transmission electron microscopy imaging with negative staining at the Emory Integrated Electron Microscopy Core. EV size and concentration was determined by NanoSight NS300. Samples were diluted 1:100 in PBS, and three, 60-second video images were captured per sample and analyzed by NanoSight NTA 3.4 software. Of note, CPCs and CPC-EVs were collected from 37 patients total, however, patient #1048 CPC-EVs were not sequenced and patients #1063, #1099, and #2013 CPC-EVs were not examined in *in vitro* assays, due to low vesicle yield.

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 negative and positive controls, respectively. Tube formation experiments were performed five times (n=5) for each CPC-EV.

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 serum free 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 in serum free media). MSC migration experiments were performed four times (n=4) for each CPC-EV.

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 (Invitrogen) 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). Relative mRNA levels were calculated using the $2^{-\Delta\Delta Ct}$ method, as compared to Glyceraldehyde 3-phosphate dehydrogenase (*Gapdh*) housekeeping gene. Primer sequences are listed in Table S2.

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

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 (no growth factors, no serum). CECs were treated with 20 μg/mL protein of EV in 500 μL of bare 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 RT-qPCR analysis. CECs without EV treatment in bare media, 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. RT-qPCR was used to evaluate transcript expression of interleukin 1β (*Il-1β*), interleukin 1α (*Il-1α*), and interleukin 6 (*Il-6*).

Next generation sequencing

In this study, we combined RNA 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 performed RNA sequencing on 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 files were aligned 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.[27] 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.

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.[28] RNA counts were normalized with edgeR's weighted trimmed mean of M-values method (TMM) using the 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.[33] Multidimensional scaling plots before and after batch correction are shown in Figure S11.

Identification of miRNA gene targets and pathway analysis

Gene targets of miRNAs were determined using the database miRTarBase.[34] Targets experimentally validated by at least three different methods (e.g. microarray, NGS, western blot, reporter assay, etc.) were considered for pathway analysis. Pathway analysis of RNAs and miRNA gene targets was performed using Metascape express analysis.[29]

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.[21] RNA and miRNA datasets were combined, mean centered and scaled using mdatools R package's prep.autoscale function. The optimal soft-threshold power was graphically determined ($\beta = 9$, Table S3) 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. 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 with multiple outcomes ($p < 0.1$) were examined for biological significance with enrichR GO: Biological Processes pathway analysis.[32]

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 VIP RNAs and VIP miRNA gene targets. 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 outcome using 5-fold cross validation with 3 repeats. 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.

Ridge, lasso, and elastic net regression was performed using the glmnet and caret R packages.[30,31] Models were constructed from the training data for each experimental outcome using 5-fold cross validation with 3 repeats. Given the high number of features and low variance in error metrics, optimal models were selected based on R^2 values, to return an informative number of features. For ridge and lasso regression models, the hyperparameter, λ, was tuned (100 values from 10^{-3} to 500). For elastic net models, a tuneLength of 10 was set to test combinations of 10 different values of λ and α. Feature coefficients were taken from the final, optimal model.

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantification and statistical analyses were performed as described in the experimental model and subject details section of the STAR Methods.

ADDITIONAL RESOURCES

This study involves data and materials collected from patients enrolled in The CHILD Trial: Hypoplastic Left Heart Syndrome Study (NCT03406884). Measurements in right ventricle fractional area change, global longitudinal strain rate, free wall longitudinal strain were made before treatment, as well as six-months and twelve-months post-surgery. Measurements from patients were ranked before use in this manuscript.