Dissecting mechanisms of chamber-specific cardiac differentiation and its perturbation following retinoic acid exposure

David M Gonzalez, Nadine Schrode, Tasneem A M Ebrahim, Nicolas Broguiere, Giuliana Rossi, Lika Drakhlis, Robert Zweigerdt, Matthias P Lutolf, Kristin G Beaumont, Robert Sebra and Nicole C Dubois
DOI: 10.1242/dev.200557

Editor: Benoit Bruneau

Review timeline
Original submission: 1 February 2022
Editorial decision: 15 March 2022
First revision received: 19 May 2022
Accepted: 26 May 2022

Original submission

First decision letter

MS ID#: DEVELOP/2022/200557

MS TITLE: Dissecting Mechanisms of Chamber-Specific Cardiac Differentiation and its Perturbation Following Retinoic Acid Exposure

AUTHORS: David Michael Gonzalez, Nadine Schrode, Tasneem Ebrahim, Nicolas Broguiere, Giuliana Rossi, Lika Drakhlis, Robert Zweigerdt, Matthias Lutolf, Kristin Beaumont, Robert Sebra, and Nicole Caroline Dubois

I have now received all the referees reports on the above manuscript, and have reached a decision. The referees' comments are appended below, or you can access them online: please go to BenchPress and click on the 'Manuscripts with Decisions' queue in the Author Area.

The overall evaluation is positive and we would like to publish a revised manuscript in Development, provided that the referees' comments can be satisfactorily addressed. Please attend to all of the reviewers' comments in your revised manuscript and detail them in your point-by-point response. If you do not agree with any of their criticisms or suggestions explain clearly why this is so.

We are aware that you may currently be unable to access the lab to undertake experimental revisions. If it would be helpful, we encourage you to contact us to discuss your revision in greater detail. Please send us a point-by-point response indicating where you are able to address concerns raised (either experimentally or by changes to the text) and where you will not be able to do so within the normal timeframe of a revision. We will then provide further guidance. Please also note that we are happy to extend revision timeframes as necessary.

Reviewer 1

Advance summary and potential significance to field

The manuscript from Gonzalez et al examines development of the heart and surrounding tissues in mouse embryos. The study defines different intermediate trajectories during heart development

and potentially new intermediate populations leading to atrial and ventricular cardiomyocyte differentiation. Moreover, new genes that may provide useful markers for different cardiomyocyte populations are identified. Importantly, the data can be used as a resource for comparison to other data sets, such as organoids, to define differentiate states. The study examines the effects of exogenous RA signaling, determining the doses used inhibit ventricular differentiation, cell cycle, and metabolism without affecting earlier progenitor field specification.
Overall, the study provides a detailed examination of cell populations in murine heart development, that can be used to define differentiation states, trajectories, and potentially new cell populations, as well as how these cell populations are affected by aberrant signaling leading to congenital heart defects.

Comments for the author

The manuscript from Gonzalez, Schrode et al, examined cardiac progenitor differentiation and the effects of excess retinoic acid (RA) signaling in mouse embryos. The authors performed single cell analysis of dissected portions of Foxa2-reporter embryos containing the heart and adjacent tissues at multiple stages. Through bioinformatic analyses of the single cell data, the authors are able to define different trajectories of the cardiac progenitor populations, identifying that posterior second heart field primarily gives rise to sinus venosus and atrial cardiomyocytes, while anterior second heart field gives rise to ventricular cardiomyocytes.
They are able to use their data to cross examine organoid populations from mouse and human cells, providing more detail insights into the differentiation states of the cells in the organoids. New markers and proposed drivers of differentiation for atrial and ventricular cardiomyocytes are picked out. The authors then evaluate teratogenic effects of exogenous RA on the cardiac progenitors and differentiated cardiomyocytes via injecting dams with low concentrations of RA. They find the concentration used impairs ventricular/outflow tract differentiation without perturbing posterior/atrial populations and affects cell cycle and metabolic gene signatures with the ventricular cells. Coupled with the ventricular defects, they find that RA induces an expansion of the pharyngeal mesoderm population.
Overall, the authors add to the growing number of studies employing single cell approaches to examine cardiac progenitor and differentiation states and their developmental trajectories. The information provided in the manuscript will certainly provide a useful resource to the field, allowing for additional comparisons to previous and future single cell data sets, as well as the generation of hypotheses regarding gene function during heart development. The main weaknesses of the paper are that fundamentally it is largely descriptive with minimal verification of key points from their analysis. Despite the abundance of information, their data primarily reinforce existing hypotheses. Additionally, there is concern about the presentation, which is very dense, and the interpretations of gene function, again without validation. Thus, while I am very enthusiastic about the information provided from the study and recognize the importance of the abundance of data generated, there are a number of areas in the present manuscript that I think can be improved.

Major issues:
1. One of the more interesting aspects of the study is cluster 0, the transcriptionally-distinct mesodermal population. This is one key area that there should be some attempt of in vivo validation of the data. Some in vivo assessment of this population and whether it is distinct spatiotemporally, rather than stating these cells will be subject of future studies (page 21) would help the impact of the paper and its utility to the field.
Additionally, as shown in Supplemental Figure 2C, it is not clear why the relative size of this population does not decrease as development progressing? Wouldn't this be an expectation if the cell population is differentiating into atrial and ventricular cardiomyocytes? The differentiated cardiomyocyte populations in the data increase over time, as would be expected. It is also interesting the relative location of this cluster changes somewhat dramatically in different UMAP data, such as Supplementary Figure 2, which I think is the same data. Even though the specific depiction of clusters in the UMAP data can change each time it is run one would think the relationships of clusters should not shift so much.
2. The authors state that they have identified "driver" genes, which are inferred to mean genes driving transitions through the identified differentiation states based on a statement on page 16. Evidence that the genes proposed to be drivers perform this function is not provided through functional validation. While some functional validation as drivers may certainly be beyond the scope of this paper, at this point they shouldn't be referred to as drivers, but as markers.

Additionally, more rationale and explanation for the focus on the select genes needs be provided. Information on the putative function of the gene products (transcription factors, enzymes, epigenetic modifiers, etc.) they have selected based on their expression patterns is not included in the text. Examination of mentioned genes from searches makes it unclear to this reader how the genes selected would be drivers of transition states. There is not even speculation in the text from studies in other contexts, organisms or available mouse KOs in the literature about functions in the heart. The full names of the genes should be included prior to the acronym, as many are not common gene names.
3. With respect to the effects of RA signaling, an implication of the treatments and effects on cardiomyocyte populations is the loss of ventricular cardiomyocytes and expansion of pharyngeal mesoderm. This is another area that could have been validated experimentally with immunohistochemistry or in situ hybridization.
However, one would think that this transition between mesodermal derivatives would be visible in the trajectory data (Figure 7A)? It does not appear to be the case based on the trajectories shown. Why is this?
Can the authors subcluster the relevant clusters to better examine the mesodermal trajectories with RA perturbation compared to controls?
4. A number of the observations seem to corroborate previous studies. However, many of these previous studies are not referenced. Examples include recent studies of Hoxb1 gene and Aldh1a2 expression in cardiac progenitors from the Kelly and Zaffran labs (Stefanovic et al, 2020) and studies of Foxf1 localization from the Moskowitz lab. In particular the former study shows these genes mark the posterior populations, which again is corroborate in the present study.
5. Similar to point 4, the authors do not quite fully address previous studies of the effects of exogenous or increased RA signaling on vertebrate heart development. The present study does extend previous observations and provide more detail of the specific molecular effects on ventricular cardiomyocyte differentiation. However, the overt defects shown here with the current doses produce a deficit of ventricular cells without perturbation of posterior/atrial cells. These results are remarkably similar to what was reported for zebrafish Cyp26a1 mutants (Rydeen et al, 2016), as well as with increased RA signaling with different doses of RA treatment in zebrafish (D'Aniello et al, 2013). Additionally, the comparisons provided in the discussion are primarily with respect to loss of RA signaling. Studies such as those from the Aldh1a2 and RXRa KO mice demonstrate there is premature differentiation of ventricular cardiomyocytes in their hearts which is overtly the opposite of defects reported here and a point not clearly articulated. Thus, although reviews covering RA signaling in the heart are referenced, the discussion of the effects of increased/exogenous RA signaling on vertebrate hearts should be covered in more detail and with specific citations.

Minor issues:
1. In many places in the manuscript the authors refer to numbers, which are the clusters. While it became apparent after going through the manuscript they were clusters, in many places they are not initially defined or referred to as the clusters when discussing the populations. Specifically defining the clusters the authors are discussing throughout the manuscript, such as by adding a "C" or some other designation in the text would help the reader.
2. It is difficult to follow the cell type of each of the clusters throughout the other figures as it is only indicated in Figure 1 for the first data set. Indicating what these populations are in other figures, if possible would be helpful.
3. Many of the colors for the clusters in the UMAPs are very difficult to tell apart, particularly when they are right next to each other in the UMAP such as clusters 9 and 10. While it may be a reflection of distortion when the figures are uploaded for review, if the colors can be modified to better distinguish them it would be helpful to the reader.
4. In many figures, there is very small text that is difficult to read. Again, this may be a consequence of lower resolution images uploaded for review. However, the names would still very small and hard to read even if they weren't so pixelated . Minimally, the authors should make sure all the names with small print are not pixelated and are legible.
5. Although the data in Supplemental Figure 2C is meant to show the contributions to the data from the different time points/conditions. It would be helpful if there was a UMAP provided for the data in Figure 1 as well as Figure 6 where just the cells from the different conditions/time points were indicated. The use of just 3 or 6 different colors would allow one to visualize their respective contributions in the same UMAP as shown with all the clusters.

6. It is stated for the first data set that there are 26 clusters (page 7). However, there appear to actually be 27 for the first data set: 0 through 26.
7. Figure 1B cluster 9 is labeled improperly. Cluster 9 is where cluster 24, which is missing, should be labeled.
8. In Figure 4, it would be helpful if the color scheme between A and C matched. They appear to be reversed for the clusters between the UMAP and the graphs.
9. In many of the Figures one can see outlines/shadows of the individual panels. Maybe this reflects something in the conversion of the figures when submitting?
10. In Figure 7, the color designations for the different populations (Control and RA) in the graphs are indicated in panel C. However, it is used first in B. This label should be indicated earlier so the reader can understand the colors indicate in the graphs when they are first presented.
11. There is not real mention of pacemaker/sinoatrial node cardiomyocytes from the data. The only indication is in the GO analysis in Figure 1G. Are all sinus venosus cells really pacemaker cardiomyocytes and expressing canonical pacemaker cardiomyocyte markers or are pacemaker cardiomyocytes a subcluster found within the sinus venosus or even atrial populations?
12. Typo – Figure 2A/B/C should be 2A-C.
13. Typo – page 14 “represente”
14. Figure 3F – part of the “S” for the upper graph is cut off. Also, u, S, s, and t are not defined from what I can find in the legend.
15. As a general thought, how useful are some of the terms from GO analysis in assessing populations? For instance terms that would connote some chamber specific identity, in particular “ventricular cardiac muscle tissue morphogenesis,” are found associated with virtually every cluster, including the atrial-posterior second heart field populations.
16. There is a black dot in Figure 6E.

Reviewer 2

Advance summary and potential significance to field

In the manuscript by Gonzalez et al, the authors present a thorough, single-cell transcriptomic interrogation into the early developmental paths cells take in the formation of chamber-specific cardiomyocytes and identified key retinoic acid-dependent processes in cardiomyocyte formation as a result. Overall, this paper is of high interest, high quality, and advances our understanding of early cardiac chamber specification and the processes giving rise to congenital heart disease to the human population.

The authors’ work presents the following important insights:
Figure 1. Single cell transcriptional profiling can distinguish many known cell types and predict many intermediates of early heart development in vivo.
Figure 2. The single cell profiling in this manuscript identifies many similar cell states reported in in vitro systems of cardiac differentiation.
Figure 3. RNA velocity analysis identifies intermediate states of differentiation of different progenitor pools.
Figure 4. This data identifies many known and novel markers distinguishing the anterior and posterior SHF progenitor pools and identifies state changes associated with intermediary cell states during cardiomyocyte differentiation.
Figure 5. Origin of cells appear to have little effect on final cell state, i.e., a ventricular cardiomyocyte looks like any other ventricular cardiomyocyte, but the intermediate cell states taken in development of that final state can vary widely.
Figure 6. Mild exogenous retinoic acid treatment has limited changes to the heart, although there is a specific switch of anterior SHF cells from ventricular cardiomyocyte to pharyngeal fates.
Figure 7. Exogenous retinoic acid signaling specifically blocks anterior SHF cells during intermediate stages of cardiomyocyte differentiation and appear to limit fate potential.

Comments for the author

In the manuscript by Gonzalez et al, the authors present a thorough, single-cell transcriptomic interrogation into the early developmental paths cells take in the formation of chamber-specific cardiomyocytes and identified key retinoic acid-dependent processes in cardiomyocyte formation

as a result. Overall, this paper is of high interest, high quality, and advances our understanding of early cardiac chamber specification and the processes giving rise to congenital heart disease to the human population.
The authors' work presents the following important insights:
Figure 1. Single cell transcriptional profiling can distinguish many known cell types and predict many intermediates of early heart development in vivo.
Figure 2. The single cell profiling in this manuscript identifies many similar cell states reported in in vitro systems of cardiac differentiation.
Figure 3. RNA velocity analysis identifies intermediate states of differentiation of different progenitor pools.
Figure 4. This data identifies many known and novel markers distinguishing the anterior and posterior SHF progenitor pools and identifies state changes associated with intermediary cell states during cardiomyocyte differentiation.
Figure 5. Origin of cells appear to have little effect on final cell state, i.e., a ventricular cardiomyocyte looks like any other ventricular cardiomyocyte, but the intermediate cell states taken in development of that final state can vary widely.
Figure 6. Mild exogenous retinoic acid treatment has limited changes to the heart, although there is a specific switch of anterior SHF cells from ventricular cardiomyocyte to pharyngeal fates.
Figure 7. Exogenous retinoic acid signaling specifically blocks anterior SHF cells during intermediate stages of cardiomyocyte differentiation and appear to limit fate potential.

Comments:
1. The manuscript starts off with the goal of identifying what determines atrial versus ventricular chamber specification. The authors describe multiple known markers and present new markers throughout the manuscript. Using the myriad of progenitor and intermediate stages eloquently described in Figures 2-5, the authors identify and propose gene regulatory networks that are driving these atrial versus ventricular specification decisions. This is very interesting and insightful.
2. The RA-treated embryos appear to have a block in differentiation towards ventricular cardiomyocyte fate with a switch in fate towards pharyngeal fate. Please comment n whether there is any indication that these blocked cells appear "normal" relative to the other pharyngeal cells?
3. Do Cluster 0 and 15 have similarities to cardiomyocytes and cardiac progenitors derived from mESC differentiation described originally in Kattman et al. (2011, Cell Stem Cell)?
4. Left/right ventricular identity was mentioned throughout the manuscript with markers defining each. It is intriguing that the atrial cells appear as a single homogenous pool. Upon closer interrogation, do the atrial data distinguish left/right identity at these stages like the ventricle or does the atrium appear un-sided at these stages?
5. Minor comment, Figure 4, panels A and C, the colors of 8 and 14 swap between the two panels.

Reviewer 3

Advance summary and potential significance to field

In this manuscript, Gonzalez, Schrode, and their colleagues present a valuable resource: a deep analysis of the transcriptomic heterogeneity present at three important early stages in the development of the mouse heart. The multifaceted nature of their analysis, which includes comparison with organoid data sets, trajectory prediction through RNA velocity analysis, examination of differential gene expression, and gene set enrichment analysis, has high value and is an important addition to the growing body of scRNA-seq work in the embryonic heart. In a particularly interesting application of their strategies for computational analysis of scRNA-seq data, the authors examine the effects of exogenous RA on transcriptome distribution in the developing heart, and their results suggest that exogenous RA may result in the aberrant differentiation of cardiac progenitor cells, shunting them away from the ventricular differentiation pathway and into other differentiation trajectories. While this study stops short of in vivo validation/testing of hypotheses suggested by the scRNA-seq data, it provides an important foundation for a number of potential future studies and will therefore be greatly appreciated by the field.

Comments for the author

A few modifications to the text will enhance the clarity and impact of this manuscript.
1) The authors offer a provocative interpretation of the lack of a "specific population of FHF progenitors" in their data set, arguing that FHF cells closely resemble differentiating cardiomyocytes. Is it also possible that even the earliest timepoint analyzed by the authors is a bit too late to capture the FHF progenitors?
Addition of commentary on this possibility would be helpful.
2) The authors should consider softening their interpretations within the section titled "Identification of a transcriptionally distinct mesoderm population that gives rise to early myocardial intermediates." Generally speaking, computational analysis of scRNA-seq data can't demonstrate a lineage relationship -- this type of analysis can suggest a hypothesis that can be tested in vivo, but it can't show that one cell type gives rise to another one. Discussing this interesting cluster and its unique attributes is warranted, but the authors would need to show where these cells are in vivo and to perform tracking experiments to show their lineage to reach firm conclusions about whether these are "myocardial intermediates" and what they contribute to. Granted, the authors say that "these data suggest..." and "predicted to contribute...", but the title and tone of the section give the impression that we know more about the biography of these cells than we currently do.
3) While the paper is generally clearly written, there are several occasions where it is difficult to link claims made in the text with data provided in the figures or legends. This is probably a consequence of the need to adhere to a reasonable word limit for the text, but it can still leave the reader unable to connect the dots. Just to cite a couple of examples from the first section of the results: The authors state that multiple Pdgfra+ mesoderm populations are more commonly found in S phase than their differentiated Nkx2-5+ progeny, but there are no concrete references to specific clusters or to the percentages of cells that justify the "more commonly found" conclusion -- simple additions to the text could make this much more clear.
In the same paragraph, the authors state that "Isl1+ SHF progenitors had a balanced contribution from all three stages", but the reader cannot easily discern which specific clusters are being considered as "Isl1+ SHF progenitors" in this context. There are other places in the text where it is easier to catch exactly what is referred to in the figures -- if the authors could go through the text and add some further annotations where needed to guide the reader through the data, this would enhance the value of the manuscript.

First revision

Author response to reviewers' comments

Reviewer 1 Advance Summary and Potential Significance to Field:
The manuscript from Gonzalez et al examines development of the heart and surrounding tissues in mouse embryos. The study defines different intermediate trajectories during heart development and potentially new intermediate populations leading to atrial and ventricular cardiomyocyte differentiation. Moreover, new genes that may provide useful markers for different cardiomyocyte populations are identified. Importantly, the data can be used as a resource for comparison to other data sets, such as organoids, to define differentiate states. The study examines the effects of exogenous RA signaling, determining the doses used inhibit ventricular differentiation, cell cycle, and metabolism without affecting earlier progenitor field specification. Overall, the study provides a detailed examination of cell populations in murine heart development, that can be used to define differentiation states, trajectories, and potentially new cell populations, as well as how these cell populations are affected by aberrant signaling leading to congenital heart defects.

Reviewer 1 Comments for the Author:
The manuscript from Gonzalez, Schrode et al, examined cardiac progenitor differentiation and the effects of excess retinoic acid (RA) signaling in mouse embryos. The authors performed single cell analysis of dissected portions of Foxa2-reporter embryos containing the heart and adjacent tissues at multiple stages. Through bioinformatic analyses of the single cell data, the authors are able to

define different trajectories of the cardiac progenitor populations, identifying that posterior second heart field primarily gives rise to sinus venosus and atrial cardiomyocytes, while anterior second heart field gives rise to ventricular cardiomyocytes. They are able to use their data to cross examine organoid populations from mouse and human cells, providing more detail insights into the differentiation states of the cells in the organoids. New markers and proposed drivers of differentiation for atrial and ventricular cardiomyocytes are picked out. The authors then evaluate teratogenic effects of exogenous RA on the cardiac progenitors and differentiated cardiomyocytes via injecting dams with low concentrations of RA. They find the concentration used impairs ventricular/outflow tract differentiation without perturbing posterior/atrial populations and affects cell cycle and metabolic gene signatures with the ventricular cells. Coupled with the ventricular defects, they find that RA induces an expansion of the pharyngeal mesoderm population.

Overall, the authors add to the growing number of studies employing single cell approaches to examine cardiac progenitor and differentiation states and their developmental trajectories. The information provided in the manuscript will certainly provide a useful resource to the field, allowing for additional comparisons to previous and future single cell data sets, as well as the generation of hypotheses regarding gene function during heart development. The main weaknesses of the paper are that fundamentally it is largely descriptive with minimal verification of key points from their analysis. Despite the abundance of information, their data primarily reinforce existing hypotheses. Additionally, there is concern about the presentation, which is very dense, and the interpretations of gene function, again without validation. Thus, while I am very enthusiastic about the information provided from the study and recognize the importance of the abundance of data generated, there are a number of areas in the present manuscript that I think can be improved.

Major issues:

1. One of the more interesting aspects of the study is cluster 0, the transcriptionally-distinct mesodermal population. This is one key area that there should be some attempt of in vivo validation of the data. Some in vivo assessment of this population and whether it is distinct spatiotemporally, rather than stating these cells will be subject of future studies (page 21) would help the impact of the paper and its utility to the field. Additionally, as shown in Supplemental Figure 2C, it is not clear why the relative size of this population does not decrease as development progressing? Wouldn't this be an expectation if the cell population is differentiating into atrial and ventricular cardiomyocytes? The differentiated cardiomyocyte populations in the data increase over time, as would be expected. It is also interesting the relative location of this cluster changes somewhat dramatically in different UMAP data, such as Supplementary Figure 2, which I think is the same data. Even though the specific depiction of clusters in the UMAP data can change each time it is run, one would think the relationships of clusters should not shift so much.

We thank the Reviewer for their comments and agree that this part of our manuscript needs improved clarity. Cluster 0 refers to a cardiomyocyte population (these cells were previously labeled as myocardial progenitor cell, this nomenclature has been edited for clarity in the revised manuscript) that expresses *Tnnt2* and is most transcriptionally similar to the atrial and ventricular clusters (C4 and C2/C17 respectively) though it appears to be less mature, based on the fact that it is not enriched for energetic processes (electron transport) and a response to hypertrophy, but instead is enriched for biosynthetic processes. Like the atrial and ventricular cardiomyocyte clusters, it is composed primarily of cells from the PHT and HT stage (Figure 1F and Supplementary Figure 4A). Cluster 15, which the Reviewer may be referencing, is a population that we identified as a transcriptionally distinct mesoderm population and was annotated as such based on shared expression of lateral plate mesoderm markers (*Pmp22*, *Hand1* among others) but this population also expressed low levels of *Nkx2-5*. The relative size of this population does indeed change over the course of development and is quantified in Figure 1F where the majority of the population comes from the CC stage, with smaller contributions from the PHT and HT stages. Supplemental Figure 2D visualizes contribution to cell clusters that are composed of the full data set (all cells sequenced) rather than just the subclustered cardiac data that is shown in Figure 1. In this UMAP projection, we do not observe these transcriptionally distinct mesoderm cells clustering separately from other cardiac cell types – this is likely due to shared transcriptional similarity with other cardiac progenitors on the background of a more transcriptionally diverse PCA space in the full

data set than that of the cardiac subset.

We share the Reviewers intrigue about the clustering location of C0 (the cardiomyocyte population) and C15 (the transcriptionally-distinct mesoderm population) in our cardiac subset population. In order to further determine the source of this, we wanted to ensure that the UMAP localization of C0 and C15 was not due to technical reasons. We plotted the nFeature_RNA (# of genes) and the nCount_RNA (# transcripts) for each cluster, and found that C0 and C15 had comparatively low levels of RNA expression. This lower expression level was also shared by other clusters in proximity to C0/15, including the endothelial populations (C9,C10,C13). While we did originally filter the data for cells with low expression (cells with less than 200 genes identified were excluded, in line with recommended pipelines for scRNAseq analysis in Seurat) and scaled our data prior to clustering to control for differences in RNA abundance, we did not specifically filter out cells for a particular number of transcripts. In order to determine if C0 and C15 would have been removed by filtering for lowly expressed transcripts, we filtered our data for cells below 2000 transcripts and 5000 transcripts and found that these populations persisted even at these more stringent cutoffs. Furthermore, when the filtered cardiac subset was reclustered, we found that cells within cluster 0 and 15 did not integrate together with other clusters expressing similar markers, suggesting that these still represent a biological distinct population. However, after filtering these cells now clustered further away from the endothelial/endocardial populations and were located closer to the myocardial intermediates, suggesting that the location in the UMAP projection may have partially been driven by differences in the number of transcripts (**Figure 1 for Reviewer**).

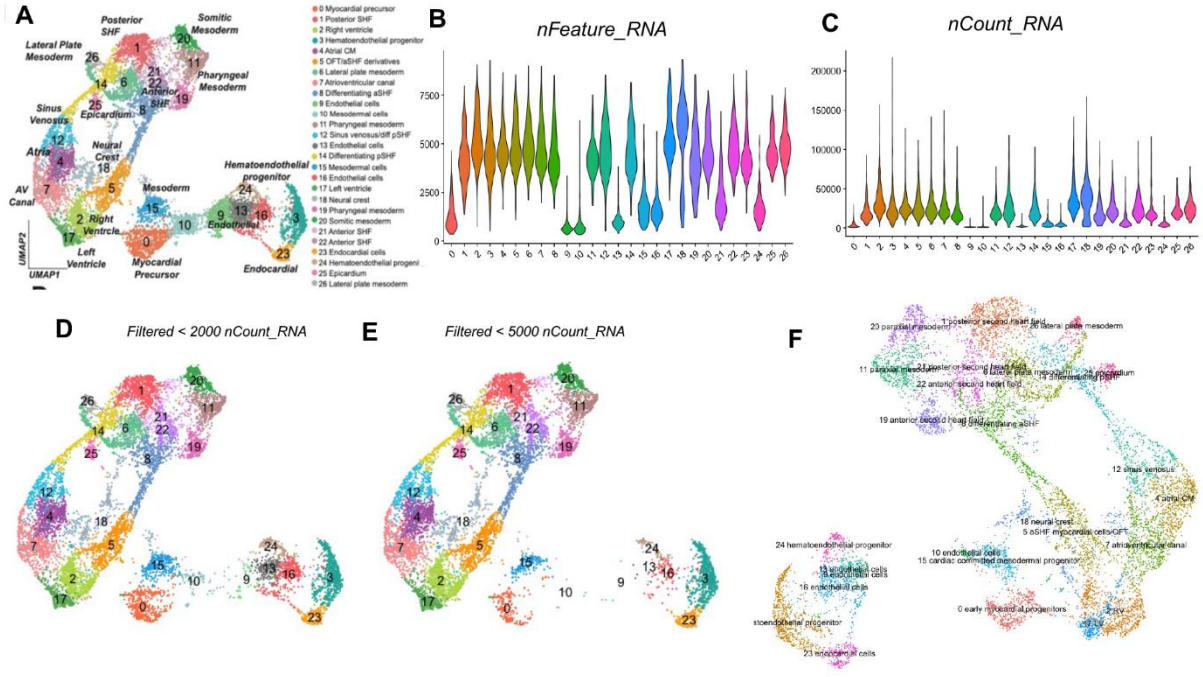


Figure 1 for Reviewer. A) UMAP of cardiac subset of interest. **B)** ViolinPlot showing the number of genes (Features) identified for each cluster in cardiac subset. **C)** ViolinPlot showing the number of RNA transcripts identified for each cluster in the cardiac subset data. **D)** UMAP of cardiac data after filtering out cells with less than 2000 transcripts identified. **E)** UMAP of cardiac data after filtering out cells with less than 5000 transcripts identified. **F)** UMAP of reclustered cardiac data, labeled with previous cell annotation. Cells with less than 5000 transcripts were filtered out as in Figure 1E, and then reclustered and a UMAP projection re-run using the same parameters as those used previously.

As the Reviewer points out, we have yet to validate this population in the embryo to determine its exact location/relationship to other mesoderm cells in the case of C15, and cardiomyocytes for C0. Due to the high overlap of markers within these and other clusters validation will likely comprise complex multi-marker analyses and ultimately functional validation and lineage-tracing studies. While we do maintain that there are meaningful biological differences in these populations, such as

differences in regulation of genes such as *Nkx2-5* and enrichment for biosynthetic processes that suggested a precursor state, the in depth validation may be beyond the scope of this paper. We have however strongly softened the language on this population in the revised manuscript and removed some of the more speculative statements in light of the absence of *in vivo* validation.

2. The authors state that they have identified "driver" genes, which are inferred to mean genes driving transitions through the identified differentiation states based on a statement on page 16. Evidence that the genes proposed to be drivers perform this function is not provided through functional validation. While some functional validation as drivers may certainly be beyond the scope of this paper, at this point they shouldn't be referred to as drivers, but as markers. Additionally, more rationale and explanation for the focus on the select genes needs be provided. Information on the putative function of the gene products (transcription factors, enzymes, epigenetic modifiers, etc.) they have selected based on their expression patterns is not included in the text. Examination of mentioned genes from searches makes it unclear to this reader how the genes selected would be drivers of transition states. There is not even speculation in the text from studies in other contexts, organisms or available mouse KOs in the literature about functions in the heart. The full names of the genes should be included prior to the acronym, as many are not common gene names.

We agree that it is important to note that these genes cannot be termed drivers without proper functional validation. We have thus modified the language in this section to better reflect our findings, and referred to these as "highly dynamically regulated genes" or in rare cases as "putative drivers" in the context that the high dynamic changes in these genes makes them candidates of interest for functional follow up. We explicitly state in the text that such experiments would be necessary for this claim.
Additionally, we have modified the text to make clear that the candidates selected are candidates that were among the top dynamically regulated genes for either a particular cluster (eg *Nebl,* is the #1 top dynamically regulated gene having highest velocity for cluster 8). For the selected candidates shown across the LPM/pSHF and aSHF differentiation stream we chose candidates that were within the list of top 300 dynamically regulated genes, and focused on candidates that had either a) high ranking on these lists, b) an unknown role in cardiac development, c) a link to heart defects in GWAS and/or animal model studies, d) cell type specific regulation dynamics or in many cases a combination of the above. We have modified the resulting section to make this more clear and have copied the section below for the Reviewer's convenience.

> *Modified section #1: "We next plotted the splicing dynamics and expression over time of the most highly dynamically regulated genes for cluster 8 and cluster 14, reasoning that these might represent genes governing early transition from the LPM/heart field progenitors towards cardiomyocytes. Cap2 a gene implicated in cardiac conduction and sudden cardiac death was among the top dynamically regulated candidates in cluster 14 (Field et al., 2015; Peche et al., 2013). Cap2showed rapid upregulation compared to cycling pSHF cells, suggesting that Cap2 may play an important early role in development, consistent with its role as a regulator of actin dynamics (**Figure 4D**). Ccd141 was also among the most highly dynamically regulated genes in cluster 14 and has been identified as a loci implicated in heart rate regulation (van den Berg et al., 2017) but no functional studies during heart development have been reported to date. Nebl and Nexn, two genes encoding components of the Z-disk were among the top 5 most dynamically regulated genes for cluster 8 and have been associated with familial cardiomyopathies (**Figure 4E**). Similar to many candidates identified here, their role during development has not yet been studied; their upregulation at the earliest stage of the differentiation stream suggests that these may play a key role early on. We also identified dynamic regulation of Mest, which is known to be expressed in the trabeculae during development but otherwise has no known function in heart development (King et al., 2002). Mest was downregulated in cluster 14 but rapidly upregulated in the aSHF derivative population (C5), perhaps indicating a transient requirement for this gene, or context specific roles at different stages of development. These examples highlight interesting patterns not only in gene expression levels, but also in the dynamic changes in the regulation of their expression within individual populations and along differentiation trajectories. ."*

> *Modified section #2: "By plotting the splicing dynamics and expression over time of*

individual candidates from this list of 300 genes, we were able to better understand how these patterns varied for individual populations. We found that genes such as Vsnl1 were most strongly upregulated in cell types found at the earlier/intermediary stages of differentiation (C12, C14) and was later downregulated in atrial/AVC cells ***(Figure 3H)****. In contrast, Ehz2, which was also highly dynamically regulated was highly expressed at earlier stages but became downregulated later during differentiation. This is consistent with previous studies demonstrating a role for Ezh2 in repressing the cardiac progenitor genes in differentiated myocardial cells (Delgado-Olguín et al., 2012).*

In addition to these examples we identified many other candidates with previously unrevealed cell type-specific regulation across the LPM/pSHF and aSHF differentiation streams. A number of the top dynamically regulated genes identified were specific to one differentiation stream, and had no previously identified role in heart development. These included candidates such as Prtg which plays roles in neurons and the developing tooth germ (Takahashi et al., 2010; Wong et al., 2010) but has not been implicated in cardiac development. Prtg was strongly upregulated in differentiating aSHF cell types but was lowly expressed in differentiated myocardial cells, similar to another dynamically regulated gene Meis2. Meis2 has been implicated in cardiac and facial defects and is essential for cranial and cardiac neural crest development (Louw et al., 2015; Machon et al., 2015), but its expression pattern within the differentiating aSHF has not previously been recognized, and raises the question of whether observed defects in patients are purely due to a neural crest etiology ***(Figure 3K)****. Other dynamically regulated candidates that showed cell type specific regulation was Smyd1 which was strongly upregulated in a subset of the differentiated aSHF derived cells (C5) (****Supplementary Figure 6D****). Smyd1 plays a well-established role in orchestrating early heart development, particularly in the OFT/SHF (Franklin et al., 2016; Rasmussen et al., 2015; Wang et al., 2021), and the heterogenous upregulation of this gene within the aSHF derivatives raises interesting questions about whether this occurs in a more granular fashion than previously appreciate. Further studies of these and other candidates identified here may shed better light on the mechanistic consequences of these genes in a cell-type specific context, as well as at which stages of differentiation these genes may be acting."*

3. With respect to the effects of RA signaling, an implication of the treatments and effects on cardiomyocyte populations is the loss of ventricular cardiomyocytes and expansion of pharyngeal mesoderm. This is another area that could have been validated experimentally with immunohistochemistry or in situ hybridization. However, one would think that this transition between mesodermal derivatives would be visible in the trajectory data (Figure 7A)? It does not appear to be the case based on the trajectories shown. Why is this? Can the authors subcluster the relevant clusters to better examine the mesodermal trajectories with RA perturbation compared to controls?

We thank the Reviewer for discussing this aspect of our findings. The RNA velocity map in 7A is a summation of individual vectors for each cell shown, and when multiple overlapping trajectories occur simultaneously it is sometimes difficult to appreciate the relationships between these cells. We have taken the Reviewers very helpful suggestion to recluster the cells from the aSHF and its derivates, namely the ventricular cardiomycoytes, as well as the pharyngeal/mesenchymal cell types identified and recalculated the RNA velocity for the simplified population in **Figure 7C-H** (copied here for the Reviewer). Additionally, to better visualize related populations and branching points between these cells, we have employed an additional computational trajectory tool (URD), which allows the user to select the beginning population (the aSHF in our case) and the ending populations (ventricular cardiomyocytes, pharyngeal/mesenchymal populations). This projection better shows the mesodermal trajectories, where we observe a block of differentiation toward the ventricular lineage (as evidenced by decreased density of RA (teal) treated cells along this branch), and instead a preponderance of this population in a neighboring branch belonging to pharyngeal/mesodermal cells.

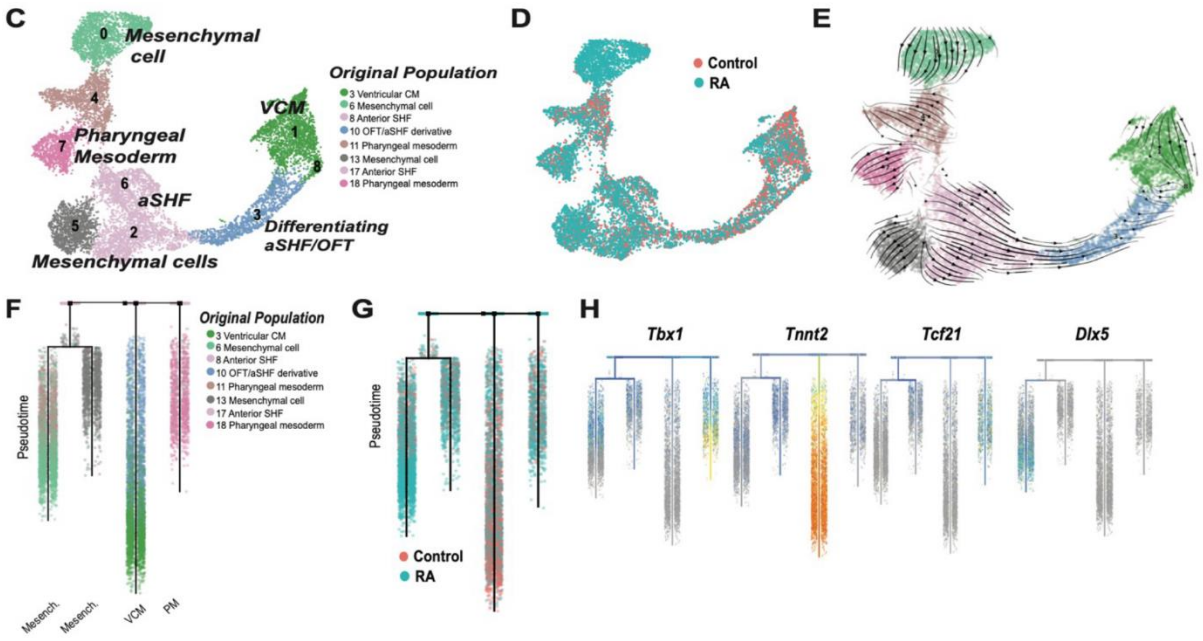


Figure 7C-H. C) Subclustering of aSHF cell types and derivatives. **D)** Contribution of Control and RA-treated embryos to UMAP projection. E) RNA velocity map of aSHF derivative UMAP projection. **F-G**) URD trajectories demonstrating cell state transitions, colored by cell type (F) and condition (G). **H)** Expression of characteristic markers across URD trajectory.

4. A number of the observations seem to corroborate previous studies. However, many of these previous studies are not referenced. Examples include recent studies of Hoxb1 gene and Aldh1a2 expression in cardiac progenitors from the Kelly and Zaffran labs (Stefanovic et al, 2020) and studies of Foxf1 localization from the Moskowitz lab. In particular the former study shows these genes mark the posterior populations, which again is corroborate in the present study.

We apologize for this oversight and thank the Reviewer for adding this important information. We have revised the relevant sections accordingly to include references to the studies indicated, which do indeed provide additional support to the findings in our work.

5. Similar to point 4, the authors do not quite fully address previous studies of the effects of exogenous or increased RA signaling on vertebrate heart development. The present study does extend previous observations and provide more detail of the specific molecular effects on ventricular cardiomyocyte differentiation. However, the overt defects shown here with the current doses produce a deficit of ventricular cells without perturbation of posterior/atrial cells. These results are remarkably similar to what was reported for zebrafish Cyp26a1 mutants (Rydeen et al, 2016), as well as with increased RA signaling with different doses of RA treatment in zebrafish (D'Aniello et al, 2013). Additionally, the comparisons provided in the discussion are primarily with respect to loss of RA signaling. Studies such as those from the Aldh1a2 and RXRa KO mice demonstrate there is premature differentiation of ventricular cardiomyocytes in their hearts, which is overtly the opposite of defects reported here and a point not clearly articulated. Thus, although reviews covering RA signaling in the heart are referenced, the discussion of the effects of increased/exogenous RA signaling on vertebrate hearts should be covered in more detail and with specific citations.

We fully agree and have modified both the discussion and results section with a more informed discussion of our findings in the context of other studies in our field, particularly the ones highlighted by the Reviewer. The specific sections of the discussion are copied below for the Reviewer's convenience:

> *"The use of scRNAseq data to understand cell-type specific effects of genetic defects has been explored previously (Kathiriya et al., 2021; de Soysa et al., 2019). We extended this to an in utero exposure model of RA, as RA has been well-characterized as a teratogen*

(Lammer et al., 1985; Piersma et al., 2017). The hypomorphic ventricular phenotype observed in RA-exposed embryos are consistent with classical studies in the chick (Osmond et al., 1991; Yutzey et al., 1994), zebrafish (Stainier and Fishman, 1992), and mouse (Xavier-Neto et al., 1999), demonstrating that exogenous RA results in changes to atrial and ventricular chamber size in a concentration-dependent fashion (Bernheim and Meilhac, 2020; Perl and Waxman, 2019; Waxman and Yelon, 2009). Further studies showing that depletion of retinoic acid receptors results in a low increase in RA concentration and increased the number of atrial CMs, while combining a Cyp26a1 depletion with it results in an intermediate increase in RA concentration, with reduced ventricular CMs and no effect on atrial cells (D'Aniello et al., 2013). The work presented here builds on these studies and aims to determine the transcriptional effects induced directly by RA signaling in a cell-type specific manner (Hochgreb et al., 2003; Lin et al., 2010; Perl and Waxman, 2020). We demonstrate that in utero exposure to RA at E7.5 did not affect the distribution of heart field progenitors between normal and RA-exposed conditions, which was interesting as previous work has shown a role for RA in restriction of the cardiac progenitor pool (Duong et al., 2021; Keegan et al., 2004, 2005). This may be because we elected to inject at a timepoint following migration of the LPM, in an effort to focus on the effect RA on cardiac mesodermal cell differentiation and expansion. Future work should examine whether earlier injection of RA at similar concentrations might recapitulate previously observed defects in progenitor size, and determine if this occurs in atrial/ventricular specific manner.

Our observation that aSHF differentiation towards a myocardial fate was negatively affected by RA exposure is consistent with previous work demonstrating posterior expansion of Fgf8+ and Isl1+ SHF progenitors and defects in OFT septation in RA-deficient mouse mutants. Antagonism of Fgf8 and RA has been well-characterized in the literature in various contexts, placing RA at the top of important signaling cascade driving growth and differentiation of the developing heart (del Corral et al., 2003; Cunningham et al., 2013; Pasini et al., 2012). Interestingly, recent work has indicated that RA-deficient zebrafish embryos display defects in differentiation of SHF-derived cardiomyocytes. It is unclear if this represents differences across species, or perhaps concentration differences between KO studies and our low-dose RA exposure model. The increase in processes involved in head morphogenesis is also consistent with the role of RA as a teratogen, and a cause for craniofacial malformations and microcephaly (Petrelli et al., 2019). Future work should determine whether an RA concentration-dependent mechanism fine tunes control of SHF progenitor expansion/differentiation across various lineages, or if the increase in pharyngeal mesoderm specification is an otherwise aberrant differentiation event driven by restriction of cardiomyocyte fate."

Minor issues:

1. In many places in the manuscript the authors refer to numbers, which are the clusters. While it became apparent after going through the manuscript they were clusters, in many places they are not initially defined or referred to as the clusters when discussing the populations. Specifically defining the clusters the authors are discussing throughout the manuscript, such as by adding a "C" or some other designation in the text would help the reader.

We agree entirely and have modified the manuscript text to reflect this comment, with each cluster indicated by C1, C2 etc whenever referring to cluster numbers.

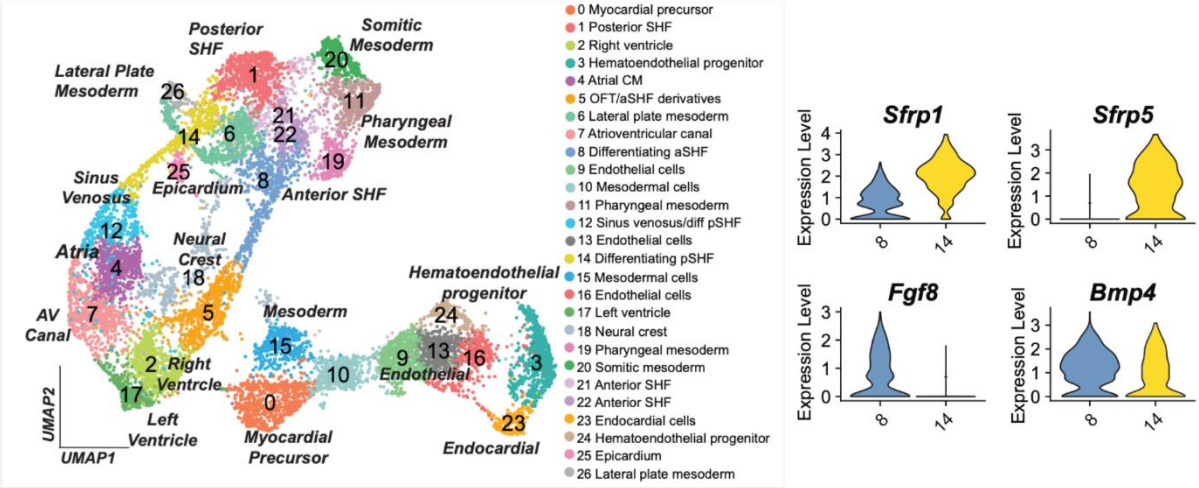
2. It is difficult to follow the cell type of each of the clusters throughout the other figures as it is only indicated in Figure 1 for the first data set. Indicating what these populations are in other figures, if possible would be helpful.

We have added the cluster identification to the UMAP in subsequent figures wherever possible, and annotated clusters of interest with their names in schematics wherever it was not possible to also include the full UMAP annotation.

3. Many of the colors for the clusters in the UMAPs are very difficult to tell apart, particularly when they are right next to each other in the UMAP such as clusters 9 and 10. While it may be a reflection of distortion when the figures are uploaded for review, if the colors can be modified to

better distinguish them it would be helpful to the reader.

We were indeed not satisfied with the initial color scheme either and have taken the opportunity in our revised manuscript, to reassign the colors of our UMAP plots to a custom color palette in order to better distinguish the clusters from one another. These also now match the colors of all UMAPs used in the subsequent figures and are used when highlighting single populations in our schematics, such as those found in Figure 4A/C. An example of the new coloring scheme and its related use for other figure panels is included below:



4. In many figures, there is very small text that is difficult to read. Again, this may be a consequence of lower resolution images uploaded for review. However, the names would still very small and hard to read even if they weren't so pixelated. Minimally, the authors should make sure all the names with small print are not pixelated and are legible.

We have taken care to ensure that the figures, in particular the smaller text has been enlarged and ensured the submitted figures also be of sufficient quality to read the smaller legends.

5. Although the data in Supplemental Figure 2C is meant to show the contributions to the data from the different time points/conditions. It would be helpful if there was a UMAP provided for the data in Figure 1 as well as Figure 6 where just the cells from the different conditions/time points were indicated. The use of just 3 or 6 different colors would allow one to visualize their respective contributions in the same UMAP as shown with all the clusters.

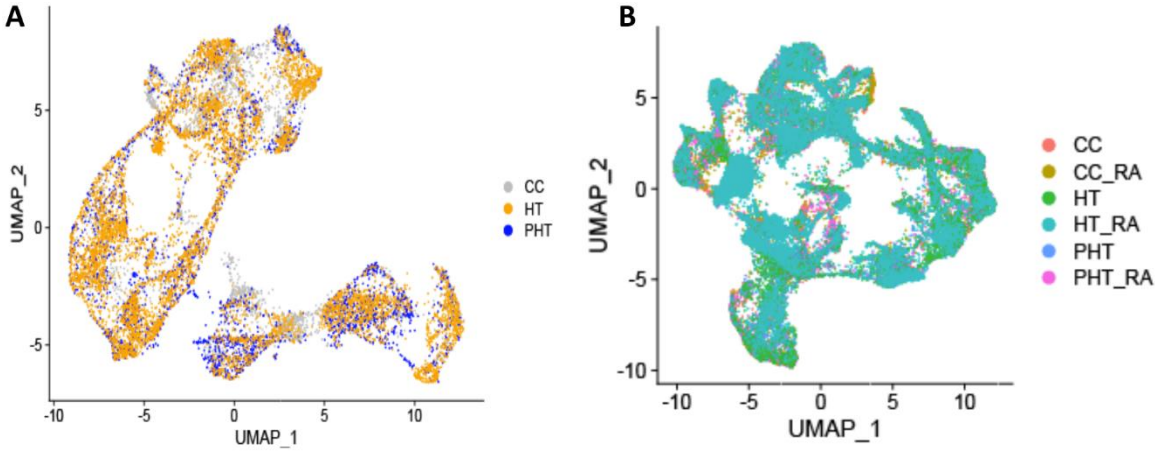


Figure 2 for Reviewer. UMAP showing distribution of cells from each timepoint sequenced in **A)** cardiac subset for normal (untreated) samples **B** cardiac subset of normal and RA-exposed samples.

Because of the density of cells in our data, it is sometimes difficult to discern cell types that are overlayed on top of each other in a satisfactory manner on UMAP plots, as overlay are not transparent and many cells will be hidden under cells of the top layer (For example see **Figure 2 for Reviewer**). We also recovered different numbers of cells for each stage sequenced (eg. 21,318 cells for the HT stage and 14,355 cells for the CC stage), which is also why we have elected to rely on the quantification of the normalized contribution to understand the contribution of individual samples. We have also included the split diagrams in the associated supplemental figures, as these avoid the overcrowding issue observed with the overlay on the UMAP plots.

6. It is stated for the first data set that there are 26 clusters (page 7). However, there appear to actually be 27 for the first data set: 0 through 26.

Thank you for this correction – we have modified the text to properly reflect the cluster numbers.

7. Figure 1B cluster 9 is labeled improperly. Cluster 9 is where cluster 24, which is missing, should be labeled.

We have modified the figure to properly reflect the cluster numbers.

8. In Figure 4, it would be helpful if the color scheme between A and C matched. They appear to be reversed for the clusters between the UMAP and the graphs.

Thank you for pointing this out – we have modified the color scheme here to correspond with the schematic, which should now match to the original color palette in the UMAP in all of our updated figures.

9. In many of the Figures one can see outlines/shadows of the individual panels. Maybe this reflects something in the conversion of the figures when submitting?
We believe that indeed this may have been due to conversions of the figures when submitting the manuscript. For submission of the revised paper we have included high quality images where this problem should no longer be present.

10. In Figure 7, the color designations for the different populations (Control and RA) in the graphs are indicated in panel C. However, it is used first in B. This label should be indicated earlier so the reader can understand the colors indicate in the graphs when they are first presented.

We agree and have modified the figure such that the legend is used first in Figure 7B.

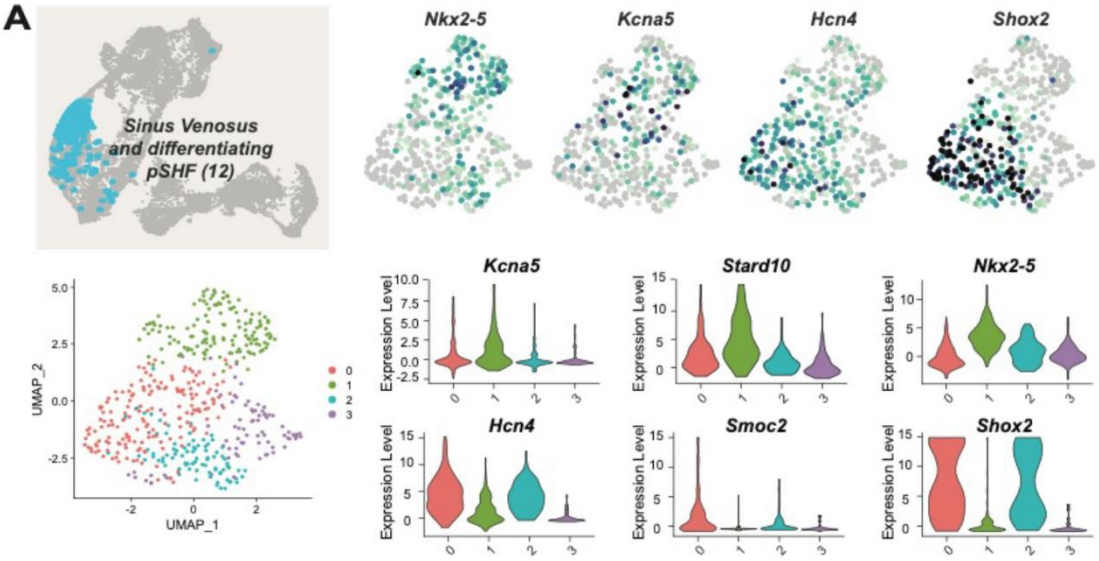
11. There is not real mention of pacemaker/sinoatrial node cardiomyocytes from the data. The only indication is in the GO analysis in Figure 1G. Are all sinus venosus cells really pacemaker

cardiomyocytes and expressing canonical pacemaker cardiomyocyte markers or are pacemaker cardiomyocytes a subcluster found within the sinus venosus or even atrial populations?

We thank the Reviewer for their comment and for bringing up this interesting question which we have addressed with additional analysis in the revised manuscript. When subclustered, the sinus venosus population does indeed reveal the existence of several populations which broadly comprise an *Hcn4+/Smoc2+/Shox2+/Nkx2-5-* population, as well as an *Nkx2-5+/Tnnt2+/Kcna5+* population which we believe represent the sinoatrial nodal pacemaker cells, as well as the atrial cell types that likely may cluster together due to shared proximity in the venous pole of the heart. The shared clustering of these cells together also reflects a similar observation as that seen in cluster 5 of our analysis, where cells of the OFT as well as differentiating cardiomyocytes from the aSHF clustered together. This analysis further illustrates that many of the current 27 clusters contain additional heterogeneity that we and others can explore further in future studies.

We have added the new analysis to Supplementary Figure 5A of our revised manuscript, and have added the following text into the results section to address this question.

"We subclustered the SV cells and found that these cells broadly segregated into a Nkx2-5- population expressing sinoatrial nodal markers such as Smoc2, Hcn4, and Shox2 and an Nkx2- 5+ population that instead expressed markers of atrial myocytes (Figure S5A)".



Supplementary Figure 5A. *Top left:* Schematic of cluster investigated (Cluster 12). Bottom left: UMAP clustering of Cluster 12 alone following re-clustering. *Top right*: FeaturePlots for markers of atrial and sinoatrial nodal identity. Bottom right: Violin plots for markers of atrial and SV lineages across individual clusters.

12. Typo – Figure 2A/B/C should be 2A-C.

Thank you – this has been corrected.

13. Typo – page 14 "represente"

Thank you – this has been corrected.

14. Figure 3F – part of the "S" for the upper graph is cut off. Also, u, S, s, and t are not defined from what I can find in the legend.

The legend has been modified to properly define each term (u, S, s, t) associated with these plots, in addition to properly positioning these such that the "S" is no longer cut off.

15. As a general thought, how useful are some of the terms from GO analysis in assessing populations? For instance terms that would connote some chamber specific identity, in particular "ventricular cardiac muscle tissue morphogenesis," are found associated with virtually every cluster, including the atrial-posterior second heart field populations.

This is true and a valid point. We have revisited this analysis and made an effort to be more selective in which GO terms are included in the data presentation to highlight the populations and processes of interest in the most specific way possible. In some instances GO terms have been called for multiple populations due to shared expression of broadly expressed genes, but we have taken care to minimize this wherever possible.

16. There is a black dot in Figure 6E. Thank you – this has been corrected.

Reviewer 2 Advance Summary and Potential Significance to Field:
In the manuscript by Gonzalez et al, the authors present a thorough, single-cell transcriptomic interrogation into the early developmental paths cells take in the formation of chamber-specific cardiomyocytes and identified key retinoic acid-dependent processes in cardiomyocyte formation as a result. Overall, this paper is of high interest, high quality, and advances our understanding of early cardiac chamber specification and the processes giving rise to congenital heart disease to the human population.

The authors' work presents the following important insights:
Figure 1. Single cell transcriptional profiling can distinguish many known cell types and predict many intermediates of early heart development in vivo.
Figure 2. The single cell profiling in this manuscript identifies many similar cell states reported in in vitro systems of cardiac differentiation.
Figure 3. RNA velocity analysis identifies intermediate states of differentiation of different progenitor pools.
Figure 4. This data identifies many known and novel markers distinguishing the anterior and posterior SHF progenitor pools and identifies state changes associated with intermediary cell states during cardiomyocyte differentiation.
Figure 5. Origin of cells appear to have little effect on final cell state, i.e., a ventricular cardiomyocyte looks like any other ventricular cardiomyocyte, but the intermediate cell states taken in development of that final state can vary widely.
Figure 6. Mild exogenous retinoic acid treatment has limited changes to the heart, although there is a specific switch of anterior SHF cells from ventricular cardiomyocyte to pharyngeal fates.
Figure 7. Exogenous retinoic acid signaling specifically blocks anterior SHF cells during intermediate stages of cardiomyocyte differentiation and appear to limit fate potential.

Comments:
1. The manuscript starts off with the goal of identifying what determines atrial versus ventricular chamber specification. The authors describe multiple known markers and present new markers throughout the manuscript. Using the myriad of progenitor and intermediate stages eloquently described in Figures 2-5, the authors identify and propose gene regulatory networks that are driving these atrial versus ventricular specification decisions. This is very interesting and insightful.

We thank the Reviewer for their enthusiasm on our work and its value to the field – we hope this work will spur additional investigation into these fate decisions and look forward to following up on these questions in the future.

2. The RA-treated embryos appear to have a block in differentiation towards ventricular cardiomyocyte fate with a switch in fate towards pharyngeal fate. Please comment n whether there is any indication that these blocked cells appear "normal" relative to the other pharyngeal cells?

We thank the Reviewer for raising this interesting question, which is one that we are in the process of pursuing further as well. When examining the genes and processes that are upregulated in these cells, the most differentially enriched processes are those involved in cell cycling, as well as biosynthetic processes such as tRNA synthesis, rRNA processing, RNA splicing, and chromatin

assembly. In an effort to adhere to space constraints in the final manuscript and simplify the overall text, we have chosen not to focus on the "blocked" aspect of these cells, as more robust *in vivo* validation experiments might be more appropriate to answer this question in the long term and more conclusively. We are including some data for the Reviewer below which demonstrates the top 10 most highly enriched GO terms for this cluster based on differential expression analysis (**Figure 3 for Reviewer**).

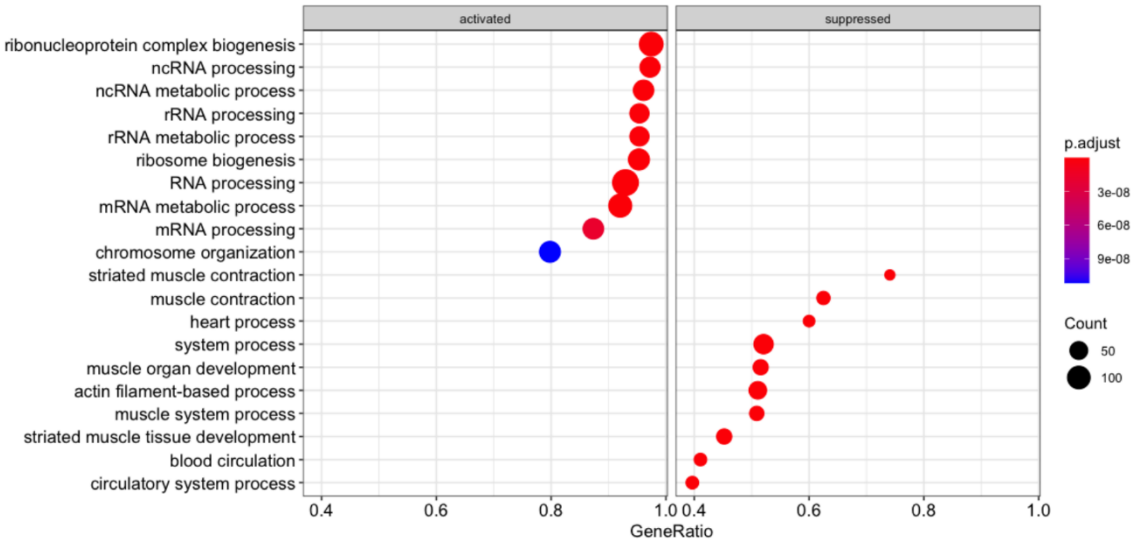


Figure 3 for Reviewer. Top 10 enriched GO terms for mesenchymal cell population (C6). (Left) Upregulated processes ; (Right) Downregulated processes

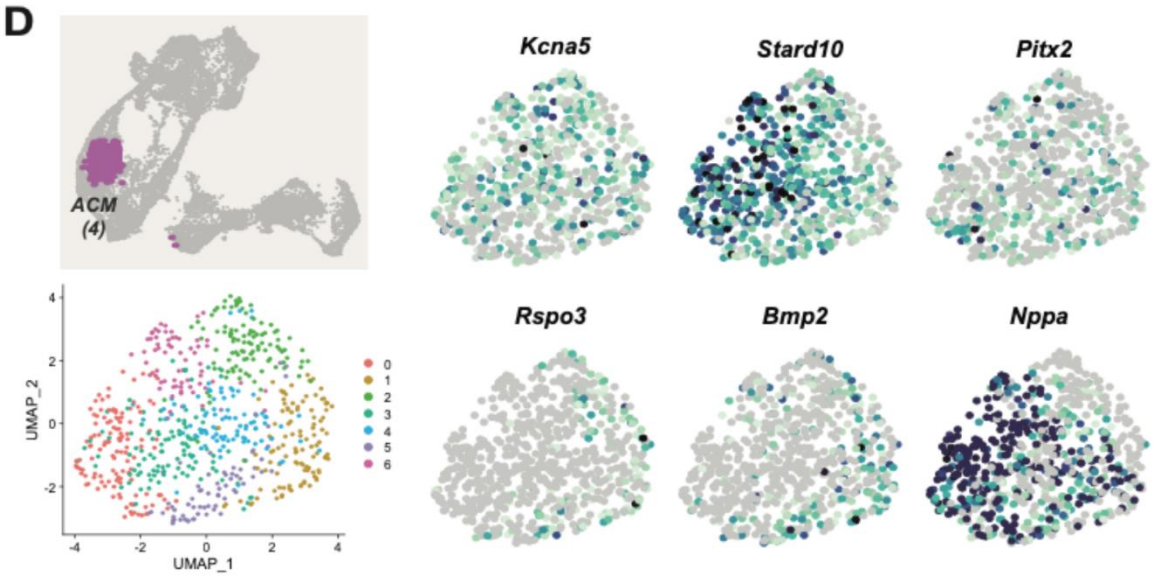
3. Do Cluster 0 and 15 have similarities to cardiomyocytes and cardiac progenitors derived from mESC differentiation described originally in Kattman et al. (2011, Cell Stem Cell)?

This is an interesting point. The cardiac progenitors described in Kattman *et al* 2011 are characterized by co-expression of both *Pdgfra* and Kdr and they represent a mesoderm population specified toward the cardiac lineage. We find that Cluster 0 and 15 demonstrate comparatively low expression of both *Pdgfra* and *Kdr* in comparison to the SHF populations (C1,C8,C14,C21,C22) identified in our studies (**Figure 1C**). Both Pdgfra and Kdr are known to be transiently expressed during cardiac specification, and lower levels in Clusters 0 and 15 might suggest that these cells have moved beyond the cardiac mesoderm stage as characterized by co-expression of *Pdgfra* and *Kdr*. For additional comments on the potential identity of Clusters 0 and 15 please also refer to comments for Reviewer 1, point 1.

4. Left/right ventricular identity was mentioned throughout the manuscript with markers defining each. It is intriguing that the atrial cells appear as a single homogenous pool. Upon closer interrogation, do the atrial data distinguish left/right identity at these stages like the ventricle or does the atrium appear un-sided at these stages?

We thank the Reviewer for this comment and we gladly followed up on this interesting question. We find that when the atrial cells are subclustered, we do not observe segregation of the resulting clusters into populations with distinct left/right identity. However, we do observe mixed expression of left/right markers such as *Pitx2* (**Supplementary Figure 4D**). This suggests that while atrial cells at this stage may have been begun to express chamber specific markers, their overall transcriptome is still very similar. This is likely due to the comparatively early developmental stages within which the atrial cells were profiled, but should and could be further explored.
We have included this data in Supplementary Figure 4C of our revised manuscript, and have added the following text in the results section of the manuscript:

> ***"Atrial cells did not segregate discretely into left/right populations, though did show differential expression of markers of left/right identity such as Pitx2."***



Supplementary Figure 4D (subset for Reviewer)**.** *Top left*: schematic of atrial cells used for subclustering *Bottom left:* Resulting UMAP after subclustering atrial cell types. Right: FeaturePlots for markers of atrial cardiomoycytes as well as markers of AVC proximity (*Bmp2+/Rspo3+)*.

5. Minor comment, Figure 4, panels A and C, the colors of 8 and 14 swap between the two panels.

Thank you for this correction. We have modified this figure to fix the error and also colored the highlighted schematic such that the colors match the original colors used in the UMAP.

Reviewer 3 Advance Summary and Potential Significance to Field:
In this manuscript, Gonzalez, Schrode, and their colleagues present a valuable resource: a deep analysis of the transcriptomic heterogeneity present at three important early stages in the development of the mouse heart. The multifaceted nature of their analysis, which includes comparison with organoid data sets, trajectory prediction through RNA velocity analysis, examination of differential gene expression, and gene set enrichment analysis, has high value and is an important addition to the growing body of scRNA-seq work in the embryonic heart. In a particularly interesting application of their strategies for computational analysis of scRNA-seq data, the authors examine the effects of exogenous RA on transcriptome distribution in the developing heart, and their results suggest that exogenous RA may result in the aberrant differentiation of cardiac progenitor cells, shunting them away from the ventricular differentiation pathway and into other differentiation trajectories. While this study stops short of in vivo validation/testing of hypotheses suggested by the scRNA-seq data, it provides an important foundation for a number of potential future studies and will therefore be greatly appreciated by the field.

We thank the Reviewer for their enthusiasm about our work and its contribution to the field, we hope this is further improved by the revisions included here and will inspire and inform future studies.

Reviewer 3 Comments for the Author:
A few modifications to the text will enhance the clarity and impact of this manuscript.

1) The authors offer a provocative interpretation of the lack of a "specific population of FHF progenitors" in their data set, arguing that FHF cells closely resemble differentiating cardiomyocytes. Is it also possible that even the earliest timepoint analyzed by the authors is a bit too late to capture the FHF progenitors? Addition of commentary on this possibility would be helpful.

We thank the Reviewer for raising this important and very interesting question. We initially did indeed entertain the possibility that the earliest timepoint might be too late to capture the FHF progenitors when first looking at the merged data. However, when we examined the populations

present at the cardiac crescent stage alone (Figure 5D-E), we were able to identify a population of Nkx2-5+/Hcn4+/Tbx5+ cells that clustered separately which we identified as the FHF. For this reason, we feel confident that we have in fact captured the FHF transcriptionally. We have added additional commentary on the identification of this population at the CC stage within the relevant results section for Figures 1/5. The specific text is included below for the Reviewers convenience:

> *"Intriguingly, while we did identify a FHF population expressing Nkx2-5/Hcn4/Tbx5 when examining the CC stage alone (Figure S3A) when merged with later time points these cells did not form a discrete cluster and instead clustered with multiple differentiating cardiomyocyte populations (C7,C12, &C14), underscoring the identity of the FHF as a population of cells that most closely resemble differentiating cardiomyocytes."*

In future studies one could further address this by incorporating earlier stages published by other groups (refs) into our analysis, or by generating new data at these earlier stages and specifically interrogate the temporal emergence and transcriptional characteristics of FHF progenitor cells.

We are very interested in doing precisely what the Reviewer suggests, embedding our data in both earlier and later time points to generate an even larger window of heart development.
Given that this current analysis has already attempted to understand a rather large number of diverse questions computationally, we have opted to not include any of these approaches here aiming to keep some focus of the present paper.

2) The authors should consider softening their interpretations within the section titled "Identification of a transcriptionally distinct mesoderm population that gives rise to early myocardial intermediates." Generally speaking, computational analysis of scRNA-seq data can't demonstrate a lineage relationship -- this type of analysis can suggest a hypothesis that can be tested in vivo, but it can't show that one cell type gives rise to another one. Discussing this interesting cluster and its unique attributes is warranted, but the authors would need to show where these cells are in vivo and to perform tracking experiments to show their lineage to reach firm conclusions about whether these are "myocardial intermediates" and what they contribute to. Granted, the authors say that "these data suggest..." and "predicted to contribute...", but the title and tone of the section give the impression that we know more about the biography of these cells than we currently do.

Thank you for raining this important point. This comment together with a similar comment from Reviewer 1 has prompted us to take an even closer look at Clusters 0 and 15. After examining these cluster more closely we concluded that differences in in sequencing depth may partially contribute to segregation of this cluster from the other myocardial cell types (detailed in our response to Reviewer #1). We have taken the suggestion to strongly soften this interpretation, and have removed the detailed analysis of this population in the original Figure S6. While we still remain interested in understanding more about this population, which does appear to have unique features and characteristics, an emphasis on this population may be better suited for future work where we will follow up on the predicted lineage relationships with *in vivo* validation experiments as suggested by the Reviewer.
For additional analysis and related new information on Clusters 0 and 15 please also refer to comments for Reviewer 1, point 1.

3) While the paper is generally clearly written, there are several occasions where it is difficult to link claims made in the text with data provided in the figures or legends. This is probably a consequence of the need to adhere to a reasonable word limit for the text, but it can still leave the reader unable to connect the dots. Just to cite a couple of examples from the first section of the results: The authors state that multiple Pdgfra+ mesoderm populations are more commonly found in S phase than their differentiated Nkx2-5+ progeny, but there are no concrete references to specific clusters or to the percentages of cells that justify the "more commonly found" conclusion -- simple additions to the text could make this much more clear. In the same paragraph, the authors state that "Isl1+ SHF progenitors had a balanced contribution from all three stages", but the reader cannot easily discern which specific clusters are being considered as "Isl1+ SHF progenitors" in this context. There are other places in the text where it is easier to catch exactly what is referred to in the figures -- if the authors could go through the text and add some further annotations where needed to guide the reader through the data, this would enhance the value of

the manuscript.

This is great feedback, thank you. We recognize that the material and content of this paper is dense. We have revised our entire manuscript in light of this comment and focused on enhancing clarity in both text and figures wherever possible. Examples of this include the color- revised and better annotated UMAPs, additional schematics, and many edit to the manuscript text.

Second decision letter

MS ID#: DEVELOP/2022/200557

MS TITLE: Dissecting Mechanisms of Chamber-Specific Cardiac Differentiation and its Perturbation Following Retinoic Acid Exposure

AUTHORS: David Matthew Gonzalez, Nadine Schrode, Tasneem Ebrahim, Nicolas Broguiere, Giuliana Rossi, Lika Drakhlis, Robert Zweigerdt, Matthias Lutolf, Kristin Beaumont, Robert Sebra, and Nicole Caroline Dubois
ARTICLE TYPE: Research Article

I am happy to tell you that your manuscript has been accepted for publication in Development, pending our standard ethics checks.