nature portfolio

Peer Review File

Large-scale Single-nuclei Profiling Identifies Role for ATRNL1 in Atrial Fibrillation

Corresponding Author: Dr Patrick Ellinor

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this manuscript Hill et al, perform scRNA-seq on samples from the left atria of non-failing hearts from a cohort of healthy organ donors and subjects with atrial fibrillation. While they found no differences in cellular composition and very limited cell-type-specific differences, they found that ATRNL1 was overexpressed among AF patients. Functional studies in hESC-derived aCM suggest a potential role of ATRNL1 in modulating action potential.
Furthermore, they used cell-type-specific expression in atria to prioritize candidate AF risk genes within AF-associated risk loci.
This study introduces a unique dataset that can provide valuable insights into cellular changes associated with AF.

However, I have several concerns with the data analysis and presentation that would require a major revision of the manuscript.

Major concerns:
1. Description of analysis is insufficient and code has not been provided (apologies if I missed this).
2. Sample processing and QC measures are described but no actual thresholds are provided. The authors should provide supplemental figures that summarize the data quality of all samples.
3. Data should be shared as appropriate. I couldn't find any description of whether/ how the data will be shared.
4. It is unclear how many cells were collected per individual (range).
5. Large variation in total cell number and high inter individual variation will affect the ability to detect differences in proportion. Supplemental Figure 1b suggests that the cell-type proportions are highly variable between individuals.
6. The patient samples should be discussed in more detail. Some metadata are captured but the severity of the phenotype and time since diagnosis are not clear to me.
7. The DEG analysis needs to be documented in detail.
a. How was pseudobulking performed? Which cut-offs were chosen for the minimal number of cells to include a sample?
b. It seems likely that both high variation and potentially a weak phenotype contribute to the limited number of DEG. Given that DEGs only came from 2 of the largest 3 clusters detection power is likely a major issue. This should be investigated and presented in the supplement as well. It's fine to have low power, however it would be very informative to see these analyses.
c. For the resulting DEG (mainly ATRNL1): how variable was their expression across samples?
8. Supplemental Tables are not well annotated (missing description of column headers etc)
9. The circular RNA analysis is confusing. While an interesting observation, the interpretation of the result seems overly confident. It would be better to keep this speculation to the discussion.

Minor concerns
1. The manuscript needs substantial editing.
2. Line 121: 'Very few substantial changes in transcription were identified after completing this comparison, consistent with our pseudo-bulk PCA results.' While the results are indeed the same (I.E. no changes), this didn't need to be the case for cell-type-specific gene expression.
3. Last sentence of abstract: "In sum, we have identified a role for ATRNL1 and other differentially regulated genes that may serve as potential therapeutic targets for this common arrhythmia." It seems that ATRNL1 was the only one identified?

4. How was the window size around sentinel variants chosen? What is the actual window size it's 1MB in the figure legend and 500kb in the method section.
Why not use LD as way to define blocks instead? Did the authors observe a distance dependency between identified genes and the variant at the center of the block? (for example, the identified genes were not uniformly distributed within the window but grouped closer to the variant?)

Reviewer #2

(Remarks to the Author)
In this study, Hill et al. investigate the cellular and transcriptional landscape of atrial fibrillation using singe nucleus RNA sequencing. By examining left atrial samples from 16 donor controls and 18 patients with atrial fibrillation the authors identified cardiomyoyctes and macrophages as displaying the greatest transcriptomic differences. Among differentially expressed genes in cardiomyocytes, the authors chose to focus on Attractin-like 1 (ATRNL1). They provide evidence that ATRNL1 is expressed in cardiomyocytes and that modulating ATRNL1 impacts the electrophysiological properties of hPSC-derived atrial cardiomyocytes. Lastly, they utilize their snRNAseq dataset to examine the expression of candidate genes associated with atrial fibrillation.

Specific comments:

1. Please provide additional information describing the location of the atrial tissue subjected to snRNAseq and include more detailed clinical information. Please also provide QC metrics (genes/cell, UMIs/cell, %mitochondrial reads/cell) to better assess the quality of the snRNAseq data.

2. The claim that cellular composition is not changed between donor and atrial fibrillation tissues is not well supported by the presented data. The authors do not provide any information pertaining to the sub-composition of major cardiac cell types between donor and atrial fibrillation samples.

3. Please expand on comments made that sex dependent differences were observed at the transcriptomic level.

4. Is ATRNL1 expressed in ventricular cardiomyocytes? Please provide quantification of the immunostaining images shown in Figure 3.

5. It is not clear whether ATRNL1 exerts its effects through the long isoform, short isoform, or circRNA species. Does knockdown or overexpression of these isoforms of ATRNL1 lead to different outcomes? The impact on electrophysiology is not very detailed and restricted to immature iPSC-derived atrial cardiomyocytes.

6. Cardiac cell type specific expression of genes associated with candidate atrial fibrillation loci should be validated across different snRNAseq datasets. For example, from this manuscript and published available data.

Reviewer #3

(Remarks to the Author)
In this manuscript by Hill et all, the authors perform snRNAseq on LA tissue from AF patients and identify a role of ATRNL1 in AF. I would like to congratulate the authors on the work performed and the well written paper. However, I do have some concerns regarding especially the patient populations included and perhaps the lack of a clear difference in phenotype (expect from the AF), which rather appears more or less to be a form of "lone AF", which no longer is an accepted for m of AF. I have the following comments:

- Supplementary table 1 should include more summarized date and statistical analysis on the patient characteristics that is also provided for the individual patients in Supplementary table 1 including for instance LVEF, LA size, co-comorbidities and other relevant information. Although it is very nice to have the individual data provided it is difficult to get a complete overview from by the individual values.

- I have a hard time figuring out what the reason for the occurrence of AF is in the AF cohort is? There is always a reason for AF to occur, but the patient cohorts seem very similar. Is it a genetic component only then? Or are there co-comorbidities the authors do not have the information on that could be the underlying cause. Please elaborate.

- Are there any information available on the AF history for these patients?

- The differential abundance testing identified not any major shifts in cell type composition between the SR and AF cohorts. However, I am no sure if I understand the conclusion that the Masson's Trichrome staining confirms the finding of similar tissue composition between the SR and AF cohorts. The staining only provides information on the collagen content and the volume of cardiomyoctes, but does not provide any information on the other cell types. For this purpose the authors would need to perform additional immunostainings with specific cellular markers.

- Have the authors quantified the fibrotic content in the tissue? Usually more fibrosis is reported in AF patients, which is a change in tissue composition.

- A recent paper (PMID: 37440641) in which the authors performed scRNA-seq of AF patients the MP/DC cluster expanded twofold, whereas endothelia and mural cells decreased in frequency. Could the authors please comment on the findings of this paper compared to their findings?

- Why do the authors only have access to LA tissue and not RA tissue? It could have been interesting to see if the same results were obtained in RA. Furthermore please specify from what exact location of the LA the samples were taken (LA posterior wall, LA free wall, RA appendage or somewhere else)? Especially the LA posterior wall is important for AF induction and would have been interesting to investigate as well.

- Please provide negative controls for the immunohistochemical staining performed in atrial tissue.

- I agree that the Cx43 and ATRNL1 co-stain in the same region, but this does not confirm a co-localization of the proteins. I would be careful concluding this as an overlap in fluorescence does not necessarily indicate co-localization. Additionally I would suggest the authors to perform Pearson's correlation coefficient (PCC) as a statistic for quantifying co-localization.

- I must disagree with the statement on line "272-273" regarding the SK channels stating: "KCNN2 and KCNN3 encode small conductance calcium-activated potassium channels. However,
the role of these channels at steady state and during AF remains unclear". Numerous studies have both in animal models and in human tissue samples provided clear evidence of the role of SK channels in action potential repolarization in SR, while changes in trafficking and function of the channels as well as transcriptional and translational changes during AF also has been demonstrated. Please include these information on the role of SK channels in SR and AF.

- Line "301-303" states that the authors performed "in vivo and in vitro" immunoflouresence. Please clarify what you mean with in vivo immunofloresence? If the reference is to tissue staining the in vivo phrasing needs to be changed.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Reviewer #2

(Remarks to the Author)
The authors have significantly revised their manuscript and it is much improved. If possible, it is recommended that the authors perform RNAscope to address expression of the ATRNL long, short, and circular RNA form in human atrial and ventricular cardiac specimens with and without AF. This information would be useful to the field moving forward.

Reviewer One

In this manuscript Hill et al, perform scRNA-seq on samples from the left atria of non-failing hearts from a cohort of healthy organ donors and subjects with atrial fibrillation. While they found no differences in cellular composition and very limited cell-type-specific differences, they found that ATRNL1 was overexpressed among AF patients. Functional studies in hESC-derived aCM suggest a potential role of ATRNL1 in modulating action potential. Furthermore, they used cell-type-specific expression in atria to prioritize candidate AF risk genes within AF-associated risk loci. This study introduces a unique dataset that can provide valuable insights into cellular changes associated with AF.

However, I have several concerns with the data analysis and presentation that would require a major revision of the manuscript.

Authors' response: We sincerely appreciate your thoughtful comments and constructive feedback.

1. Description of analysis is insufficient and code has not been provided (apologies if I missed this).

Authors' response: We thank you for pointing out this deficiency. Overall, we conducted sample collection, snRNA-seq, and downstream computational analysis very similar to our previous studies[1,2]. For our analyses we utilized previously described and widely used analytical packages, and so we do not have any custom code that was used in this manuscript. We appreciate the Reviewer's point and to better describe our analysis better up-front, we have edited the Results section to read as follows:

> "**snRNA-seq of human left atrial tissue from AF cases and controls**
> To characterize the cellular and molecular characteristics associated with AF, we performed snRNA-seq on samples from either the anterior wall of the LA near the appendage, or from the posterior LA wall near the annulus of human patients with AF who were not in heart failure (n=19). We carried out strict quality control, including sample sex check, unique molecular identifier (UMI) decay curve analysis, and a crosscheck of genotype fingerprinting with genome sequencing data (see **Methods**). A total of 2 samples failed our quality control standards and were removed from all subsequent analysis. Our final data set included 18 AF cases and 16 controls (**Fig. 1a**, **Supplementary Table 1**, and **Supplementary Table 2**). Next, batch correction was performed with the single-cell variational inference framework (scVI)[3]. Low-quality nuclei, doublets, and misclassified cells were then removed following clustering, leaving us with 179,697 nuclei that occupied a total of 15 transcriptionally distinct clusters (**Fig. 1b**, and **Supplementar**y **Fig. 1**). Differential expression analysis was performed to determine the top representative genes for each cluster (**Supplementary Table 3**), and then clusters were annotated based on the expression of canonical cell

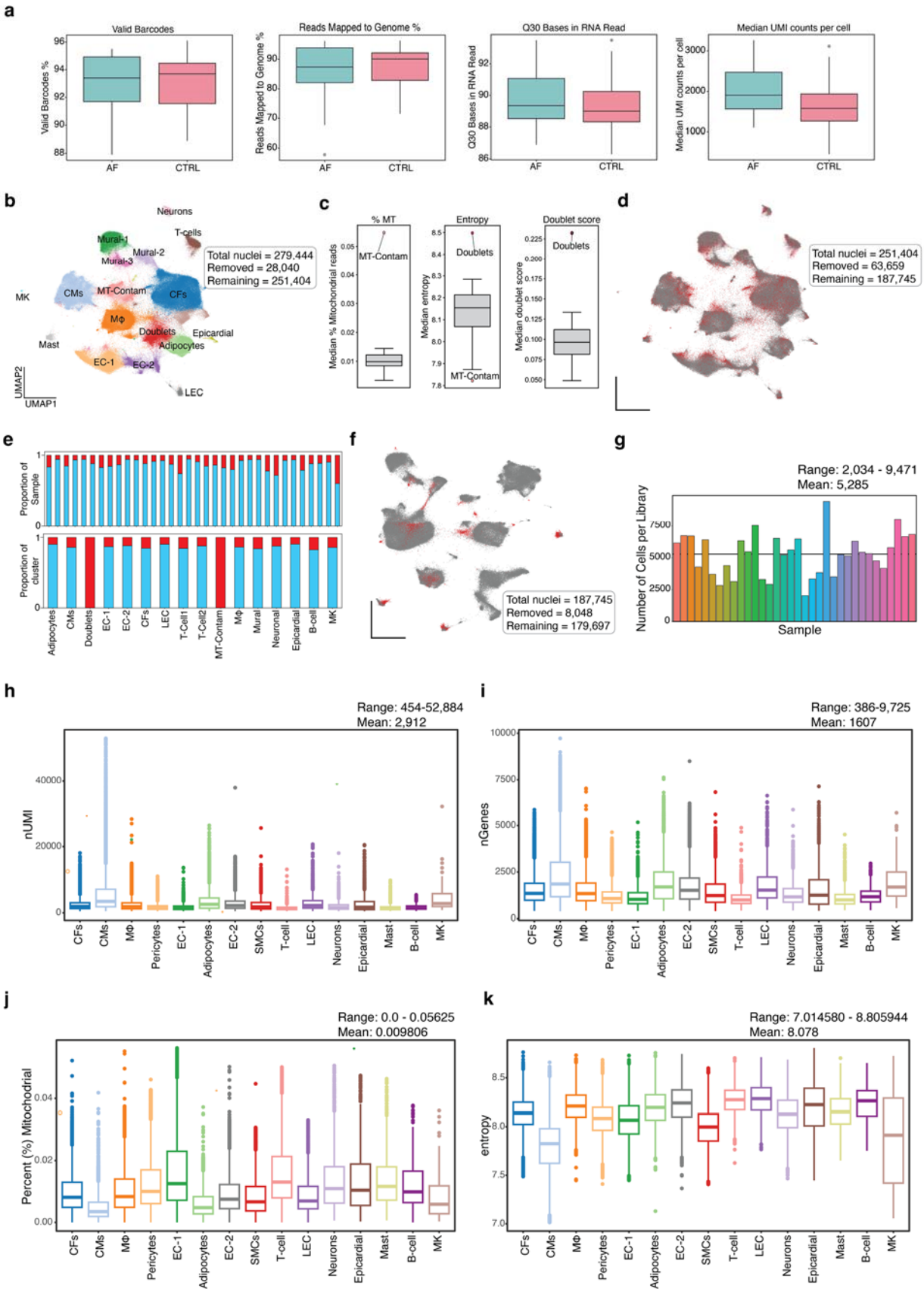
markers (**Fig. 1c**). For example, *TBX5* was enriched in cardiomyocytes (CMs), *ERG* in endothelial cell clusters (EC-1, EC-2, and LEC), and *ADIPOQ* was highly expressed in adipocytes (**Fig. 1d**). ”

2. Sample processing and QC measures are described but no actual thresholds are provided. The authors should provide Supplementary figures that summarize the data quality of all samples.

Authors’ response: Thank you for raising this important point. Firstly, we have added **Supplementary Table 12** to our manuscript, which contains all the quality control metrics output by CellRanger for each sample library.

We have also added a new figure (**Response Figure 1**, below) as **Supplementary Figure 1** to the updated manuscript. Firstly, we have provided QC metrics for each individual library as output from CellRanger (**Response Figure 1a**). Overall, the number of valid barcodes, percentage of reads mapped to the human (hg38) genome, and the median UMI counts per cell were comparable between AF and CTRL samples. Secondly, we included a full accounting of our post-sample QC process (**Response Figure 1b-g**). After sample quality control, we were left with a dataset containing 279,444 non-empty droplets (nuclei) (**Response Figure 1b**). Next, batch correction with scVI and Leiden clustering was performed. Looking at nuclei quality control metrics like percentage mitochondrial reads, doublet score (output from Scrublet) ,and entropy we identified 2 obvious clusters to be removed, the Doublets and MT-Contam clusters (**Response Figure 1c**). We then performed quality control for all nuclei, as described in the methods, identifying an additional 63,659 nuclei for removal (**Response Figure 1d**). Overall, the nuclei identified from this quality control process were relatively evenly distributed across samples, and clusters (**Response Figure 1e**). Finally, we performed sub-clustering and marker gene analysis to identify misclassified cells for removal leaving us with our final dataset which was re-clustered (nuclei= 179,697)(**Response Figure 1f**). After quality control, our dataset contained on average 5,285 nuclei per sample (range = 2,034 – 9,471) (**Response Figure 1g**).

We also included post-quality control metrics, including nUMI, nGenes, percent mitochondrial reads, and entropy for each cluster (**Response Figure 1h-k**). Moreover, we reported both the range and mean for each of these metrics across all cell clusters in the figure to make it more readily accessible.


a
Valid Barcodes
Valid Barcodes %
AF
CTRL
Reads Mapped to Genome %
Reads Mapped to Genome %
Q30 Bases in RNA Read
Q30 Bases in RNA Read
Median UMI counts per cell
Median UMI counts per cell
b
Neurons
T-cells
Mural-1
Mural-2
Mural-3
Total nuclei = 279,444
Removed = 28,040
Remaining = 251,404
MK
CMs
MT-Contam
CFs
MΦ
Doublets
Epicardial
Mast
Adipocytes
EC-1
EC-2
LEC
UMAP2
UMAP1
c
% MT
Median % Mitochondrial reads
MT-Contam
Entropy
Median entropy
Doublets
MT-Contam
Doublet score
Median doublet score
Doublets
d
Total nuclei = 251,404
Removed = 63,659
Remaining = 187,745
e
Proportion of Sample
Proportion of cluster
Adipocytes
CMs
Doublets
EC-1
EC-2
CFs
LEC
T-Cell1
T-Cell2
MT-Contam
MΦ
Mural
Neuronal
Epicardial
B-cell
MK
f
Total nuclei = 187,745
Removed = 8,048
Remaining = 179,697
g
Range: 2,034 - 9,471
Mean: 5,285
Number of Cells per Library
Sample
h
Range: 454-52,884
Mean: 2,912
nUMI
CFs
CMs
MΦ
Pericytes
EC-1
Adipocytes
EC-2
SMCs
T-cell
LEC
Neurons
Epicardial
Mast
B-cell
MK
i
Range: 386-9,725
Mean: 1607
nGenes
j
Range: 0.0 - 0.05625
Mean: 0.009806
Percent (%) Mitochodrial
k
Range: 7.014580 - 8.805944
Mean: 8.078
entropy

Response Figure 1. snRNA-seq sample quality control metrics. (**a**) Distribution of snRNA-seq quality control metrics for each library derived from CellRanger count as box boxplots grouped by patient category. (**b**) UMAP embedding of all nuclei post-sample QC (n = 279,444) colored by Leiden clustering and labelled according to predicted cell type. (**c**) Distribution of the median of three quality control metrics for each cluster shown as boxplots, including percent of unique molecular identifiers (UMIs) mapping to mitochondrial genes (% MT),entropy, and the Scrublet estimated doublet score. Clusters containing cells to be removed are highlighted in red (n= 28,040 nuclei). (**d**) UMAP embedding of all nuclei after removal of low-quality clusters identified in **b** and **c** (n = 251,404). Additional low-quality nuclei as detected per cluster and per-sample are colored in red. (**e**) Proportion of each sample and cluster removed during the quality control procedure shown as stacked bar graph. (**f**) UMAP representation of remaining nuclei after removal of low-quality clusters and per-cluster quality control. Red nuclei were deemed as misclassified or low-quality following sub-cluster analysis. (**g**) Number of nuclei per library included in final snRNA-seq embedding. (**h**) Distribution of number of unique molecular identifiers (nUMI). (**i**) Distribution of number of unique genes (nGene). (**j**) Distribution of percentage of mitochondrial transcripts per cell. (**k**) Distribution of entropy across nuclei of each unique cell type. Center line, median; box limits, upper and lower quartiles; whiskers,1.5x interquartile range.

3. Data should be shared as appropriate. I couldn't find any description of whether/ how the data will be shared.

Authors' response: We completely agree with the spirit of the Reviewer's comment, and we have a long track record of publicly distributing out genetic and snRNA data. We have made the data available, both raw fastq and processed data matrices, though the gene expression omnibus (GEO) as well as through the Broad Institute's Single Cell Portal (https://singlecell.broadinstitute.org/single_cell). Upon publication the Single Cell Portal users will have the ability to explore our dataset interactively online without the need to download and analyze the large count matrices. Further, all metadata, including quality control metrics, will be available through the Single Cell Portal. We have updated our data availability statement to reflect these changes, and it reads as follows:

> **"Data Availability**
> Processed single-nucleus transcriptomic data are available through the Broad Institute's Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) under project ID SCP2489. Raw and processed next-generation sequencing data have been deposited at the NCBI Gene Expression Omnibus with accession number GSE255992."

4. It is unclear how many cells were collected per individual (range).

Authors' response: Thank you for pointing this out, please see above response for Reviewer-1 question-2 and **Supplementary Figure 1g** (**Response Figure 1g**). We captured an average of 5,285 nuclei per library after sample QC and as it states in the methods, we loaded each 10X lane aiming to achieve 6,000 nuclei per library.

5. Large variation in total cell number and high inter individual variation will affect the ability to detect differences in proportion. Supplementary Figure 1b suggests that the cell-type proportions are highly variable between individuals.

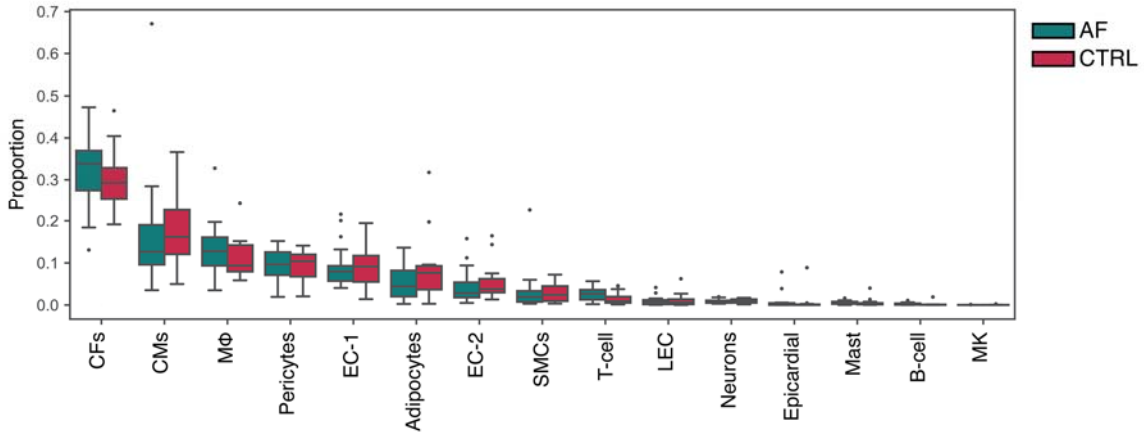
Authors' response: Indeed, and this is a common issue with snRNA-seq analysis that we[1,2] and others[4] have had to overcome using various state-of-the art computational methods. To overcome the many issues associated with tissue sampling, we chose Milo for performing differential abundance testing[5]. Milo is constructed to be good for modeling complex experimental designs where samples may be of different ages, backgrounds, sex, disease severity, as well as other sources of variation. Milo largely overcomes challenges associated with common differential abundance testing by using k-nearest neighbor graphs as opposed to clustering of cells into discrete groups. For our Milo-based differential abundance analysis we utilized a k=50, to get at least 1 cell from each patient into every neighborhood resulting in a neighborhood embedding of 7,499 neighborhoods. And as stated in the manuscript, we observed no statistically significant changes in cell composition between controls and patients with AF.

To further evaluate changes in tissue composition with a different statistical approach we have decided to utilize a Bayesian model that we have successfully utilized and experimentally validated previously[1,2], the single-cell compositional data analysis (scCODA) framework. scCODA models cell-type counts with a hierarchical Dirichlet-Multinomial distribution that accounts for the uncertainty in cell-type proportions and the negative correlative bias via joint modeling of all measured cell-type proportions instead of individual ones[6]. We applied scCODA to our AF dataset to assess differences in tissue composition between control donors and patients with AF (**Response Figure 2a**). We observed no statistical differences between any cell cluster for AF or control LA tissue.

To address the concerns of the reviewer we have added **Response Figure 2** to Supplementary Figure 2 (**Supplementary Figure 2c**) and discussed in the results section. Further, we have added the following line to the discussion to address your valid point:

> "A potential limitation of our study would that the observed tissue compositional variability among the patients profiled here may be difficult to properly account for even with modern analytical frameworks (e.g. Milo and scCODA) and could require larger sample sizes with greater statistical power to determine the distinctions more accurately between patient groups."

a



Response Figure 2. Cell composition of controls and patients with atrial fibrillation. (**a**) Box plot displaying the differences in cell composition for each cardiac cell type in controls (red) and patients with atrial fibrillation (AF) (blue). Boxplot represented as: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

6. The patient samples should be discussed in more detail. Some metadata are captured but the severity of the phenotype and time since diagnosis are not clear to me.

Authors' response: We thank the Reviewer for their comment. We focused on patients with permanent AF to better control for the many diverse etiologies associated with this common arrythmia. In response to comments from all Reviewers, we have cleaned up and updated **Supplementary Tables 1 and 2** to reflect all the information that was available regarding time of diagnosis in relation to transplantation. To describe the patient cohorts with more clarity we have updated the Results to read as follows:

> "To characterize the cellular and molecular characteristics associated with AF, we performed snRNA-seq on samples from the LA of human patients with AF who were not in heart failure (n=17) as well as non-AF controls (n=19)(**Supplementary Table 1**, and **Supplementary Table 2**). We selected patients without heart failure to better control for the many diverse etiologies associated with this common arrythmias. Overall, our controls (CTRL) were 63 % female, the mean age was 68 (SD = 7.5), 25% had a history of taking beta-blockers, and 6% had taken anticoagulants. The AF cases were 61% female, the mean age was 66 (SD = 8.2), 56% had a history of taking beta-blockers, and 44% were being administered anticoagulants."

7. The DEG analysis needs to be documented in detail.

Authors' response: Thank you for this comment. To clarify our approach, we implemented a formal differential expression model controlling for the correlation amongst nuclei from the same individual by summing gene counts across all nuclei in a cluster within an individual patient and treating the data as a bulk RNA sequencing experiment before running limma-voom, in a similar manner as mentioned in the methods for marker gene detection. This approach, based on Lun et. al. (2017), is performed to avoid systematic batch effects and restore type I error control.[7]

We have updated the DEG section of the methods to read as follows:

> **"Differential expression analysis**
> A formal differential expression model controlling for the correlation amongst nuclei from the same individual was performed by summing gene counts across all nuclei in a cluster within an individual patient and treating the data as a bulk RNA sequencing experiment before running limma-voom, in a similar manner as mentioned above. The differential expression analysis across cases and controls was performed with the method limma-voom using the R-package limma v3.40.6. First, the expression counts were summarized by cluster and sample, including each observation with at least 20 nuclei per sample/cluster combination. For each cluster, genes were filtered using the algorithm implemented in filterByExpr from the R-package edgeR v3.26.8[77] to retain genes with sufficiently large counts with default settings. The data was then normalized using DESeq2 v1.24.0[78] normalization. Differential expression was then calculated using the limma-voom pipeline[79] with the model '~af + sex'', followed by extraction of the contrast comparing expression in cases versus controls. Genes were considered significant in the differential expression analyses when the adjusted P-value (by Benjamini Hochberg) was < 0.05. For visualization of the differential expression results in violin plots, the filtered and normalized count data was converted to the unit counts per million (CPM) as implemented in the cpm function from the edgeR R-package. Since there can be notable background contamination present in single nuclei RNA-sequencing experiments we generated a flag that would identify genes that have a high probability of coming from the background. We followed a procedure that was previously described[10] to calculate the flag based on CellRanger and also CellBender counts."

8. How was pseudobulking performed? Which cut-offs were chosen for the minimal number of cells to include a sample?

Authors' response: For our pseudo-bulking approach implemented in the DEG analysis, nuclei were only aggregated in an individual if there were more than 20 nuclei of the given cluster. As outlined in point 7 above, we have edited the methods to make this point clearer (See above response to point 7).

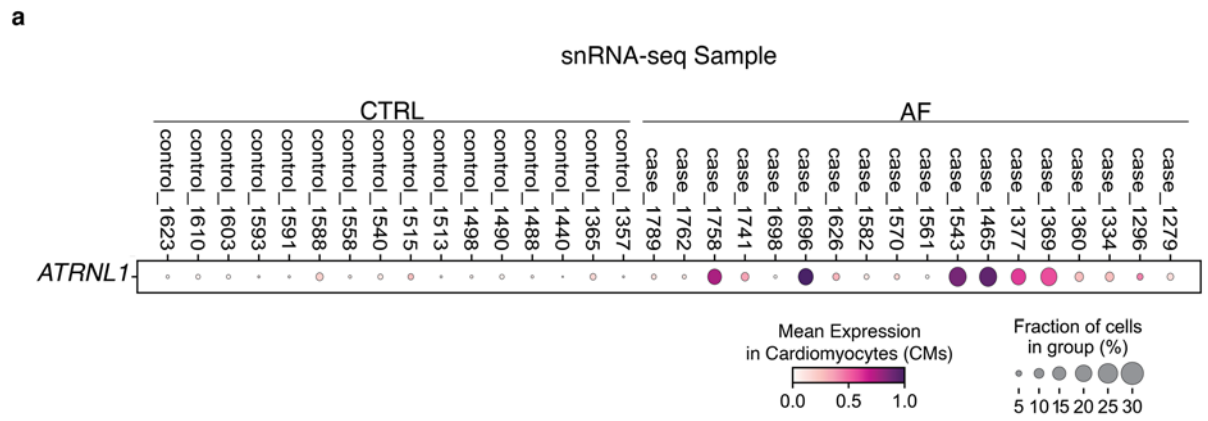
9. It seems likely that both high variation and potentially a weak phenotype contribute to the limited number of DEG. Given that DEGs only came from 2 of the largest 3 clusters detection power is likely a major issue. This should be investigated and presented in the supplement as well. It's fine to have low power, however it would be very informative to see these analyses.

Authors' response: The variation we observe is similar to what's found in other studies[1,2,4]. However, we do agree that the phenotype is more subtle than our prior work in end-stage heart failure and likely contributes to the lower number of DEGs observed in our study. Power or sensitivity for DEG analysis in scRNA-seq studies is overall a function of the number of cells profiled per sample, number of samples included in the dataset, expression level of each gene, and depth of sequencing. Several studies have performed computationally intensive simulations to determine the optimal experimental setup for improving DEG analysis[8,9]. Overall, the common recommendation for improving power and reducing false positives from such analysis is to increase the number of cells/nuclei profiled per sample. To address this, we have added the following limitation to the discussion:

> "Moreover, profiling more patients at a greater depth (nuclei per sample) could improve our statistical power to accurately identify differentially expressed genes across more cell types[8]."

10. For the resulting DEG (mainly ATRNL1): how variable was their expression across samples?

Authors' response: As with any snRNA-seq study we did observe variability in the expression of *ATRNL1* in cardiomyocytes across samples (**Response Figure 3a**). Although, we do observe that all samples with high levels of *ATRNL1* expression are from AF cases. We have added **Response Figure 3a** to **Supplementary Figure 4** to highlight this variability.



Response Figure 3. snRNA-seq expression of *ATRNL1*. (**a**) Dot plot displaying the mean expression of *ATRNL1* in cardiomyocytes across all patient samples. The size of the dot indicates the percentage of cells from each patient expressing *ATRNL1*.

11. Supplementary Tables are not well annotated (missing description of column headers etc)

Authors' response: We are sorry for the trouble, and we have now added keys for each column header to the bottom or right side of the tables where appropriate.

12. The circular RNA analysis is confusing. While an interesting observation, the interpretation of the result seems overly confident. It would be better to keep this speculation to the discussion.

Authors' response: We thank you for your feedback. We have removed the following text from the results:

> "Among these genes was *THBS1* or thrombospondin 1. Importantly, *THBS1* is known to be inhibited by miR-18a[10], and interestingly, among our putative miRNAs to be bound by circATRNL1 is miR-18a. The consistent decrease in *THBS1* expression (**Fig. 4c-d**) when *ATRNL1* and circATRNL1 are depleted suggests that *THBS1* expression is being reduced because of excess miR-18a being released from circATRNL1."

13. The manuscript needs substantial editing.

Authors' response: We appreciate your careful examination of our work and we have thoroughly edited the revised manuscript.

14. Line 121: 'Very few substantial changes in transcription were identified after completing this comparison, consistent with our pseudo-bulk PCA results.' While the results are indeed the same (I.E. no changes), this didn't need to be the case for cell-type-specific gene expression.

Authors' response: Thank you, we have edited this sentence.

15. Last sentence of abstract: "In sum, we have identified a role for ATRNL1 and other differentially regulated genes that may serve as potential therapeutic targets for this common arrhythmia." It seems that ATRNL1 was the only one identified?

Authors' response: We have revised this sentence to focus on *ATRNL1.*

16. How was the window size around sentinel variants chosen? What is the actual window size it's 1 MB in the figure legend and 500kb in the method section. Why not use LD as way to define blocks instead?

Authors' response: We thank the Reviewer for pointing out this discrepancy and suggesting an alternative method. The window size was chosen as 1 megabase indexed around the sentinel variant. There is a typo in the methods section, and it was corrected to "+/- 500 kb". The sentinel variant was defined as the variant with the smallest P-value at the genetic locus. Indeed, another option for choosing the window could be using LD. An LD approach is less straightforward as it requires choosing an appropriate reference to calculate LD and defining an LD cutoff. Additionally, there are many genetic loci with multiple independent signals, an LD window around the sentinel variant would not capture those complex signals. Furthermore, since the GWAS results are a meta-analysis across different datasets (some with different imputation references), the availability of the sentinel variant in an LD reference is not guaranteed and a different strategy for those variants would have to be implemented. Furthermore, the choice of the population for the LD reference would influence the LD pattern and therefore the window. Given that the GWAS meta-analysis is multi-ancestry we did not want to restrict this approach by using for example a European only LD reference. With

these issues in mind, we decided that a distance-based approach would be the less biased approach for these analyses.

17. Did the authors observe a distance dependency between identified genes and the variant at the center of the block? (for example, the identified genes were not uniformly distributed within the window but grouped closer to the variant?)

Authors' response: We thank the Reviewer for this interesting question. We took a closer look at the location of the 59 identified genes relative to the sentinel variant and restricted to protein coding genes at the loci. 37% of these genes were also the nearest protein coding gene and 63% of the genes were within the 3 closest protein coding genes. This points towards an enrichment of the identified genes closer to the sentinel variant.

Reviewer Two

In this study, Hill et al. investigate the cellular and transcriptional landscape of atrial fibrillation using singe nucleus RNA sequencing. By examining left atrial samples from 16 donor controls and 18 patients with atrial fibrillation the authors identified cardiomyoyctes and macrophages as displaying the greatest transcriptomic differences. Among differentially expressed genes in cardiomyocytes, the authors chose to focus on Attractin-like 1 (ATRNL1). They provide evidence that ATRNL1 is expressed in cardiomyocytes and that modulating ATRNL1 impacts the electrophysiological properties of hPSC-derived atrial cardiomyocytes. Lastly, they utilize their snRNAseq dataset to examine the expression of candidate genes associated with atrial fibrillation.

Authors' response: We would like to extend our appreciation for your careful examination of our work, and contributions toward improving the manuscript.

1. Please provide additional information describing the location of the atrial tissue subjected to snRNAseq and include more detailed clinical information. Please also provide QC metrics (genes/cell, UMIs/cell, %mitochondrial reads/cell) to better assess the quality of the snRNAseq data.

Authors' response: Samples for this study were collected from either the anterior wall of the LA near the appendage, or from the posterior wall near the annulus. We have added this information to the results.

To address the Reviewers concerns, as well as those of Reviewer One (similar response also stated above) regarding the snRNA-seq data we have added **Supplementary Table 12** to our manuscript, which contains all the quality control metrics output by CellRanger for each sample library.

Moreover, we have also added a new figure (**Response Figure 1**) as **Supplementary Figure 1, panels a-d** (above) to the updated manuscript Firstly, we have provided QC metrics for each individual library as output from CellRanger (**Supplementary Figure 1a**). Overall, the number of valid barcodes, percentage of reads mapped to the human (hg38) genome, and the median UMI counts per cell were comparable between AF and CTRL samples. Secondly, we included a full accounting of our post-sample QC process (**Supplementary Figure 1b-g**). After sample quality control, we were left with a dataset containing 279,444 non-empty droplets (nuclei) (**Supplementary Figure 1b**). Next, batch correction with scVI and Leiden clustering was performed. Looking at nuclei quality control metrics like percentage mitochondrial reads, doublet score (output from Scrublet) ,and entropy we identified 2 obvious clusters to be removed, the Doublets and MT-Contam clusters (**Supplementary Figure 1c**). We then performed quality control for all nuclei, as described in the methods, identifying an additional 63,659 nuclei for removal (**Supplementary Figure 1d**). Overall, the nuclei identified from this quality control process were relatively evenly distributed across samples, and clusters (**Supplementary Figure 1e**). Finally, we performed sub-

clustering and marker gene analysis to identify misclassified cells for removal leaving us with our final dataset which was re-clustered (nuclei= 179,697)(**Supplementary Figure 1f**). After quality control, our dataset contained on average 5,285 nuclei per sample (range = 2,034 – 9,471) (**Supplementary Figure 1g**).

We also included post-quality control metrics, including nUMI, nGenes, %mitochondrial reads, and entropy for each cluster (**Supplementary Figure 1h-k**). Further, we reported both the range and mean for each of these metrics across all cell clusters in the figure to make it more readily accessible.

2. The claim that cellular composition is not changed between donor and atrial fibrillation tissues is not well supported by the presented data. The authors do not provide any information pertaining to the sub-composition of major cardiac cell types between donor and atrial fibrillation samples.

Authors' response: To address the Reviewer's concerns, as well as those of Reviewer One (above, question 5) regarding cellular composition analysis, we have the same response to both points (stated both here and above). We agree that this is a common issue with snRNA-seq analysis that we[1,2] and others[4] have had to overcome using various state-of-the art computational methods. To overcome the many issues associated with tissue sampling, we chose Milo for performing differential abundance testing[5]. Milo is constructed to be good for modeling complex experimental designs where samples may be of different ages, backgrounds, sex, disease severity, as well as other sources of variation. Milo largely overcomes challenges associated with common differential abundance testing by using k-nearest neighbor graphs as opposed to clustering of cells into discrete groups. For our Milo-based differential abundance analysis we utilized a k=50, to get at least 1 cell from each patient into every neighborhood resulting in a neighborhood embedding of 7,499 neighborhoods. And as stated in the manuscript, we observed no statistically significant changes in cell composition between controls and patients with AF.

To further evaluate changes in tissue composition with a different statistical approach we have decided to utilize a Bayesian model that we have successfully utilized and experimentally validated previously[1,2], the single-cell compositional data analysis (scCODA) framework. scCODA models cell-type counts with a hierarchical Dirichlet-Multinomial distribution that accounts for the uncertainty in cell-type proportions and the negative correlative bias via joint modeling of all measured cell-type proportions instead of individual ones[6]. We applied scCODA to our AF dataset to assess differences in tissue composition between control donors and patients with AF (**Response Figure 2a**). We observed no statistical differences between any cell cluster for AF or control LA tissue.

To address the concerns of the reviewer we have added **Response Figure 2** to Supplementary Figure 2 (**Supplementary Figure 2c**) and discussed in the results section. Further, we have added the following line to the discussion to address your valid point:

"A potential limitation of our study would that the observed tissue compositional variability among the patients profiled here may be difficult to properly account for even with modern analytical frameworks (e.g. Milo and scCODA) and could require larger sample sizes with greater statistical power to determine the distinctions more accurately between patient groups."

3. Please expand on comments made that sex dependent differences were observed at the transcriptomic level.

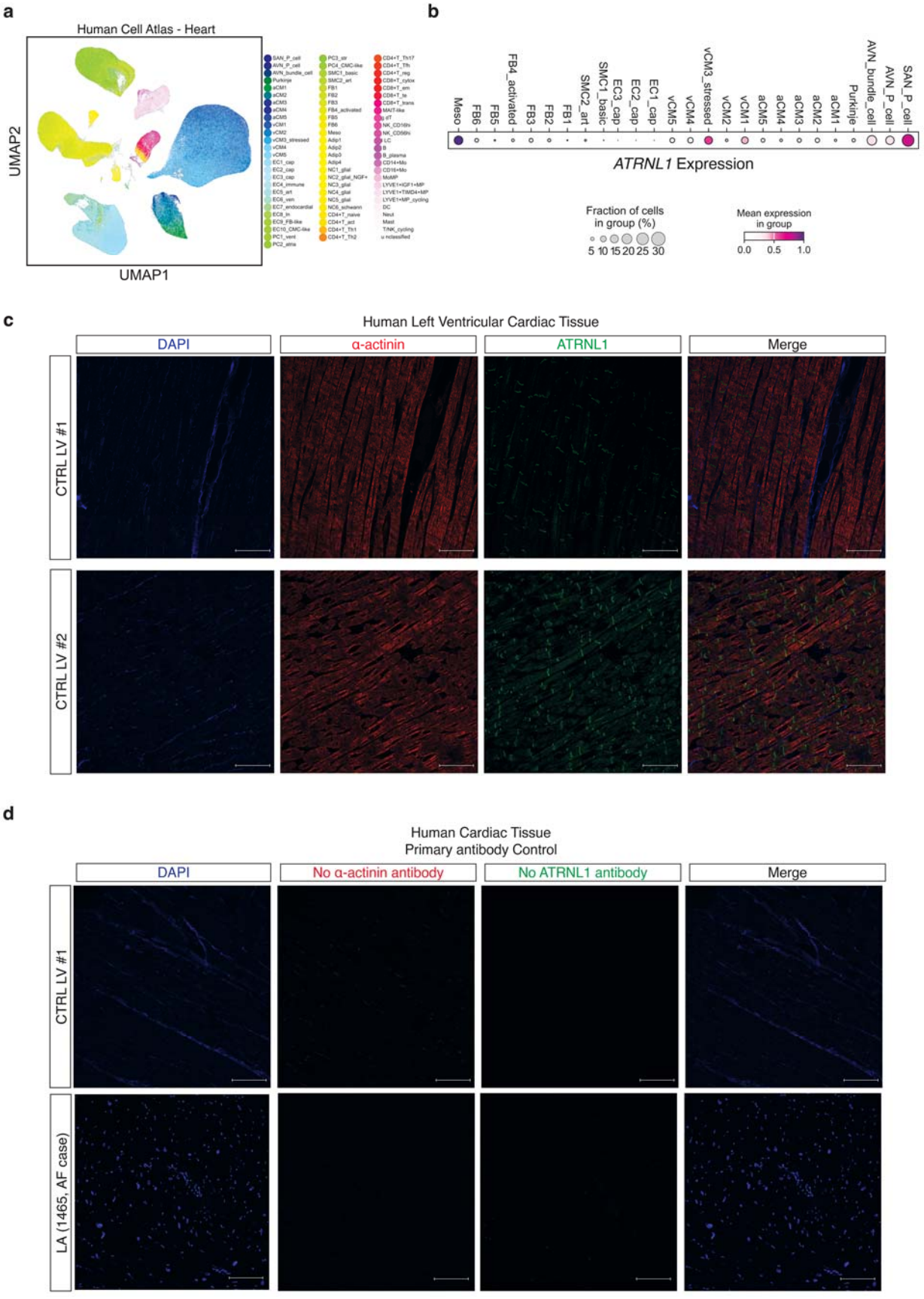
Authors' response: For standard RNA-seq and pseudobulk RNA-seq analysis, sex will always contribute as a strong source of transcriptional variation[2]. To clarify our statement, we have added the following to the results:

"The greatest source of variation separating samples, regardless of disease status, was sex, as genes from X and Y chromosomes were included in our analysis."

4. Is ATRNL1 expressed in ventricular cardiomyocytes?

Authors' response: We were also curious to know whether ATRNL1 was expressed in other regions of the heart. To look at the expression of *ATRNL1* across different anatomical regions of the human heart we investigated the Human Heart Atlas (**Response Figure 4a**)[11]. We found that *ATRNL1* was expressed in both ventricular and atrial cardiomyocytes (**Response Figure 4a**, and **4b**). Further, we found that *ATRNL1* was expressed in pacemaker cells and epicardial cells (mesothelial). To confirm the presence of ATRNL1 in ventricular cardiomyocytes we performed immunofluorescence staining in non-failing LV samples and observed that ATRNL1 was localized at the intercalated disk in ventricular cardiomyocytes of the LV (**Response Figure 4c**). We have added **Response Figure 4** to the manuscript as an additional Supplementary Figure (**Supplementary Figure 4**) and added the following to the results:

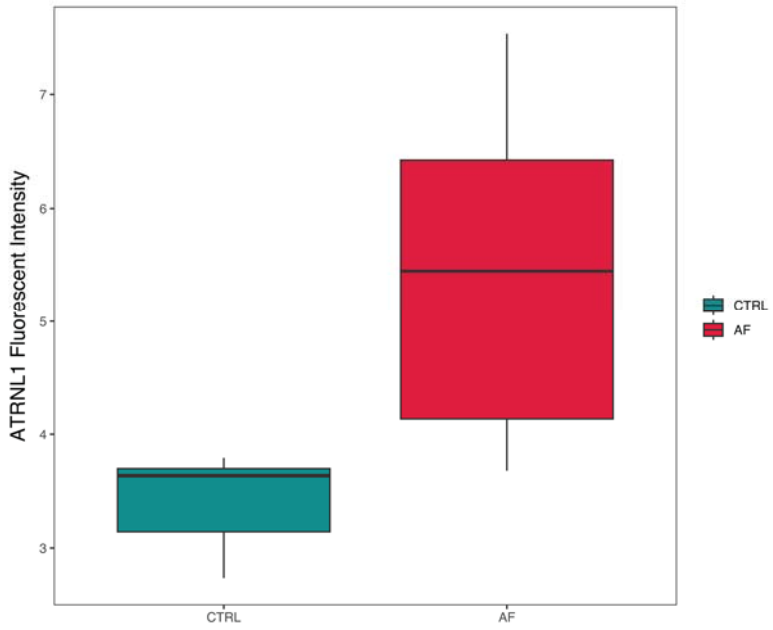
"To assess at the expression of *ATRNL1* across different anatomical regions of the human heart we investigated the Human Heart Atlas[11]. We found that *ATRNL1* was expressed in both ventricular and atrial cardiomyocytes (**Supplementary Fig. 4a**, and **4b**). Further, we found that *ATRNL1* was expressed in pacemaker and epicardial cells (mesothelial). To confirm the presence of ATRNL1 in ventricular cardiomyocytes we performed immunofluorescence staining in non-failing LV samples and observed that ATRNL1 was localized at the intercalated disk in ventricular cardiomyocytes of the LV (**Supplementary Fig. 4c**)."


a
Human Cell Atlas - Heart
UMAP2
UMAP1
b
Meso
FB6
FB5
FB4_activated
FB3
FB2
FB1
SMC2_art
SMC1_basic
EC3_cap
EC2_cap
EC1_cap
vCM5
vCM4
vCM3_stressed
vCM2
vCM1
aCM5
aCM4
aCM3
aCM2
aCM1
Purkinje
AVN_bundle_cell
AVN_P_cell
SAN_P_cell
ATRNL1 Expression
Fraction of cells in group (%)
5 10 15 20 25 30
Mean expression in group
0.0 0.5 1.0
c
Human Left Ventricular Cardiac Tissue
DAPI
α-actinin
ATRNL1
Merge
CTRL LV #1
CTRL LV #2
d
Human Cardiac Tissue
Primary antibody Control
DAPI
No α-actinin antibody
No ATRNL1 antibody
Merge
CTRL LV #1
LA (1465, AF case)

Response Figure 4. Tissue expression of ATRNL1.(**a**) UMAP embedding of Human Heart Cell Atlas colored by cell state. (**b**) Dot plot showing expression of *ATRNL1* across all common cardiac cell states (non-immune cells only).(**c**) Immunostaining of human left ventricular (LV) tissue. (**d**) Immunostaining of human left ventricular (LV) tissue and left atrial (LA) tissue controlling for primary antibody. Samples treated stained with DAPI and secondary antibodies only.

5. Please provide quantification of the immunostaining images shown in Figure 3.

Authors' response: We have quantified the fluorescent intensity in all control LA tissue (n=2 samples, and n=6 images), and LA tissue from AF patients (n=2 samples, and n=8 images) masked for cardiomyocytes (**Response Figure 5**). Overall, we observed high fluorescence intensity in hearts with AF compared to controls. These results are consistent with our snRNA-seq findings.



Response Figure 5. Quantification of ATRNL1 expression in cardiac tissue. Box plot of ATRNL1 fluorescence intensity from immunofluorescence analysis. Center line, median; box limits, upper and lower quartiles; whiskers,1.5x interquartile range.

6. It is not clear whether ATRNL1 exerts its effects through the long isoform, short isoform, or circRNA species. Does knockdown or overexpression of these isoforms of ATRNL1 lead to different outcomes? The impact on electrophysiology is not very detailed and restricted to immature iPSC-derived atrial cardiomyocytes.

Authors' response: We agree this is an interesting and important question. From our RNA-seq experiments in hESC-aCM cells we were able to partially disentangle the different functions of the long and short isoforms. With siRNA-1, which only targets the long isoform of *ATRNL1*, we observed a slightly different transcriptional response compared to an siRNA (siRNA-2) that targets both long and short isoforms (**Figure 4c-d**, and **4f-g**). However, we acknowledge that siRNA-based knockdown experiments are not the most experimentally precise way to decipher the different functions of the different *ATRNL1* isoforms. We were able to overexpress the short isoform in hESC-aCMs, which confirmed a role for the short isoform of ATRNL1 in promoting processes like glucose metabolism, the ER stress response, response to hypoxia, and MTORC1 signaling (**Figure 4h-i**). Further, it appears that short ATRNL1 expression also represses cell proliferation and regulates cell junction organization. However, we were not able to overexpress the long isoform as it's too long to express in a lentivirus. Consistent with the findings made in cancer studies[12], we found that the most prominent circATRNL1 isoform in CMs possess putative miRNA binding sites for genes involved in cell cycle progression, TGFB production, and glycolysis. Understanding the roles played by each isoform not only in human cardiomyocytes, but across all cell types that express *ATRNL1* would be immensely informative and will be our goal going forward with future studies.

From or immunofluorescence-based analysis of ATRNL1 in human cardiac tissue we found that it localizes along the longitudinal borders of CMs like the protein components of the intercalated disk. The antibody used for this analysis recognizes an antigen only contained in the long isoform. Thus, the long isoform of ATRNL1 is what we are detecting within or near the intercalated disk. New reagents that recognize each species of RNA specifically will need to be developed and validated for follow-up work.

We agree that the electrophysiology experiments in hESC-aCMs do have some limitations. And that there are alternatives (e.g. engineered heart tissues) which could improve the interpretation of our results and further establish the role played by ATRNL1 in cardiomyocytes.

To address the reviewers concerns we have added the following to the discussion:

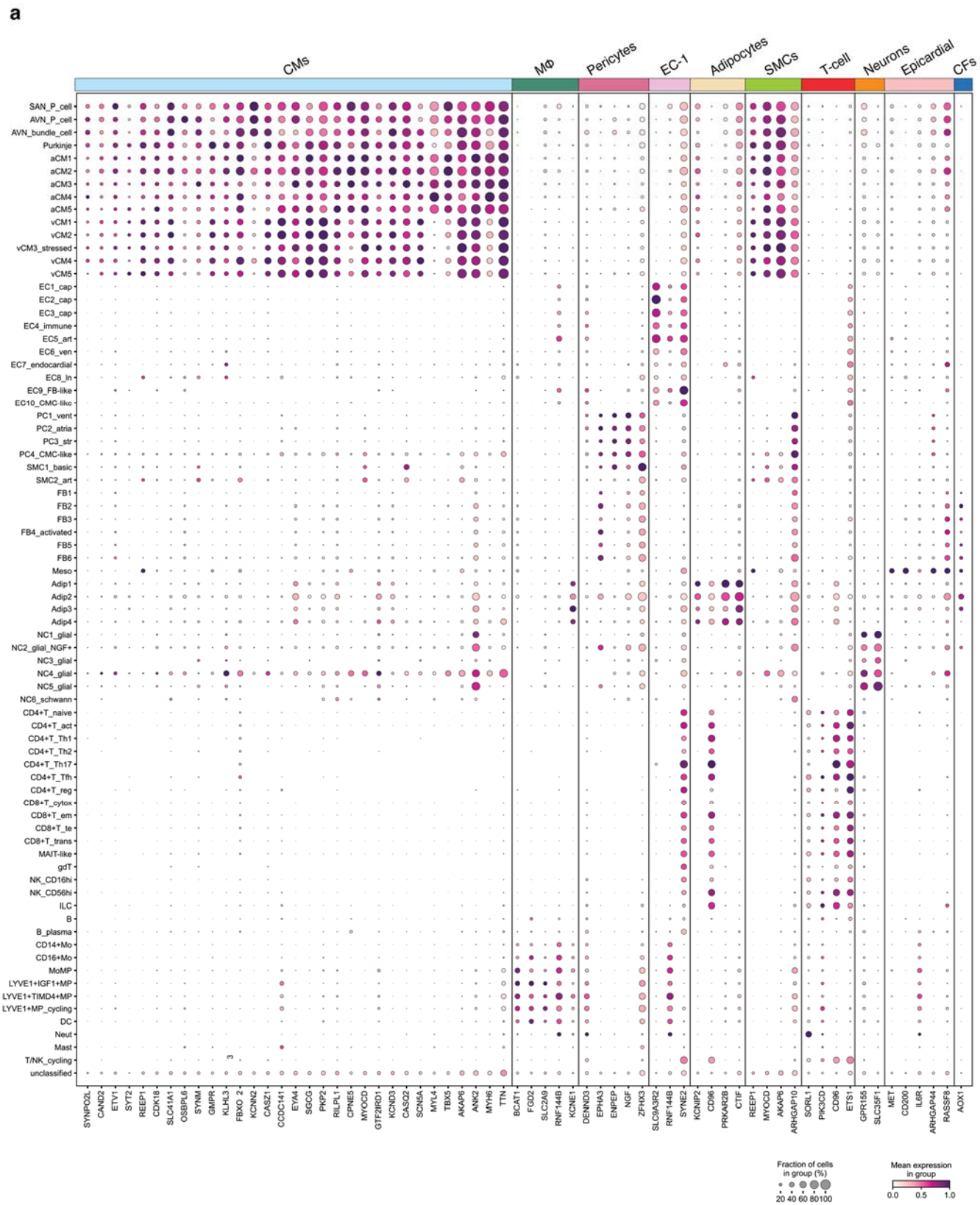
> "One limitation of our study is the use of hESC-aCMs, which are not as mature as adult human CMs in respect to features like sarcomere structure, Ca^{2+} kinetics, and ion channel density. Recent developments in 3D cardiac tissue technologies and methods have led to improvements in CM maturation and cardiac tissue modeling, including engineered heart tissue (EHT)[13,14]. Future work evaluating the physiological role of ATRNL1 in CMs will need to be conducted in EHTs or other 3D cardiac tissue models containing more mature and therapeutically relevant cell populations."

7. Cardiac cell type specific expression of genes associated with candidate atrial fibrillation loci should be validated across different snRNAseq datasets. For example, from this manuscript and published available data.

Authors' response: We thank the Reviewer for their suggestion. We have evaluated the expression of the candidate atrial fibrillation loci across the cell states found in the Human Heart Cell Atlas (**Response Figure 6**)(https://www.heartcellatlas.org/)[11]. This dataset includes cells from 6 different anatomical regions of the heart and contains 704,290 cells and nuclei combined (**Response Figure 4a**). We found that our data matched their cell state/type annotations very well. The CM AF loci were expressed across the different subtypes of atrial and ventricular CMs, as well as pacemaker cells, and Purkinje cells. Some candidate genes did display differential expression among the CM clusters, including known atrial-enriched transcripts like *TBX5* and *MYL4*[15]. Further, with the large enrichment of immune cells in their dataset we observe an increase in resolution among our immune cell-enriched candidate AF loci. For example, the macrophage marker gene *RNF144B* also exhibited high expression in neutrophils. RNF144B is an E3 ubiquitin ligase known to promote lipopolysaccharide-inducible IL-1b expression and inflammasome priming in human macrophages[16]. The interconnection of inflammasome activation and AF have been recently established[17]. Overall, the Human Heart Atlas data agrees with our cell type annotations and provides improved cell-type-specific expression information for the identified AF GWAS loci.

We have added **Response Figure 8** as **Supplementary Figure 8** and incorporated the following to the results:

> "We also identified several known candidates for AF that have been previously described, including *TTN*[18]**,** *ZFHX3*[19], TBX5[20] and *KCNIP2*, which supports the validity of our approach. Next, we evaluated the expression of all candidate AF loci across the cell states found in the Human Heart Cell Atlas (**Supplementary Fig. 8a**)[11]. Our data matched the Heart Cell Atlas cell state annotations well. The CM AF loci were expressed across the different subtypes of atrial and ventricular CMs, as well as pacemaker cells, and Purkinje cells. Some candidate genes did display differential expression among the CM clusters, including known atrial-enriched transcripts like *TBX5* and *MYL4*[15]. Further, with the large enrichment of immune cells in the Human Heart Atlas dataset we found improved resolution among our immune cell-enriched candidate AF loci. For example, the macrophage marker gene *RNF144B* exhibited high expression in neutrophils. RNF144B is an E3 ubiquitin ligase known to promote lipopolysaccharide-inducible IL-1b expression and inflammasome priming in human macrophages. The interconnection of inflammasome activation and AF have been recently established[16]. Overall, the Human Heart Atlas data agrees with our cell type annotations and provides improved cell-type-specific expression information for the identified AF GWAS loci."



Response Figure 6. Expression of AF GWAS loci from Human Heart Cell Atlas. (**a**) Dot plot displaying the mean expression of AF GWAS genes across all cell states

identified by the Human Heart Cell Atlas. Top, bars are colored according to the clusters presented in our study (**Figure 6c**). SAN, sinoatrial node; AVN, atrioventricular node; P cell, pacemaker cell; aCM, atrial cardiomyocyte; vCM, ventricular cardiomyocyte; EC, endothelial cell; PC, pericyte; SMC, smooth muscle cell; FB, cardiac fibroblast; meso, mesothelial cell (epicardial cell); Adip, adipocyte; NC, neural cell; NK, natural killer cell; Mo, monocyte; MP, macrophage; DC, dendritic cell; Neut, neutrophil.

Reviewer Three

In this manuscript by Hill et all, the authors perform snRNAseq on LA tissue from AF patients and identify a role of ATRNL1 in AF. I would like to congratulate the authors on the work performed and the well written paper. However, I do have some concerns regarding especially the patient populations included and perhaps the lack of a clear difference in phenotype (expect from the AF), which rather appears more or less to be a form of "lone AF", which no longer is an accepted form of AF. I have the following comments:

Authors' response: We thank the Reviewer for their kind words and thoughtful comments. The population of patients profiled here represents an advanced age group over 60 (mean age > 65 for both males and females) affected predominantly with permanent AF.

1. Supplementary table 1 should include more summarized date and statistical analysis on the patient characteristics that is also provided for the individual patients in Supplementary table 1 including for instance LVEF, LA size, co-comorbidities and other relevant information. Although it is very nice to have the individual data provided it is difficult to get a complete overview from by the individual values.

Authors' response: In response to comments from Reviewers 1 and 2, we have cleaned up and updated **Supplementary Tables 1 and 2** to include more relevant information on these patients, where available. Unfortunately, LA size information was not available for these patients.

We focused on patients with permanent AF to better control for the many diverse etiologies associated with this common arrythmia. In response to comments from all Reviewers, we have cleaned up and updated **Supplementary Tables 1 and 2** to reflect all the information that was available regarding time of diagnosis in relation to transplantation. To describe the patient cohorts with more clarity we have updated the Results to read as follows:

> "To characterize the cellular and molecular characteristics associated with AF, we performed snRNA-seq on samples from the LA of human patients with AF who were not in heart failure (n=17) as well as non-AF controls (n=19)(**Supplementary Table 1**, and **Supplementary Table 2**). We selected patients without heart failure to better control for the many diverse etiologies associated with this common arrythmias. Overall, our controls (CTRL) were 63 % female, the mean age was 68 (SD = 7.5), 25% had a history of taking beta-blockers, and 6% had taken anticoagulants. The AF cases were 61% female, the mean age was 66 (SD = 8.2), 56% had a history of taking beta-blockers, and 44% were being administered anticoagulants."

2. I have a hard time figuring out what the reason for the occurrence of AF is in the AF cohort is? There is always a reason for AF to occur, but the patient cohorts seem very similar. Is it a genetic component only then? Or are there co-comorbidities the authors do not have the information on that could be the underlying cause. Please elaborate.

Authors' response: This is a great question, and important to acknowledge for the conclusions made from this study. Here, we are evaluating patients predominately diagnosed with permanent AF without heart failure. Likely these patients are at risk due to unknown genetic or environmental contributions; however, we are not sure of their etiologies and are evaluating in an unbiased manner based on clinical diagnosis of AF. Single cell studies biased on genetics do have value[21], however, they don't represent a normal sample size of the population of patients afflicted with AF. Future studies aimed at analyzing prioritized cohorts with known etiologies of AF are certainly a logical next step toward further characterizing the molecular features of this disease.

3. Are there any information available on the AF history for these patients?

Authors' response: Thank you for raising this important point. In response to this quesry, we performed a manual chart review and when available, we added more information about the history of AF. We have included this data in the "other relevant information" in **Supplementary Table 2**. While this information was not available for all patients, all our AF cases were confirmed to have AF, with no heart failure.

4. The differential abundance testing identified not any major shifts in cell type composition between the SR and AF cohorts. However, I am not sure if I understand the conclusion that the Masson's Trichrome staining confirms the finding of similar tissue composition between the SR and AF cohorts. The staining only provides information on the collagen content and the volume of cardiomyoctes, but does not provide any information on the other cell types. For this purpose the authors would need to perform additional immunostainings with specific cellular markers.

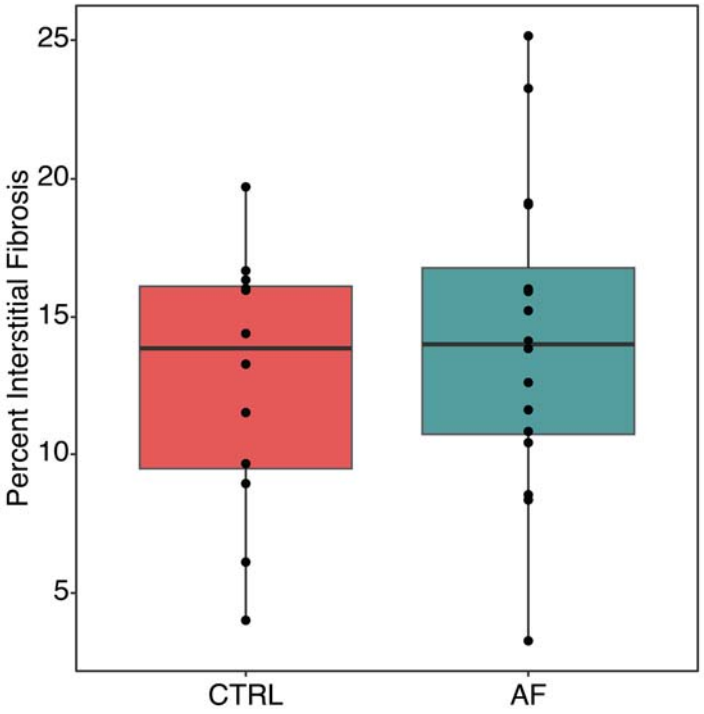
Authors' response: While we agree that immunostaining is a valid method for evaluating tissue composition, it's beyond the scope of this study. We also agree that Masson's Trichome staining doesn't fully support all changes in tissue composition apart from fibrosis, fat deposition, and muscle tissue. To our knowledge, snRNA-seq is a valid method that doesn't rely on the specificity of antibodies for profiling cardiac cell diversity and tissue composition. However, we do agree that it's not without its own bias

and comes with several limitations as we have mentioned in previous snRNA-seq studies[1,2,22].

5. Have the authors quantified the fibrotic content in the tissue? Usually more fibrosis is reported in AF patients, which is a change in tissue composition.

Authors' response: We were able to successfully section and carry out Masson Trichrome staining on 28/34 samples for this study and quantified the amount of total (including the epi- and endocardium) and interstitial fibrosis found in these samples (**Response Figure 7**). We observed similar results to the snRNAseq data, with no marked or significant changes in the amount of interstitial fibrosis in AF vs. controls.

Quantification of fibrosis was carried out using ImageJ in a blinded fashion. Trichrome staining images were converted to RGB stacks and thresholded for the total area of the section. The total area of fibrosis was measured using thresholding, using a color image as a guide to identify fibrotic regions. Areas of tissue that were folded, or incorrectly identified as fibrosis were subtracted from the total fibrosis measurements. If present, the area of the epicardium and endocardium were also measured. To calculate total fibrosis, the total area of fibrosis (minus any incorrectly regions) was divided by the total area of the tissue and a percentage was calculated. To calculate the interstitial fibrosis only, the epicardial and endocardium measurements (if present) were subtracted from the total fibrosis and the remaining area was then divided by the total area of the tissue minus the epi and endocardium.



Response Figure 7. Quantification of Interstitial Fibrosis in Human LA Tissue. Boxplot of percent interstitial fibrosis quantified from Masson's Trichrome stained LA tissue sections. Boxplot represented as: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

6. A recent paper (PMID: 37440641) in which the authors performed scRNA-seq of AF patients the MP/DC cluster expanded twofold, whereas endothelia and mural cells decreased in frequency. Could the authors please comment on the findings of this paper compared to their findings?

Authors' response: Indeed, members of our team were a part of that collaboration which stemmed from looking at patients with mitral valve regurgitation (MR) and persistent AF undergoing heart surgery. They were able to obtain LA appendage samples from these patients (n=7) as well as controls without AF (n=5) and then enzymatically digest the cardiac tissue, FACS sort live cells, and finally perform scRNA-seq. So, both the patient population and methodologies differ greatly between their study and ours. Further, they only performed a student's t-test for changes in cell composition, which is not a well-supported model for proper differential abundance testing of scRNA-seq data. Although, the HOMER mouse model that they characterized is quite compelling and does present with a phenotype more commonly associated with persistent AF.

To address these results in relation to our study we have added the following to the discussion:

> "Further studies on a broader range of AF cases classified from mild to severe (e.g., AF with heart failure) could help identify subtle changes in intercellular signaling and tissue composition that accompany AF. Interestingly, a recent scRNA-seq study conducted on patients undergoing open heart surgery with AF and mitral valve regurgitation found evidence of macrophage expansion and a commensurate decrease in endothelial and mural cell composition in the left atrial appendage compared to controls[23]. The flow cytometry-based cellular enrichment approach taken in that study may be an ideal approach for characterizing the changes in all non-cardiomyocytes from living tissue, especially rare immune cells that our nuclei-based profiling approach may not properly account for."

7. Why do the authors only have access to LA tissue and not RA tissue? It could have been interesting to see if the same results were obtained in RA. Furthermore please specify from what exact location of the LA the samples were taken (LA posterior wall, LA free wall, RA appendage or somewhere else)? Especially the LA posterior wall is important for AF induction and would have been interesting

to investigate as well.

Authors' response: We agree that profiling of the RA would also be interesting, and certainly we will address the RA in our future work. Samples for this study were collected from either the anterior wall of the LA near the appendage, or from the posterior wall near the annulus (see above; Reviewer 2, response 1). We have added the following to the discussion to highlight the need for future work to include the RA:

> **"**Importantly, here we only focused on the LA, an important tissue for evaluating AF. In AF the right atrium (RA) is also affected, however, the cellular and molecular changes to the RA in AF are not well understood. The pulmonary vein (PV) is known to play a role in the pathogenesis of AF, and catheter ablation of foci in the PV is an effective treatment for arrhythmias[24,25]. Future studies interrogating the single cell characteristics of the PV-LA junction, and the RA are also warranted.**"**

8. Please provide negative controls for the immunohistochemical staining performed in atrial tissue.

Authors' response: We have provided negative controls from both the LA and LV (**Response Figure 4d**). This data has been added to our manuscript as **Supplementary Figure 4**.

9. I agree that the Cx43 and ATRNL1 co-stain in the same region, but this does not confirm a co-localization of the proteins. I would be careful concluding this as an overlap in fluorescence does not necessarily indicate co-localization. Additionally, I would suggest the authors to perform Pearson's correlation coefficient (PCC) as a statistic for quantifying co-localization.

Authors' response: We thank the Reviewer for their comment. And we agree that to prove true co-localization of two proteins requires more sophisticated biochemical assays and higher resolution microscopy. We have modified the text to simply say 'localize' instead of 'co-localize. Further, we carried out Pearson's correlation coefficient analysis on all images stained for ATNRL1 and Cx43 (2 AF, 2 NF sections, 4 images per section) and averaged the Pearson's correlation coefficient to get an average of 0.87. We have added this information into the results and methods as follows:

> "**Results:** To confirm this subcellular localization, we co-stained the same LA tissues with the intercalated disc component connexin 43 (Cx43/GJA1) (Fig. 3b, Supplementary Fig. 3b). We found that ATRNL1 and Cx43 localize at

intercalated discs in human LA tissue samples (Pearson's correlation coefficient = 0.87).

Methods: Pearson's correlation coefficient measurements were carried out on images co-stained for ATNRL1 and Cx43 using the Image J plug in Just Another Colocalization Plugin[26]. Images were converted to 16 bit and ATRNL1 and Cx43 images were analyzed. The average of all analyzed images was reported in the results section."

10. I must disagree with the statement on line "272-273" regarding the SK channels stating: "KCNN2 and KCNN3 encode small conductance calcium-activated potassium channels. However,the role of these channels at steady state and during AF remains unclear". Numerous studies have both in animal models and in human tissue samples provided clear evidence of the role of SK channels in action potential repolarization in SR, while changes in trafficking and function of the channels as well as transcriptional and translational changes during AF also has been demonstrated. Please include these information on the role of SK channels in SR and AF.

Authors' response: We appreciate the Reviewer's point and we have modified this section of the results as follows

"*KCNN2* and *KCNN3* encode small conductance calcium-activated potassium channels (SK). The channels are known to regulate cardiac excitability and are gated by changes in intracellular Ca^{2+} derived from the sarcoplasmic reticulum[27]. Importantly, SK channels are known to contribute to arrythmias, as it has been found that both gain- and loss-of-function of SK channels can increase AF susceptibility[27–30]. "

11. Line "301-303" states that the authors performed "in vivo and in vitro" immunoflouresence. Please clarify what you mean with in vivo immunofloresence? If the reference is to tissue staining the in vivo phrasing needs to be changed.

Authors' response: Thank you for catching this. We have changed to the following:

"We performed immunofluorescence analysis of human LA tissue and hESC derived cardiomyocytes, transcriptomic profiling of siRNA and overexpression cells, and electrophysiological phenotyping to characterize the role of ATRNL1 in AF and cardiomyocyte physiology."

References

1. Simonson, B. *et al.* Single-nucleus RNA sequencing in ischemic cardiomyopathy reveals common transcriptional profile underlying end-stage heart failure. *Cell Rep.* **42**, 112086 (2023).
2. Chaffin, M. *et al.* Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* **608**, 174–180 (2022).
3. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
4. Koenig, A. L. *et al.* Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure. *Nat. Cardiovasc. Res.* **1**, 263–280 (2022).
5. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
6. Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 6876 (2021).
7. Lun, A. T. L. & Marioni, J. C. Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. *Biostat. Oxf. Engl.* **18**, 451–464 (2017).
8. Schmid, K. T. *et al.* scPower accelerates and optimizes the design of multi-sample single cell transcriptomic studies. *Nat. Commun.* **12**, 6625 (2021).
9. Vieth, B., Ziegenhain, C., Parekh, S., Enard, W. & Hellmann, I. powsimR: power analysis for bulk and single cell RNA-seq experiments. *Bioinforma. Oxf. Engl.* **33**, 3486–3488 (2017).

10. Gc, van A. *et al.* MicroRNA-18 and microRNA-19 regulate CTGF and TSP-1 expression in age-related heart failure. *Aging Cell* **10**, (2011).
11. Kanemaru, K. *et al.* Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801–810 (2023).
12. Chen, G. *et al.* Upregulation of Circular RNA circATRNL1 to Sensitize Oral Squamous Cell Carcinoma to Irradiation. *Mol. Ther. Nucleic Acids* **19**, 961–973 (2020).
13. Tani, H. & Tohyama, S. Human Engineered Heart Tissue Models for Disease Modeling and Drug Discovery. *Front. Cell Dev. Biol.* **10**, 855763 (2022).
14. Hansen, A. *et al.* Development of a drug screening platform based on engineered heart tissue. *Circ. Res.* **107**, 35–44 (2010).
15. Wang, T. Y. *et al.* Human cardiac myosin light chain 4 (MYL4) mosaic expression patterns vary by sex. *Sci. Rep.* **9**, 12681 (2019).
16. Ariffin, J. K. *et al.* The E3 ubiquitin ligase RNF144B is LPS-inducible in human, but not mouse, macrophages and promotes inducible IL-1β expression. *J. Leukoc. Biol.* **100**, 155–161 (2016).
17. Xing, Y., Yan, L., Xu, Z., Ma, X. & Liu, J. The relationship between atrial fibrillation and NLRP3 inflammasome: a gut microbiota perspective. *Front. Immunol.* **14**, (2023).
18. Choi, S. H. *et al.* Association Between Titin Loss-of-Function Variants and Early-Onset Atrial Fibrillation. *JAMA* **320**, 2354–2364 (2018).
19. Gudbjartsson, D. F. *et al.* A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* **41**, 876–878 (2009).
20. Postma, A. V. *et al.* A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. *Circ. Res.* **102**, 1433–1442 (2008).

21. Reichart, D. *et al.* Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies. *Science* **377**, eabo1984 (2022).
22. Tucker, N. R. *et al.* Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466–482 (2020).
23. Hulsmans, M. *et al.* Recruited macrophages elicit atrial fibrillation. *Science* **381**, 231–239 (2023).
24. Haïssaguerre, M. *et al.* Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N. Engl. J. Med.* **339**, 659–666 (1998).
25. Mahida, S. *et al.* Science Linking Pulmonary Veins and Atrial Fibrillation. *Arrhythmia Electrophysiol. Rev.* **4**, 40–43 (2015).
26. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
27. Zhang, X.-D., Thai, P. N., Lieu, D. K. & Chiamvimonvat, N. Cardiac small-conductance calcium-activated potassium channels in health and disease. *Pflugers Arch.* **473**, 477–489 (2021).
28. Mahida, S. *et al.* Overexpression of KCNN3 results in sudden cardiac death. *Cardiovasc. Res.* **101**, 326–334 (2014).
29. Li, N. *et al.* Ablation of a Ca2+-activated K+ channel (SK2 channel) results in action potential prolongation in atrial myocytes and atrial fibrillation. *J. Physiol.* **587**, 1087–1100 (2009).
30. Ozgen, N. *et al.* Early electrical remodeling in rabbit pulmonary vein results from trafficking of intracellular SK2 channels to membrane sites. *Cardiovasc. Res.* **75**, 758–769 (2007).

Point-by-point Response

Reviewer #2 (Remarks to the Author):

The authors have significantly revised their manuscript and it is much improved.

Authors' response: We sincerely appreciate your help improving this manuscript.

If possible, it is recommended that the authors perform RNAscope to address expression of the ATRNL long, short, and circular RNA form in human atrial and ventricular cardiac specimens with and without AF. This information would be useful to the field moving forward.

Authors' response: We completely agree with the Reviewer that the *in situ* validation of each of these forms of ATRNL, in health and disease, and across the cardiac chambers will be a logical and helpful next step. In addition to the development of RNAscope probes for each isoform, we are also working on the development of an antibody targeting the short form of ATRNL, but these studies will extend well into the next year. Given the scope of these studies, we respectfully believe that such work extends beyond the current manuscript.