Peer Review File

Manuscript Title: Spatial multi-omic map of human myocardial infarction

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have deployed a multi-omits approach to generate a spatially-resolved resource of gene expression and chromatin accessibility. The authors are commended on the work, which will no doubt be useful to many. The question is whether or not there are truly new insights that inform the biology of myocardial infarction. Certainly the spatial data helps delineate the specific populations of cells, and the interesting diversity of their responses. Could there be more of a concerted effort to quantitatively compare control to MI samples for example. As it is the presentation is sample to sample, and a much more extensive analysis could be deployed across the samples and cell types. For example, what are differentially expressed genes between a normal cardiomyocyte and the two types that are seen in the border zone. Is the myocardial trans-ventricular gradient of ion channels retained? Can one build a quantitative GRN that shows the retiring of CM or fibroblast GRNs? As it is the very rich dataset is not exploited to its maximum efficiency and unfortunately remains superficial in nature.

I have some comments of a technical nature.

1. It would be useful for the reader to have a diagram of the location of the samples in the heart, perhaps as a cartoon. Are these all LV free wall (no), so then where with respect to the overall anatomy of the heart are they from?

2. The authors may want to use the co-integration feature of ArchR to help define their cell types.

3. Myod1 (mentioned at bottom of P8) is not a cardiogenic TF and is not expressed in the heart.

4. One cannot deploy trajectory analysis without a time course. A “trajectory” can be created by linking different cell types to one another, but this analysis give the false impression that there is a “path”. Rather it only provides a measure of a cellular continuum.

Referee #2 (Remarks to the Author):

Kuppe et al. present single-nuclei and spatial transcriptomics based insights into molecular and cellular events related to human myocardial infarction (MI). Generating a spatio-temporal cell atlas of cardiac remodelling processes, the authors suggest specific cell subpopulation of cardiomyocytes

and fibroblasts as well as specific signalling pathways to shape cell composition and architecture during injury, repair and remodeling. Through the integration of matched scATACseq data, the authors provide further mechanistic explanations of gene regulatory events and potential master regulators of the remodelling process. The study applies sophisticated technologies and state-of-the-art computational tools to integrate, analyze and interpret their datasets in respect to prior knowledge and to generate testable hypotheses of disease-related processes and cellular states. Overall, the study advances our understanding of the biology of MI. However, the below comments could help to go further beyond the previously known players and to generate a more data-driven and robust view of the chain of cellular and molecular events after MI.

Major comments:

1. Nuclei isolation conditions can introduce significant biases in cell type composition. How does the composition in this work compare to the recently published heart atlases that use comparable technologies and sampling strategies?
2. What was the rationale behind the cluster resolution that resulted in 24 transcriptional subpopulations? Is the resolution defined by data-driven approaches or related to the previous described biology of the heart. If the latter, the study might miss more subtle disease-specific events.
3. The variable cell type composition between the samples can also be of technical origin. Do the authors have support that missing cell types are absent or less frequent in the respective samples? This can be tested using the spatial information or complementary profiling methods. This is important to clarify as subsequent comparative analysis can be affected by technical nuclei sampling biases.
4. It is not clear what the authors mean by stating “our integrative single-cell analysis defined a consistent and non-redundant catalog of cell types that comprise the adult human heart”. Indeed, there is considerable variability between the samples and the fact that the authors used label transfer to annotate the snATACseq data makes the catalogue redundant between the modalities.
5. The Visium platform has a limited size of the capture area. How was the targeted area pre-selected?
6. Instead of relying on the top DEG for the spatial localization of cell types, the authors should consider the use of deconvolution tools to better predict the composition of each capture site. This should result in a more accurate and unbiased spatial composition of the control and diseased hearts.
7. In line, clustering the ST spots and subsequent annotation based on marker gene expression of selected cells type is a very hypothesis-driven strategy that could be strongly biased and exclude additional biological and potentially clinical relevant interpretation. Hence, the interpretation of the ST data (e.g. Figure 3) should follow a more data-driven strategy, such as spot deconvolution based on the matched snRNAseq datasets.
8. The interpretation and validation of the endothelial cell states is confusing. No detailed interpretation is added for Endo2 and the selection of ISH markers is not clearly related to previous described cell state markers. In general, the MISTy analysis is intriguing and an additional validation of the co-localisation between Fibro2 and Macro2 would further underline the validity of the predictions.
9. The authors argue that the lower gene count in ST and the reduced recovery of cell types from ischemic areas is due to an increased cell death. Although this is certainly expected given the

underlying pathology, do the authors find any molecular proof for such assumption?
10. It would be important to validate the zonation of cell type distribution in respect to the ischemic and scaring areas in multiple biological replicate samples (i.e. different donors), as a key finding in this work, and to extract general conclusion about composition dynamics. Some of the zonation events should be validated with alternative cell-resolution methods. This would be particularly interesting for the events of neo-angiogenesis and scaring-specific fibroblast populations.
11. Integration efforts of the snATAC data to extract regulatory and mechanistic information is very sparse and anecdotal. Here, the authors should make an effort to better explain regulatory activity responsible for the tissue remodelling.
12. The central regulatory role of NFE2L1 in cardiomyocytes 1 is intriguing, but entirely based on correlation analysis and prediction based on binding motifs. Since the authors highlight NFE2L1 as potential key regulator to derive disease-specific cardiomyocyte subpopulations, an experimental validation of the regulatory role of NFE2L1 in cardiomyocytes would be appreciated.

Minor:
This statement is too broad for the related section: “In summary, the data indicated distinct spatial gene regulation in response to the ischemia associated cell-death with gene regulation driving the acute cardiac injury response.”

Referee #3 (Remarks to the Author):

This study by Kuppe et al. investigates the cellular and local gene expression changes in response to ischemic injury in human heart samples at different time points after injury. They do so by performing single cell sequencing, ATAC sequencing and spatial transcriptomic profiling and different bioinformatic approaches to further mine the data.

While these are all state-of-the art methods and bioinformatic approaches that provide insightful data there are currently several issues with the experimental design and the presented data that make it difficult to determine the value and generalizability of the findings.

Major comments
- A major concern is the reproducibility and validity of the data. The authors have an n=1 for the different conditions (including control), except for the time point more immediate after myocardial infarction. Here they have tissue from both 2 and 5 days after infarction that represent the more immediate response after injury, however these samples seem to vary a lot from each other in cellular composition, local remodeling and gene expression changes and are also representing different regions.
- Is the sample for spatial transcriptomic take transmurally? The manuscript would benefit from a better description of the exact sample collection procedure.
- Where are the data for the ischemic zone from patient 3 in Extended Data Figure 1 and 3d?
- It is unclear how well the different datasets for patient 2 and 3 overlap. This should be clarified more as these are the only 2 patients from the same condition/group. Are the same cell cluster

found when comparing these 2 patients and are the same gene expression correlations found between the 2? It would be better for clarity to compare the same type of analyses for the different samples.
- It currently is unclear which portion of the control heart was taken for the analyses.
- The authors combine spatial transcriptomics on a 10um section with snRNA seq and ATAC seq from an adjacent portion of the heart. However, the results of these three give very different cellular compositions and gene expression profiles. The spatial transcriptomic for example shows an overrepresentation of cardiomyocytes (likely because of their size), while the cellular composition based on snRNA seq versus ATAC seq also gives a very different image (which can for example be seen in Extended Data Fig 3d). What would the Extended Data Fig 3d look like for ATAC seq only? If the dataset differ so much it is hard to generate an integrative molecular map as the authors indicate in the abstract.
- How many cells are roughly represented per spot in the spatial transcriptomics?
- The snRNA seq and ATAC seq data generated on samples taken 2 or 5 days after ischemic injury appear to be strongly influenced by the presence of cell death and only 4 cell types can be distinguished. However, in looking at extended Data Figure 6e all 4 cell types appear to be expressing TNNT2, a cardiomyocyte marker.
- While the validation studies in heart failure samples and functional follow-up studies support the relevance of the sequencing data, Runx1 has already been linked to TGFb signaling and myofibroblast differentiation and fibrosis.
- How do these snRNA seq data compare to previously published studies on human heart tissue?

Referee #4 (Remarks to the Author):

Tanevski et al report the results of deploying cutting edge spatial and single cell genomics analyses on precious, rare human ex vivo cardiac samples derived from individuals post myocardial infarction. Certainly the rarity of this sample set, and the promise of the technological tools used, intrigued and excited me about the possible biological insights offered. However, after reading the paper carefully I remain unsure what the authors have actually learned from all of this data generation. Is this paper reporting a biological discovery? I do not study heart biology, so I am genuinely uncertain, but the organization of the paper, and the details of the claims made (see specific points below) do not suggest that a clear, novel insight about myocardial infarction has emerged from this work. And if this biological insight has not been gained, and this manuscript is being presented as a reference atlas of MI, then greater effort needs to be made in codifying, organizing and releasing these data to the scientific public for consumption (for example, the generation of a web-based tool for plotting genes, performing individual analyses, etc).

In general, a lot of analyses presented are suggestive, but not clearly demonstrative, of the claims. In particular, there is a logical leap made by the authors between the results of MISTy and the conclusion that the genes identified are somehow directly mediating interactions between cell types. The algorithm is uncovering spatial correlations at different length scales, with certainly could suggest a causal interaction, but may also simply by a correlative effect due to, for example, the developmental patterning of the tissue (certain cells get positioned closer to others, but do not

necessarily interact in a causal fashion). Although certainly many of the nominated interactions are intriguing, it's really hard to know what the reader should do with these specific examples, without some sort of perturbation in an animal or organoid model to provide support for causal mechanisms.

Here are some specific points where I identified novelty, but where additional follow up work might help to bolster it to a solid mechanistic hypothesis:

One very intriguing insight the authors report is the presence of additional molecular/cell state heterogeneity in acute MI tissue beyond what can be seen histologically (Fig 4). This seems like a great example of how especially the spatial data was able to nominate additional molecular processes beyond what can be observed by conventional pathology, but the analysis ends at suggestive correlations. Can the authors take this a step further, and perhaps nominate a small number of markers for these heterogeneous states, and perform immunohistochemistry on a larger set of cardiac MI samples (for example FFPE samples), more deeply explore whether there are diagnostic, prognostic, or other clinically relevant implications for these different identified cell states?

The analysis presented does not specifically point towards a graded, pseudotime-ordered progression in the fibroblasts. In the UMAP in Fig 6a, I see an unclear trajectory structure, perhaps with branching occurring, but it is also possible the pseudotime model is overfitting and the actual gene expression landscape is far more complex (multiple distinct populations, or multiple transitions occurring within multiple populations). The heat map presented would also appear to support this—I do not see a whole lot of marker overlap, but rather what appears to be three rather distinct populations with few examples of cells in transition. What additional evidence is there to support that these fibroblasts are all progressing along a single trajectory?

The final insight of the paper, for which there is one in vitro validation experiment performed—is the nomination of RUNX1 as a key effector of myofibroblast differentiation. Again, I am not in this field, but a quick google search for RUNX1 and myofibroblast did reveal literature that suggests this insight is not entirely novel. For example, how does the observation made by the authors meaningfully differ from the work reported in PMID 32341028 and PMID 25313057?

Smaller points:

- Some cardiomyocytes are multinucleated. Can the authors comment on how snRNA-seq and snATAC-seq might be affected by the (likely) additional correlation between nuclei derived from the same cell?
- Have the authors tried localizing cardiac-associated GWAS signals to their clusters and spatial data? This might be an interesting way of leveraging the human data in away that is unique, since it may be much harder to make credible conclusions about this from existing mouse datasets.

Author Rebuttals to Initial Comments:

R1: NATURE 2020-11-21461

We would like to thank all the reviewers for their careful assessment of our manuscript which we feel has helped us significantly to improve the impact of our work. We have taken all the comments into account and present a substantially modified and improved manuscript, providing extensive new data in response to the reviewers comments.

In response to the reviewers' comments, we have now included the following salient additional data and analysis:

1. We have increased the multi-omic dataset considerably to a total of 28 human multi-omic myocardial datasets of patients following acute myocardial infarction including single nuclei (sn) RNA-seq, snATAC-seq and spatial transcriptomics (Visium) spanning 196,497 nuclei (snRNA-seq), 46,086 nuclei (snATAC-seq) and 91,517 spots (spatial transcriptomics) (New Fig.1; New Extended Data Fig. 2a-c).
2. We additionally added three novel, unpublished snRNA-seq samples of human acute myocardial infarction as external reference datasets (including 19,722 cells; New Extended Data Fig. 4f-g).
3. We included single-cell RNA-seq (scRNA-seq) data from a myocardial infarction time-course lineage tracing experiment of PDGFRβCreER;tdTomato mice (n=4), time points 0, 4, 7 and 14 days after MI (including 32,852 cells) to compare fibroblasts states between human and mouse and verify the directionality of the trajectory analysis of myofibroblast differentiation (New Extended Data Fig. 12i-l).
4. We revised the snRNA-seq and snATAC-seq data integration to identify major cell types using state-of-the-art methods including Harmony and ArchR (as suggested by the reviewers). Instead of annotating the cell types sample by sample as in our previous analysis, for each modality we first integrated the data from all samples and then clustered the cells. Clusters were annotated independently for snRNA-seq and snATAC-seq. To validate major cell-type annotations, we performed cross-modality comparison between snRNA-seq and snATAC-seq data within our samples, and between the human heart cell atlas that profiled healthy samples and an external reference snRNA-seq dataset of ischemic samples (point 2 above) (New Extended Data Fig. 4a,e).
5. We now provide extensive descriptions of the spatial transcriptomics slides that include per spot: i) cell-type compositions (estimated using the deconvolution method cell2location (Kleshchevnikov et al. Nat Biotech 2022, PMID: 35027729)), leveraging the paired snRNA-seq data, as suggested by reviewer #2, and ii) signalling pathway activities, transcription factor (TF) binding activities (mapped from snATAC-seq data) and GWAS signals (New Fig. 1h, New Extended Data Fig. 5j) (as suggested by reviewer #4). Moreover, we used computational modelling to identify spatial cell-type dependencies and relations to signalling pathway activities (New Fig. 2e-g, New Extended Data Fig. 7a-c). Finally, we described cardiac "niches", representing structural building blocks that are shared between different slides and could facilitate patient comparison. These niches were generated by clustering spatial transcriptomics spots using cell-type compositions (New Fig. 2 a-d) or gene expression profiles (New Fig. 3) (referred to in the text and revision notes as molecular niches) (New Fig. 1h-j; New Fig. 2a-i).
6. We revised the comparative analysis of the different sampled regions and patient groups at the molecular, compositional, and spatial level. i) We contrasted the

compositions of each major cell-type across patient groups with information provided by all omic layers (New Fig. 2i). ii) We contrasted the compositions of niches from spatial transcriptomics and demonstrated consistent remodelling events characteristic of fibrotic and ischemic samples (New Fig. 2j-o). iii) We identified molecular differences in myogenic regions that differentiated control samples from border zones and remote zones of myocardial infarction patients (New Fig. 3g-i).

7. We have revised our integration strategy to define cell-states of cardiomyocytes, fibroblasts, endothelial and myeloid cells (New Fig. 4-6). Compared to our previous analysis, we here first integrated the data from all samples and modalities for each of the major cell types mentioned above and then performed sub-clustering analysis (New Fig. 4a; New Fig. 5a; New Fig. 6a and New Fig. 6j). We quantified the importance of cell microenvironments in the prediction of disease relevant cell-states from spatial transcriptomics and performed in situ hybridization validations of the cell-state zonations in an independent patient cohort (235 images from 43 patients) (New Fig. 4c; New Fig. 6n; New Extended Data Fig. 9d; New Extended Data Fig. 14g, h, i, and New Extended Data Fig. 15g).
8. We built quantitative enhancer-based gene regulatory networks (eGRN) using the integrated snRNA-seq and snATAC-seq data by combining information from transcription factor (TF) binding activity, TF expression, enhancer-to-promoter links, and target gene expression for both cardiomyocytes (New Fig. 4h-j) and fibroblasts (New Fig. 6g-h) (as suggested by reviewer #1). We identified important putative regulators (i.e., TFs) using network analysis. We were able to map predicted TFs and target genes into space to support their role in cardiac remodelling and fibrogenesis.

Referee comments:

Referee #1:
The authors have deployed a multi-omits approach to generate a spatially-resolved resource of gene expression and chromatin accessibility. The authors are commended on the work, which will no doubt be useful to many. The question is whether or not there are truly new insights that inform the biology of myocardial infarction. Certainly the spatial data helps delineate the specific populations of cells, and the interesting diversity of their responses. Could there be more of a concerted effort to quantitatively compare control to MI samples for example. As it is the presentation is sample to sample, and a much more extensive analysis could be deployed across the samples and cell types. For example, what are differentially expressed genes between a normal cardiomyocyte and the two types that are seen in the border zone. Is the myocardial trans-ventricular gradient of ion channels retained? Can one build a quantitative GRN that shows the retiring of CM or fibroblast GRNs? As it is the very rich dataset is not exploited to its maximum efficiency and unfortunately remains superficial in nature.

We thank the reviewer and appreciate the overall positive comments. We agree that a comparison between the different time points and sample groups is necessary to gain new insights into the biology of human myocardial infarction and the different cell populations involved in the heart remodelling processes. For this reason, we have increased the number of tissue samples in this study from initially 8 to 28 from 20 patients over four different cardiac tissue regions, as well as control samples (control = 4, ischemic region = 12, border zone = 3, remote zone = 6, fibrotic zone = 6) and additionally 3 snRNA-Seq datasets from human acute myocardial infarction for cross-validation. 28 of these samples have a multi-omic profiling with single nuclei (sn) RNA-seq, snATAC-seq, and spatial transcriptomics (New Fig. 1a, New Extended Data Fig. 1-3). In addition, we also included scRNA-seq from a myocardial infarction time-course lineage tracing experiment of PDGFRβ-reporter mice at distinct time points after MI which was used in our trajectory analysis of myofibroblast differentiation.

This extended dataset allowed us to describe heart remodelling processes by performing quantitative comparisons of different sample groups at the compositional, molecular, and spatial level, as the reviewer suggested. We have included these results in two new sections: “Spatial and compositional variation of histological human cardiac tissue classifications” (New Fig. 2) and “Molecular variation of human cardiac tissue following acute myocardial infarction” (New Fig. 3).

Additionally, we have redefined the sub-clusters for several major cell-types including cardiomyocytes (New Fig. 4), endothelial cells (New Fig. 5), fibroblasts, and myeloid cells (New Fig. 6) by integrating snRNA-seq and snATAC-seq from all samples and associated them with the different disease stages. We mapped the different functional states of cell-types to spatial transcriptomics to investigate their spatial relationship with other major cell types and their changes in different patient groups (New Figs. 4k-o, 5e-f, 6l). We included these results in three new sections: “Identification of disease specific cardiomyocyte states ” (New Fig. 4), “Analysis of endothelial cell heterogeneity at spatial resolution ” (New Fig. 5), and “Spatial organisation of fibro-myeloid cell states in cardiac remodelling” (New Fig. 6).
Regarding cardiomyocytes functional states, we describe 5 different cell-states with differential compositions between samples. The top 10 differentially expressed genes of the newly

described cardiomyocyte (CM)-states (vCM1-5) are shown in New Extended Data Figure 9b and the complete results are provided in New Supplementary Table 10. Among other findings, we observed a higher correlation between ion-channel related gene expression and the marker expression of "healthy" vCM1 gene markers compared to the expression of "stressed/failing" vCM3 (New Extended Data Fig. 9e-g).

As suggested by the reviewer we now built quantitative enhancer-based gene regulatory networks (eGRN) using the integrated snATAC-seq and snRNA-seq data for both cardiomyocytes (New Fig. 4h-j) and fibroblasts (New Fig. 6g-h). This allowed us to describe regulators of different cell-states found in these major cell-types. Moreover, we were able to map the predicted regulators and target genes to spatial transcriptomics and could confirm that these regulatory programs drive tissue patterns associated with cardiac remodelling and fibrogenesis.

We have reorganised the manuscript so that the presentation of results not only highlights the richness of the data but also shows how the different combination of omics technologies and resolutions can help describe myocardial remodelling and show the relations between celltype location, organisation and function.

I have some comments of a technical nature.

1. It would be useful for the reader to have a diagram of the location of the samples in the heart, perhaps as a cartoon. Are these all LV free wall (no), so then where with respect to the overall anatomy of the heart are they from?

We agree with the reviewer on this important point. We have now increased the overall datasets of our atlas including in total 31 datasets (28 multi-omic datasets of snRNA-seq/snATAC-seq/Visium and 3 snRNA-seq) of human myocardial infarction. 20 of these datasets were generated from the left ventricular (LV) free-wall and 11 from LV-apex. We have now included a diagram of where the samples have been taken as suggested by the reviewer and additional macroscopic and microscopic images of all samples from which these were available (New Extended Data Fig. 1). Additionally, we provided all available clinical covariate data including location of the biobank, tissue localization, infarct location and a detailed pathologists' description of the tissue annotation in New Supplementary Table 1 of all H&E stainings of the previous and added spatial transcriptomics datasets.

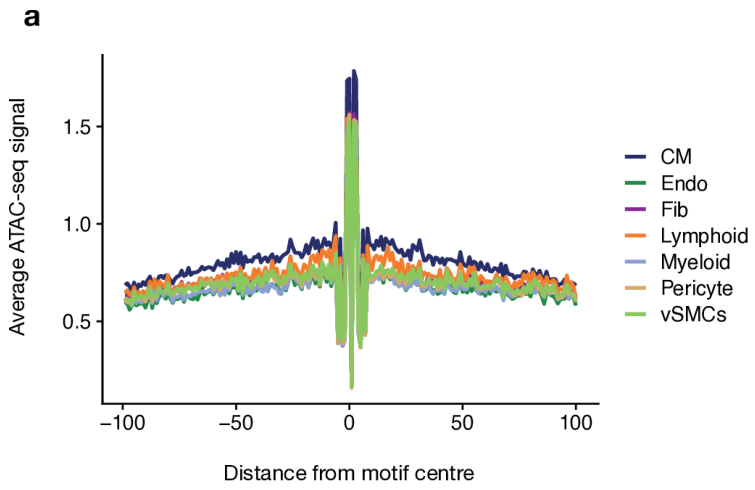
2. The authors may want to use the co-integration feature of ArchR to help define their cell types.

We appreciate this suggestion and have now used ArchR (Granja *et al.*, Nat Gen 2020, PMID: 33633365) to analyse our snATAC-seq data from all samples. ArchR was not readily available when this project started (New Extended Data Fig. 4a for detailed workflow). Briefly, we used ArchR to filter the low-quality cells based on the number of unique fragments (>3,000) and transcriptional start site (TSS) enrichment score (>4). We also used ArchR to remove doublets for each snATAC-seq library. We used a non-negative factorization method, which was recently proposed by us (scOpen) for snATAC-seq dimension reduction (Li. *et. al.*, Nat Com 2021, PMID: 34737275). The data were integrated using the Harmony algorithm (Korsunsky, I. *et. al.* Nat Methods, PMID: 31740819) and clusters were obtained using graph-based

algorithms. To annotate the clusters, we identified marker genes based on gene activity scores estimated by ArchR. To further verify our annotations, we also integrated the snATAC-seq with snRNA-seq using the Seurat package and observed a high similarity across the modalities (New Extended Data Fig. 5b).

3. Myod1 (mentioned at bottom of P8) is not a cardiogenic TF and is not expressed in the heart.

We thank the reviewer for pointing this out. We previously selected Myod1 by first predicting the transcription factor binding sites (TFBSs) and then compared the chromatin accessibility of these TFBSs between different cell types as measured by snATAC-seq. Now, by using our extended dataset, we again observed that cardiomyocytes showed a higher chromatin accessibility around the binding sites of Myod1 compared with other cell types, consistent with our previous results (see Reply Letter Fig. 1a). However, since this approach was solely based on chromatin accessibility profiles, it was not able to distinguish TFs that have similar motifs, which can cause false positives. Indeed, we observed no expression of Myod1 in our snRNA-seq (only 10 cells expressed Myod1 from the unfiltered integrated data across all samples). For this reason, we now removed Myod1 from the revised manuscript. Also, our new approach for estimations of eGRNs considers both gene expression and chromatin accessibility of TFs, which avoids this issue.



Reply Letter Fig. 1. (a) Footprinting profile of Myod1 between different major cell types. The x-axis represents distance from the Myod1 motif centre and the y-axis represents the average snATAC-seq signal around the predicted binding sites of Myod1. Colours refer to different cell types.

4. One cannot deploy trajectory analysis without a time course. A “trajectory” can be created by linking different cell types to one another, but this analysis give the false impression that there is a “path”. Rather it only provides a measure of a cellular continuum.

We appreciate the suggestion by the reviewer and agree that in principle a trajectory analysis should only be done on time course and disease progression data. Indeed, a trajectory across cell types by itself does not suggest a direction (i.e., the assignment of start and end cells along

the predicted pseudotime can be arbitrary). The current manuscript now has two cellular continuum analyses, leveraging the clinical and temporal information: one involving cardiomyocyte remodelling and another including fibroblast to myofibroblast differentiation. We added multiple analyses and novel data to justify the direction of these analyses.

We have now increased the overall dataset to provide data from different human disease groups (myogenic group comprising control specimens, remote zone and border zone as well as ischemic and fibrotic specimens) and time points. Furthermore clinical data of these groupings, e.g. time of tissue sampling after initial symptoms of the myocardial infarction, allow us to better characterise MI disease progression (New Fig. 1a-b).

To find sub-clusters of cardiomyocytes (CM) and fibroblasts (Fib), we first integrated the single nucleus data from multiple modalities (i.e., snRNA-seq and snATAC-seq) and samples, and performed sub-clustering analysis. For cardiomyocytes, we identified five states (vCM1-5) and observed significant up-regulation of *ANKRD1 and NPPB* in vCM3 (New Fig. 4a-b). Since both *ANKRD1* and *NPPB* have been reported to be upregulated in the border zone after myocardial infarction in mice (Mikhailov, A. T., & Torrado, M. 2004, Int J Dev Biol PMID: 18956313; Hama, Norio, et al., Circulation 2020, PMID: 7664440), we annotated this sub-cluster as "stressed" CM state. Comparison of cell proportions (including snRNA-seq and snATAC-seq data) revealed that this "stressed" ANKRD1+/NPPB+ vCM3 was significantly enriched in the ischemic samples, while the “healthy” vCM1 was enriched in the myogenic samples (New Fig 4d; New Extended Data Fig. 9h). Additionally, we identified a *NPPB*- but *ANKRD1*+ vCM2-state, which we annotated as “intermediate” based on expression of these marker genes and the fact that this state was associated with both myogenic and ischemic samples (New Extended Data Fig. 9e). Of note, *ANKRD1*, a transcription co-inhibitor and a member of the titin-N2A mechanosensory complex that translocates to the nucleus in response to stress, is known to be upregulated in cardiac failure and hypertrophy (Aihara et al., Hypertension 2000 PMID:10904011; Miller et al., J Mol Bio 2003 PMID: 14583192). We validated the three main cardiomyocyte states (vCM1-3) by in-situ hybridization using *TNNT2*, *NPPB* and *ANKRD1* (New Fig. 4c; New Extended Data Fig. 9d). Quantification of the in-situ hybridisation revealed in an independent cohort of 98 human cardiac tissue sections from 17 patients a significant enrichment in tissue following human myocardial infarction as compared to control tissue (non-transplanted donor hearts) providing evidence for the directionality of the cellular continuum (New Fig. 4c-d, New Extended Data Fig. 9d). We thus defined vCM1 as origin, vCM2 as intermediate, and vCM3 as the terminal cardiomyocyte state and inferred a cellular continuum from vCM1 to vCM3 using the function *addTrajectory* from ArchR based on the diffusion map representation (New Extended Data Fig. 10a).

For fibroblasts, we identified four cell states (Fib1-4) (New Fig. 6a). The Fib1 state demonstrated marker gene expression including *SCARA5* and *PCOLCE2*, which we have previously identified in myofibroblast progenitor cells of human kidney fibrosis (Kuppe et al., Nature 2021, PMID: 33176333). The Fib2 state revealed known marker gene expression reminiscent of myofibroblasts including *POSTN*, *TNC* and *COL1A1* (New Fig. 6b, New Extended Data Fig. 12a-b) and also showed the highest expression of extracellular matrix (ECM) genes (New Extended Data Fig. 12c). Spatial analysis of the marker gene for these two fibroblast states (New Fib1+2) revealed a mutually exclusive gene expression pattern across the spatial transcriptomic slides (New Fig. 6c; New Extended Data Fig. 12f). Compositional comparison between the defined patient groups revealed a significant enrichment of Fib1 in

myogenic samples and Fib2 (myofibroblasts) in ischemic samples. To further understand the differentiation of fibroblasts and to experimentally validate the findings of the computationally inferred pseudotime trajectories, we performed a transgenic mouse experiment using PDGFRβCreER;tdTomato inducible fate-tracing of all mesenchymal cells in myocardial infarction (LAD ligation) with subsequent scRNA-sequencing of tdTomato sorted (FACS) cells at different timepoints (day 0, 4, 7, 14) (New Extended Data Fig. 12i-l). We integrated and clustered the cells from all timepoints. To annotate the clusters, we integrated the mouse and human fibroblasts and performed label transfer using Seurat, uncovering three sub-clusters in mouse data (i.e., Fib1-3) (New Extended Data Fig. 12m). We observed that the Fib1 (*SCARA5*+) population decreased over time while the Fib2 (*POSTN*+) population increased and showed a higher ECM score in all time points (New Extended Data Fig. 12o-p). Based on these observations, we inferred a pseudotime trajectory from Fib1 to Fib2 (myofibroblasts) in the human samples which was further supported by an increased enrichment of extracellular matrix (ECM) score and of ECM-related biological Gene Ontology (GO) processes (New Extended Data Fig. 12q).

These analyses were used as the basis of the eGRNs (see above), which in turn explained molecular changes associated to cardiac remodelling and fibrogenesis detected in spatial transcriptomics. We additionally adjusted the paper accordingly to clarify that our main objective is to analyse cellular disease states in a cellular continuum and added the following sentences:

"[...] To this end, we paired the cells between snATAC-seq and snRNA-seq data and studied gene-regulatory changes along the cellular continuum from vCM1 to vCM3 (Extended Data Fig. 10a).[...]"

"[...] To precisely understand differentiation trajectories of fibroblasts and transfer this knowledge to the human data we performed inducible lineage tracing in mice using the pan-mesenchymal Cre driver PdgfrβCreER combined with scRNA-seq at different time points following myocardial infarction (Extended Data Fig. 12i-l). [...]".

Referee #2:

Kuppe et al. present single-nuclei and spatial transcriptomics based insights into molecular and cellular events related to human myocardial infarction (MI). Generating a spatio-temporal cell atlas of cardiac remodelling processes, the authors suggest specific cell subpopulation of cardiomyocytes and fibroblasts as well as specific signalling pathways to shape cell composition and architecture during injury, repair and remodeling. Through the integration of matched scATACseq data, the authors provide further mechanistic explanations of gene regulatory events and potential master regulators of the remodelling process. The study applies sophisticated technologies and state-of-the-art computational tools to integrate, analyze and interpret their datasets in respect to prior knowledge and to generate testable hypotheses of disease-related processes and cellular states. Overall, the study advances our understanding of the biology of MI. However, the below comments could help to go further beyond the previously known players and to generate a more data-driven and robust view of the chain of cellular and molecular events after MI.

We thank the reviewer for the overall very positive evaluation of our work and appreciate the suggestions made by the reviewer to further improve our manuscript.

Major comments:
1. Nuclei isolation conditions can introduce significant biases in cell type composition. How does the composition in this work compare to the recently published heart atlases that use comparable technologies and sampling strategies?

We thank the reviewer for this important question. We have now compared the cell-types of our intact myocardial tissue specimens (remote zones, border zones, and controls) with the reported cell-types of the single nuclei samples in the healthy human cardiac cell atlas by Litvinuková et al. (Nature 2020, PMID: 32971526) in terms of molecular profiles and cellular compositions (New Extended Data Fig. 4e). We observed a significant overlap between the marker genes inferred from these two datasets for the same cell types and a significant correlation (Pearson correlation = 0.88, *p*-value = 0.0016) between the median cell-type composition across patients of the two datasets (New Extended Data Fig. 4e). Additionally, we compared the cell-types of the ischemic samples with a novel validation dataset of three human ischemic samples from an independent cohort we now include in our analysis (New Extended Data Fig. 4f). Again, marker genes overlapped significantly and the Pearson correlation between the median cell-type compositions across patients of the two atlases was 0.79 (*p*-value = 0.0035) (New Extended Data Fig. 4g). These results suggested that significant biases could not be detected in our data compared to already published healthy human heart single cell atlas data and samples from acute myocardial infarction patients. Overall, we agree that nuclei isolation does potentially lead to biases in regards to distinct immune cells as previously described (Denisenko et al., Genome Biol. 2020, PMID: 32487174). Processing of non-frozen, fresh, human myocardial tissues from acute infarction patients with tissue quantities that allow multi-omic profiling including single-cell RNA-Seq, single-cell ATAC-Seq and spatial transcriptomics was however not possible. Overall our comparative analysis revealed a high degree of correlation with previously published heart atlases.

In response to this important point we have added the following paragraph to the revised manuscript:

"[...] To validate the annotations, we compared the data with a recent study on healthy human hearts[7] and an independent novel dataset of ischemic heart samples (n=3, part of this study) and observed a high agreement and correlation in terms of molecular profiles and cellular composition (Extended Data Fig. 4e-g). Of note, the cycling cells were also captured in the independent ischemic dataset (Extended Data Fig. 4f).[...]"

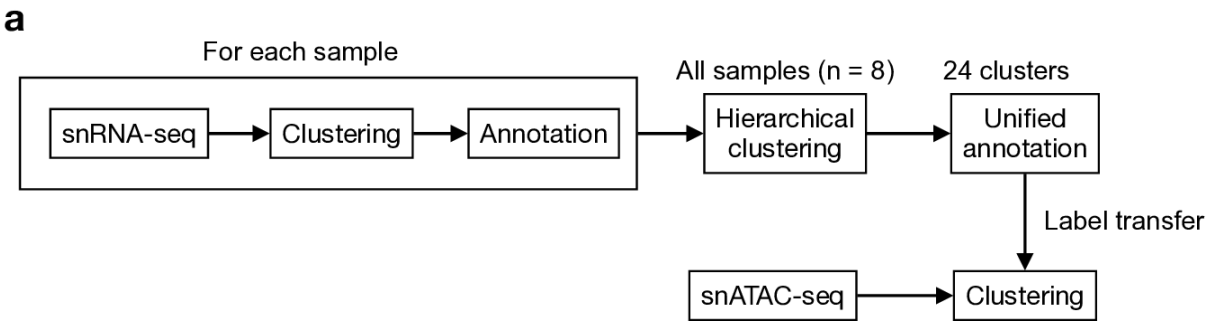
2. What was the rationale behind the cluster resolution that resulted in 24 transcriptional subpopulations? Is the resolution defined by data-driven approaches or related to the previous described biology of the heart. If the latter, the study might miss more subtle disease-specific events.

The reviewer raised a valid point that we have now investigated in detail. In our previous manuscript, we first clustered the cells from snRNA-seq data using a graph-based clustering method from Seurat with the default resolution of 0.8 and annotated the clusters using canonical markers. This analysis was done for each sample (n = 8) independently and

identified a distinct number of cell types (including sub-clusters) per sample. To unify the labels across samples, we then performed hierarchical clustering based on cluster-specific pseudo-bulk transcriptional profiles from all samples and defined 24 clusters based on prior knowledge of the heart (Litviňuková, M. et al., Nature 2020, PMID: 32971526; Wang, L. *et al*. Nat Cell Biol 2020, PMID: 31915373; Tucker, N. R. *et al.*, Circulation PMID: 32403949). We transferred the labels from snRNA-seq to snATAC-seq data (Reply Letter Fig. 2a).

However, we realised that this approach had the following limitations when we increased our dataset:

- It is impractical to apply this method to a large number of samples, as our extended dataset
- The results can be affected by batch effects since no batch-correction is performed prior to clustering
- The rare cell types might be hidden due to the low number in individual samples.



Reply Letter Fig. 2. (a) Computational workflow used for cell annotation in our previous manuscript.

We now have improved the computational strategy in our revised manuscript to identify major cell types and states in the snRNA-seq and snATAC-seq data (New Extended Data Fig. 4a).

To identify major cell types:

- First, we integrated the data from all samples and corrected batch effects using the Harmony algorithm (Korsunsky *et al.,* Nat Methods 2019, PMID: 31740819) and clustered the cells using the graph-based clustering approach from Seurat with a resolution of 1 to generate a large collection of clusters (New Extended Data Fig. 4b and New Extend Data Fig. 5a). This was done for snRNA-seq and snATAC-seq independently.
- Second, we filtered the clusters based on data quality and obtained in total 30 and 25 clusters for snRNA-seq and snATAC-seq, respectively (New Extended Data Fig. 4b and New Extended Data Fig. 5a).
- Then, we identified cluster-specific marker genes for each cluster. For snATAC-seq data, we used the gene activity score as estimated by ArchR.
- Finally, we annotated the clusters based on canonical markers from literature (Litviňuková, M. *et al.*, Nature 2020, PMID: 32971526; Wang, L. *et al.*, Nat Cell Biol 2020, PIMD: 31915373; Tucker, N. R. *et al.*, Circulation PMID: 31915373). The clusters expressing similar marker genes were merged. This uncovered ten and eight major cell types for snRNA-seq and snATAC-seq, respectively (New Fig. 1d and New Fig.

1f). The cell-type-specific markers are provided in Supplementary Table 8 (from snRNA-seq) and 9 (from snATAC-seq).

To identify cell states for cardiomyocytes, endothelial cells, fibroblasts, and myeloid cells across multiple samples and modalities:

- For each major cell type mentioned above, we integrated the snRNA-seq and snATAC-seq data using Seurat and corrected the batch effects using the Harmony algorithm (New Extended Data Fig. 9a; New Extended Data Fig. 11a; New Extended Data. 12a; New Extended Data Fig. 14a).
- Next, we clustered the cells using the graph-based clustering approach from Seurat with a high resolution (i.e., 0.9 for fibroblasts and endothelial cells and 1 for cardiomyocytes and myeloid cells)to generate a large number of clusters to capture subtle disease-specific events. We excluded the clusters that were (1) enriched in a single sample; (2) enriched in a single modality (3) having low data quality; (4) having a high doublet score.
- Finally, we merged the clusters based on cluster-specific marker genes and uncovered 5 states for cardiomyocytes, 5 states for endothelial cells, 4 states for fibroblasts, and 5 states for myeloid cells. The markers for these states are provided in Supplementary Table 8, 10, 11, and 13. The statistical results for sub-clustering analysis were provided in Supplementary Table 16.

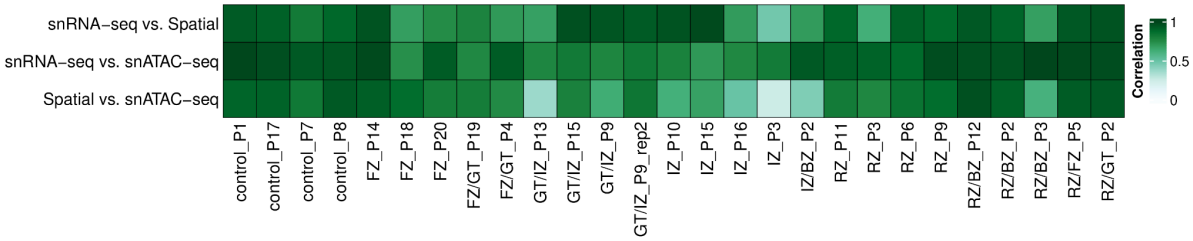
It is worth pointing out that mapping cell types and states at a particular level of resolution of interest remains a challenge for single-cell data (Lähnemann *et al.*, Genome Biol 2020, PMID: 32033589) and there is no consensus on the correct method for choosing the resolution of clustering (Kiselev et al., Nat Rev Genet 2019, PMID: 30617341). We therefore believe that our current approach (i.e., generating a large number of clusters using a high resolution and then merging them manually based on prior knowledge) is appropriate given the state-of-the-art and the complexity of our data.

3. The variable cell type composition between the samples can also be of technical origin. Do the authors have support that missing cell types are absent or less frequent in the respective samples? This can be tested using the spatial information or complementary profiling methods. This is important to clarify as subsequent comparative analysis can be affected by technical nuclei sampling biases.

We thank the reviewer for this important suggestion. We agree that cell type compositions can show technical variability that could bias our interpretations, especially in our previous analysis where we analysed the data sample by sample. We have now increased the overall numbers of samples per patient group and region that allowed a thorough comparison of the cell-type compositions detected across all modalities and conditions. We have also improved our computational analysis strategy to define major cell types and sub-clusters across all samples and modalities. To provide evidence that cell type compositions are stable between technologies in the analysed samples, we leveraged cell type quantification provided by the different technologies used (single nuclei RNA-seq and ATAC-seq, and deconvolution of spatial transcriptomics data). All major cardiac cell types were represented in both snRNA and snATAC technologies (New Extended Data Fig. 5e)

Median Spearman correlation of cell-type compositions estimated from sn-RNA-seq and sn-ATAC-seq data was 0.92, while median Spearman correlation between snRNA-seq and snATAC-seq with spatial transcriptomics was 0.83 and 0.75, respectively. In general we observed stability between the cell-type compositions of each sample across technologies, with exceptions in some ischemic samples (Reply Letter Figure 3). Mean pathway expression of BioCarta's "Death Pathway" and Reactome's "Regulated cell death Pathway" revealed a significant increase in these samples (New Extended Data Fig. 2d). We report these findings in the Results section "An integrative single-nucleus multi-omic map of the human heart" where we have added the following paragraph:

"[...] IZ samples had the lowest abundance in nuclei and an enriched cell death and regulated necrosis pathway expression suggesting increased necrotic cell death (Extended Data Fig. 2d). [...]"



Reply Letter Figure 3. Spearman correlations of cell-type compositions between different technologies across all analysed patients (duplicated samples are summarised together).

To support our comparative analyses on cell-compositions between conditions we have expanded the number of samples and used the mean compositions across technologies to avoid biassing towards a single technology or sample (including snRNA-seq, snATAC-seq and spatial transcriptomics in New Fig. 2i). Moreover, to validate compositional changes of cell-states of cardiomyocytes and myeloid cells, we performed quantification using in-situ hybridization (RNAscope) targeting failing ventricular cardiomyocytes 3 (NPPB +) and SPP1 + macrophages (New Figs 4c, 6n, New Extended Data Figs. 9d, 14n).

4. It is not clear what the authors mean by stating "our integrative single-cell analysis defined a consistent and non-redundant catalog of cell types that comprise the adult human heart". Indeed, there is considerable variability between the samples and the fact that the authors used label transfer to annotate the snATACseq data makes the catalogue redundant between the modalities.

We apologise for this unclear statement and appreciate this question. We have improved our computational workflow to identify major cell types and sub-clusters across all samples and modalities. Briefly, we now integrated the data from all samples and then clustered the cells after batch correction for snRNA-seq (New Extended Data Fig. 4b). The clusters were annotated based on canonical markers and ten major cell types were uncovered (New Fig. 1d). We also identified an additional cluster enriched with the cell-cycle marker MKI67 and we validated its biological relevance using an independent novel dataset of ischemic heart samples (New Extended Data Fig. 4f-g). For snATAC-seq, we also integrated the data and

clustered the cells after batch correction (New Extended Data Fig. 5a). The clusters were annotated using the same markers as snRNA-seq based on gene activity score estimated by ArchR. This identified eight cell types, matching all cell types from snRNA-seq with the exception of two rare cell types (i.e., mast cells and adipocytes) (New Fig. 1f-g).

To compare the identified cell types between snRNA-seq and snATAC-seq, we performed label transfer and observed a high agreement as measured by adjusted Rand index (ARI = 0.98) (New Extended Data Fig. 5b). In addition, we also estimated the proportion correlation of the cell type per sample between snRNA-seq and snATAC-seq and observed a high correlation (New Extended Data Fig. 5d). Finally, we inspected the cell type detection per patient across snRNA-seq and snATAC-seq and observed that almost all cell types were presented in all samples and modalities (New Extended Fig. 5e). All these results indicated that our annotation of major cell types were consistent between samples and modalities.

Of note, we observed that snATAC-seq data tend to show a lower cell coverage (total cell number) compared to in-parallel generated single cell RNA-seq datasets. One explanation could be that for the snATAC-seq reaction (10X Genomics), remaining nuclei after performing the snRNA-seq reactions, have to be concentrated in a small buffer volume of 5-10 µl for the tagmentation reaction when compared to the snRNA-seq workflow. This process can lead to a considerable nuclei loss through centrifugation and resuspension. Indeed, a similar ratio of snRNA/snATAC-seq cells (5/1) is also seen in the human fetal cell atlas, which is based on a combinatorial indexing protocol (Domcke et al, Science 2020, PMID: 33184180). We have now removed the statement from the current manuscript to avoid this confusion and changed the text to:

"[...] Together, our integrative analysis of transcriptomic and chromatin accessibility data defined a robust catalogue of cell types in the adult human heart across multiple modalities and samples.[...]"

5. The Visium platform has a limited size of the capture area. How was the targeted area pre-selected?

We thank the reviewer for this important question. Overall, 20 of these datasets were generated from the left ventricular (LV) free-wall and 11 from LV-apex (New Extended Data Fig. 1). The visium target area is 6.5x6.5 mm and the LV wall was for many specimens thicker than 6.5 mm, therefore we could not get transmural data. Furthermore, many of the specimens were biobanked for years and while the region of the left ventricle was known the directionality of the sample within the cryovial of the biobank was not always clear. We performed an assessment by an experienced and blinded cardiac pathologist (after H&E staining) to confirm the specimen type (ischemic, border zone etc.) and target area.

We have added this description in detail in the Methods section "Human tissue processing and screening". We have now included a diagram from where the samples have been taken as suggested by the reviewer and additional macroscopic and microscopic images of all samples from which these were available (New Ext. Data Fig. 1). Additionally, we provide all available clinical covariate data including location of the biobank, tissue localization, infarct location and a detailed pathologists' description of the tissue annotation in Supplementary Table 1 of all H&E stainings of the previous and added spatial transcriptomics datasets.

Based on this question we have added the following new paragraph to the revised manuscript:

"[...] We applied an integrative single cell genomics strategy with single nucleus RNA sequencing (snRNA-seq) and single nucleus Assay for Transposase-Accessible Chromatin sequencing (snATAC-seq) together with spatial transcriptomics from the same tissue mapping human cardiac cells in homeostasis and after MI at unprecedented spatial and molecular resolution (Fig. 1a-c; Supplementary Table 1). We profiled in total 31 samples from 23 patients including four non-transplanted donor hearts as controls and samples from tissues with necrotic tissue areas (ischemic zone, IZ), border zone (BZ), and the non-affected left ventricular myocardium (remote zone, RZ) of patients with acute MI (Fig. 1a). These acute MI specimens were collected from heart tissues obtained at different timepoints after the onset of clinical symptoms (chest pain), before the patients received a total artificial heart or a left-ventricular assist device due to cardiogenic shock and as a bridge to transplantation (Extended Data Fig. 1a-c). [...]"

We have also further added the following new paragraph to the methods:

"[...] Heart tissues were sampled by the surgeon and immediately frozen in liquid nitrogen. Tissues were dounced in liquid nitrogen and 7-10 mm^3 pieces were embedded in O.C.T. compound (Tissue-Tek) and frozen on dry-ice. 10 µm cryo tissue sections were HE stained and the appropriate tissue regions were selected for further processing. In total 52 human tissue samples were screened this way and evaluated by a cardiac pathologist. For RNA quality control we minced a 3x3 mm^3 heart tissue piece in liquid nitrogen and isolated the RNA using Qiagen RNeasy Mini kit (Qiagen) using a proteinase K digestion step as suggested in RNeasy Fibrous Tissue Mini Kit (Qiagen, 74704). RNA Integrity Number analysis (Agilent) was performed using Bioanalyzer RNA 6000 Nano kits (Agilent, No. 5067). RIN ranged from >2 to maximum 8.8. [...]"

6. Instead of relying on the top DEG for the spatial localization of cell types, the authors should consider the use of deconvolution tools to better predict the composition of each capture site. This should result in a more accurate and unbiased spatial composition of the control and diseased hearts.

We appreciate this suggestion and we agree with the reviewer that deconvolution tools can better predict the cell-type composition of Visium spots compared to using differential expression analysis. Previously robust deconvolution tools were not readily available when we analysed our data for the first version of our manuscript. We have now performed deconvolution of major cell-types in all of the slides using the annotated integrated snRNA-seq data from the respective samples with cell2location (Kleshchevnikov *et al.*, Nat Biotech 2022, PMID: 35027729) (New Fig 1h, Methods Section: "Characterization of spatial transcriptomics data sets.")

By performing deconvolution, we were able to evaluate the biases of nuclei isolation (New Extended Data Fig. 5h) and unbiasedly evaluate the spatial organisation of the different cell-types across patients (section "Modelling tissue organisation of the human heart with spatial transcriptomics", New Fig. 2). The deconvoluted data allowed us to compare the cell-type composition of the control and disease hearts (section "Modelling tissue organisation of the

human heart with spatial transcriptomics", New Fig. 2) and todescribe the spatial organisation of the subtypes of endothelial cells (section "Analysis of endothelial cell heterogeneity at spatial resolution", New Fig. 5). Additionaly, we associated a "stressed" cardiomyocyte functional state to different tissue structures that contain different compositions of vascular smooth muscle cells (VSMCs, fibroblasts, myeloid cells or adipocytes (section "Gene-regulatory and spatial variability of cardiomyocyte states", New Fig. 1-6).

7. In line, clustering the ST spots and subsequent annotation based on marker gene expression of selected cells type is a very hypothesis-driven strategy that could be strongly biased and exclude additional biological and potentially clinical relevant interpretation. Hence, the interpretation of the ST data (e.g. Figure 3) should follow a more data-driven strategy, such as spot deconvolution based on the matched snRNAseq datasets.

We agree with the reviewer and appreciate the suggestion. As mentioned in the previous response, we have performed deconvolution of the spatial transcriptomics slides to better quantify the cell-type compositions in each location.

In the revised manuscript, we have now performed integration and clustering of the spatial transcriptomics spots of all of the slides to identify shared molecular niches between the slides (New Figs. 2-3). We hypothesised that these niches represent potential structural building blocks that are shared between different slides and could facilitate patient comparison. We built two different definitions of these niches based on either i) cell-type compositions, leveraging the deconvolution results, or on ii) the expression of shared variable genes across slides, defined as molecular niches.

Although both representations allowed us to identify differential compositions of fibrotic, inflammatory and myogenic tissue structures between sample groups, only the niches described by gene expression i.e. molecular niches, were able to capture differences between remote zones, border zones and control samples (New Fig. 3g-i). Furthermore the molecular niches defined distinct vascular structures including a molecular niche comprising VSMCs and endothelial cells/pericyte (niche 10 and 11, respectively)(New Fig. 3a-c).

We discuss the utility of both niche representations in two new results sections of the manuscript: "Structural and compositional variation of histological human cardiac tissue classifications" and "*Identification of disease specific cardiomyocyte states*" (New Figs. 2-3).

8. The interpretation and validation of the endothelial cell states is confusing. No detailed interpretation is added for Endo2 and the selection of ISH markers is not clearly related to previous described cell state markers. In general, the MISTy analysis is intriguing and an additional validation of the co-localisation between Fibro2 and Macro2 would further underline the validity of the predictions.

We thank the reviewer for this point and apologise for the confusion. The in-situ hybridisation markers presented in our previous Extended Data Fig. 5f (New Extended Data Fig. 11c), were selected to capture arterial and endocardial endothelial cells. We revised our analyses and the entire manuscript to make the description of endothelial cells clearer and have included the results in a new section: "Analysis of endothelial cell heterogeneity at spatial resolution" (New Fig. 5).

In this section we described with our expanded data a redefined collection of *"[...] 5 subtypes of endothelial cells from all major vascular beds, namely capillary endothelial cells (FABP5, FLT1, and AQP1), arterial endothelial cells (SEMA3G, PCSK5, and GJA5), venous endothelial cells (ACKR1, TPO, and FAM155A), lymphatic (MMRN1, FLT4, and PROX1) and endocardial endothelial cells (NRG3, POSTN, and PKHD1L1)"*. These subtypes were obtained by co-embedding and clustering the snRNA- and snATAC-seq data after correction of batch effects (New Fig. 5a-c, New Extended Data Fig. 11a). We used immunofluorescent staining of *SEMA3G* (marker of arteriolar endothelial cells) and *ACTA2* (marker of vascular smooth muscle cells) in order to validate that indeed *SEMA3G* is a protein expressed by arteriolar endothelial cells in our cardiac tissues (New Extended Data Fig. 11c) as previously suggested in Litvinuková et al. (Nature 2020, PMID: 32971526). For the endocardial endothelial cells we combined *POSTN* and *PECAM1* (pan-endothelial cells marker) for an in-situ hybridization validation (New Extended Data Fig. 11c). Lymphatic endothelial cells expressed bona fide lymphatic marker genes (for example *PROX1*) as described e.g. in Wilting et al. (FASEB Journal 2002, PMID:12060670).

We have expanded the description of all endothelial cell subtypes by identifying unique epigenetic profiles of all endothelial cell-types and described their spatial distribution with other cell-types leveraging our spatial transcriptomics data (New Figs. 5a-c; New Extended Data Fig. 11b-c). Our spatial analyses validated the expected colocalization of venous endothelial cells with vascular smooth muscle cells, and the distribution of capillary endothelial in areas enriched with pericytes and cardiomyocytes (New Fig. 5e-g). Moreover, we performed comparative analysis of capillary endothelial cells between our different sample groups, suggesting a significant reduction in capillary endothelial cells in ischemic samples and an increased cell proportion of lymphatic endothelial cells which fits to the overall hypoxic and inflammatory environment of these samples.

Regarding the Fibro2 and Macro2 colocalization, we have expanded our analysis by modelling the spatial interactions of fibroblast and myeloid states across all slides using MISTy (results section "Spatial organisation of fibro-myeloid cell states in cardiac remodelling", New Fig. 6l-n). First, we redefined the cell-states of myeloid and fibroblast populations using co-embedding and clustering of the expanded snRNA- and snATAC-seq data after batch correction (New Fig. 6a, New Fig. 6j). Next we fitted a MISTy model to predict the marker expression of fibroblasts cell-states in terms of the expression of markers of myeloid cell-states in regions of interest enriched by fibroblasts (see new methods section: "Estimation of the impacts of the spatial context in gene expression") (New Fig. 6l, New Extended Data Fig. 15a). We found that SPP1+ macrophages within spots and in the local neighbourhood were the best predictors of the fibroblasts' cell-states, particularly of myofibroblasts (New Fig. 6l, New Extended Data Fig. 15a). To further validate the prediction of this spatial colocalization, we performed in situ hybridisation experiments in a larger cohort of samples (n= 137 in-situ hybridisation slide from n=27 patients) for SPP1, CD163, and POSTN which demonstrated close spatial association of these cell types (New Fig. 6n, New Extended Data Fig. 15g).

9. The authors argue that the lower gene count in ST and the reduced recovery of cell types from ischemic areas is due to an increased cell death. Although this is certainly expected given the underlying pathology, do the authors find evidence for such assumption?

We appreciate this comment and have performed additional analyses to evaluate the quality of ischemic samples.

We revised our sample integration and atlas annotation strategy for both snRNA-Seq and snATAC-seq and compared to our initial annotation and, despite low gene counts and reduced number of recovered nuclei in ischemic samples, we are now able to identify the eight most abundant major cell-types in most of the samples (New Extended data Fig. 5e).

However, to support our assumption that a lower number of nuclei recovered from ischemic snRNA-Seq samples is associated with increased cell death, we quantified the transcriptional signal of cell-death in the spatial transcriptomics slides across different patient groups (New Extended data Fig. 2d). We expected that cell-death and necrosis transcriptional signals would be higher in ischemic samples compared to the rest. We added details of the estimation of the cell-death and necrosis transcriptional signals in spatial transcriptomics in our methods section:

"[...] *To associate the differences in nucleus capture in snRNA-seq between the different samples to cell-death processes, we leveraged the information from spatial transcriptomics to estimate the general expression of genes associated to cell death for each sample. For each unfiltered slide we estimated per spot the normalised gene expression of BioCarta's "Death Pathway" and Reactome's*[36] *"Regulated Necrosis Pathway" using decoupleR's*[21] *(v1.1.0) wmean method. To have a final pathway score per slide, we calculated for each slide the mean "pathway expression" across all spots.[...]"*

We observed that ischemic samples in general had higher mean pathway scores of "Death Pathway" and "Regulated Necrosis Pathway" compared to the other sample groups (New Extended Data Fig. 2d). Moreover, we observed an anticorrelation between the mean number of recovered nuclei per group and the mean slide pathway scores for these two gene sets (Pearson correlation = -0.938, *p*-value = 0.0184, for "Death Pathway", and Pearson correlation = -0.885, *p*-value = 0.0460 for "Regulated Necrosis Pathway"). In some ischemic slides we could also observed an anticorrelation between the number of unique molecular identifiers and the pathway scores (Extended data Fig. 2d, right panels). These results support the hypothesis that the quality and reduced cell recovery of ischemic samples is likely due to increased levels of cell-death.

10. It would be important to validate the zonation of cell type distribution in respect to the ischemic and scaring areas in multiple biological replicate samples (i.e. different donors), as a key finding in this work, and to extract general conclusion about composition dynamics. Some of the zonation events should be validated with alternative cell-resolution methods. This would be particularly interesting for the events of neo-angiogenesis and scaring-specific fibroblast populations.

We agree with the reviewer and we have considerably expanded the number of samples for each region and patient group to perform comparative analyses at the compositional, molecular and spatial levels in two new results sections: "Spatial and compositional variation of histological human cardiac tissue classifications" (New Fig. 2) and "Molecular variation of human cardiac tissue following acute myocardial infarction" (New Fig. 3). In these sections,

we leverage the spatial transcriptomics extended data to describe general patterns of cell-type zonation across all conditions and specific remodelling events that occur in ischemic and scaring areas in multiple biological replicate samples. Specifically we present the following results:

1) Definition of cell-type colocalization niches: By integrating and clustering the spatial transcriptomics spots of all slides based on their cell-type compositions, we were able to unbiasedly identify nine different cell-type niches that describe myogenic, vascular, inflammatory, and fibrotic tissue structures (New Fig 2a-d). We demonstrated that these cell-type niches can be found in all of the analysed specimens.

2) Spatial dependencies between different cell-types across the samples: We analysed how cell-types generally relate to each other across all the analysed samples, by fitting spatially contextualised models to predict the abundance of each major cell-type (New Fig. 2e).

"[...] We evaluated three different neighbourhood area sizes using MISTy[21]: 1) the importance of cell-type abundances within a spot (colocalization) (Fig. 2e), 2) in the direct (immediate) neighbourhood, and 3) in an extended neighbourhood that expanded to a radius of 15 spots. We observed that endothelial cells were the most predictive of the abundance of vSMCs, pericytes, adipocytes, and cardiomyocytes within all spots, likely reflecting dependencies between cell types of the vasculature (Fig. 2e). Lymphoid and myeloid cells showed strong dependencies with each other in line with zones of immune cell infiltration and inflammation - similarly captured by cell-type niche 5 (Fig. 2e). [...] "

"[...] Interestingly, we observed strong dependencies between myeloid cells and fibroblasts, which were strongly co-enriched in niche 4 (Fig. 2e, Extended Data Fig. 6e), in line with a known important role of macrophages in fibroblast activation[22] and fibroblasts in macrophage attraction[23]. Between immediate and extended neighbouring spots (Extended Data Fig. 6f-h) we observed stronger dependencies between cells associated with the cardiac vasculature (vSMCs, EC, PC, Fibs) indicating that the vascular myocardial cell network dominates cardiac tissue structure organisation. [...] "

3) Identification of differential spatial cell-type dependencies between myogenic, fibrotic, and ischemic sample groups:

"[...] To this end we contrasted the previously computed MISTy importances of each major cell-type in predicting the others in the three different neighbourhood area sizes (colocalization, immediate and extended neighbourhood) between the three different sample groups (Extended Data Fig. 7h). We observed an increased spatial dependency in the immediate neighbourhood between lymphoid and myeloid cells in ischemic samples compared to myogenic-enriched samples, reflecting the expected role that immune cell interactions have of on cardiac repair following myocardial infarction[28](Extended Data Fig. 7i). Moreover, an increased colocalization of cardiomyocytes and pericytes in fibrotic-enriched samples revealed an exclusion of pericytes from scar tissue areas (Fig. 2j). Similarly, the distribution of fibroblasts was better predicted by the presence of vSMCs in the immediate neighbourhood only in myogenic-enriched samples, where fibroblasts surrounded the vasculature[29], compared to ischemic and fibrotic tissue specimens, where more extensive tissue scarring processes were captured (Fig. 2k)." [...]

4) Differential compositions of cell-type niches:

"[...] We next compared compositions of cell-type niches between groups and observed differences in six out of nine cell-type niches (Fig 2l-o; Extended Data Fig. 7j). Cell-type niche 8 (Fig. 2m) and 9 (Extended Data Fig. 7k), mostly representing cardiac muscle structures, were more present in myogenic- and fibrotic-enriched samples compared to ischemic-enriched samples, while cell-type niche 7, enriched in CM and pericytes (Fig. 2l), was reduced in fibrotic-enriched samples. Niche 4, mainly associated with fibrotic structures (more fibroblasts than myeloid cells and thus termed fibrotic-niche), was observed in higher proportions in fibrotic-enriched samples (Fig. 2n), while niche 5 (more myeloid cells than fibroblasts and thus termed inflammatory-niche) was mainly present in ischemic-enriched samples (Fig. 2o)."

Moreover, we describe in three new sections the functional states of cardiomyocytes, endothelial cells, fibroblasts and myeloid cells ("Identification of disease specific cardiomyocyte states" in New Fig. 4, "Analysis of endothelial cell heterogeneity at spatial resolution" in New Fig. 5, and "Spatial organisation of fibro-myeloid cell states in cardiac remodelling" in New Fig. 6). In these sections we characterise the epigenetic and transcriptional differences of each cell-state. We evaluated how cardiac cell states are influenced by their tissue microenvironment and for disease-relevant cell-states we performed validations using in-situ hybridization. Specifically, we present the following results:

1) Identification of a stressed cardiomyocyte cell-state that was overrepresented in ischemic samples (New Fig. 4a-d). Validation of differential state compositions was performed in a collection of 98 images from 17 patients with in-situ hybridisation (New Fig. 4c)

"[...] To further investigate distinct CM-states, we aimed to understand the molecular heterogeneity of cardiomyocytes after myocardial infarction. We co-embedded the snRNA-seq and snATAC-seq data from cardiomyocytes into a common low-dimensional space and clustered the cells (Extended Data Fig. 9a). This uncovered five cell-states of cardiomyocytes (vCM1-vCM5), spanning multiple samples, regions, and modalities (Fig. 4a). [...]"

"[...] Cellular composition comparison between sample groups revealed that vCM1 was associated with myogenic-enriched samples and vCM3 was significantly associated with ischemic-enriched samples. This was validated in an independent cohort using in-situ hybridisation, suggesting that these CM states represent distinct cellular stress states within the acute myocardial infarction phase. (i.e., vCM1; "non-stressed", vCM2 "pre-stressed" and "stressed" vCM3) (Fig. 4c-d; Extended Data Fig. 9h-i)."[...]

2) Differential zonation of the stressed cardiomyocyte state vCM3 across patient groups (New Fig. 4k-o).

"[...] We next estimated the cell dependencies of the stressed cardiomyocyte state -vCM3 with other cell types within each spatial spot and its local neighbourhood (radius of 5 spots) between sample groups (Fig. 4k-o). We observed that the importance of vSMCs in predicting vCM3 within a spot was higher in myogenic and ischemic samples (Fig. 4k), while the importance of fibroblasts and myeloid cells increased in fibrotic samples (Fig. 4k). The local

neighbourhood modelling of vCM3 revealed that the abundance of fibroblasts better explained vCM3 in myogenic enriched samples compared to fibrotic samples (Fig. 4l, Extended data Fig. 10i). To gain further insight, we visualised the dependencies of vSMCs and fibroblasts with vCM3 in myogenic enriched samples, and observed that their co-localization occurred in the perivascular niches (Fig. 4n). Overall this demonstrates that the "stressed" CM-state vCM3 occurs in the perivascular niche of larger blood vessels, highlighting the interaction of mesenchymal cells[49] of the perivascular niche with stressed cardiomyocytes in this tissue area. Furthermore we noticed that when comparing RZ with control samples, stressed vCM3s are best predicted by myeloid cells (Fig. 4o). This underlines the importance of immune-CM interactions that could additionally explain the increased arrhythmia susceptibility in the remote regions of the post-infarct heart, since it has been shown that cardiac macrophages influence normal and aberrant cardiac conduction[50,51]. Our results showed that the "stressed"-CM-vCM3 can be found in distinct spatial cell-type neighbourhoods enriched by different compositions of vSMCs, fibroblasts, adipocytes or myeloid cells."[...]

3) Differential compositions of capillary and lymphatic endothelial cells between patient groups.

"[...] Co-embedding of snRNA- and snATAC-seq data identified 5 subtypes of endothelial cells from all major vascular beds, namely capillary endothelial cells, arterial endothelial cells, venous endothelial cells, lymphatic and endocardial endothelial cells (Fig. 5a-b; Extend Data Fig. 11a-b). Subtype-based pseudo bulk ATAC-seq signals also revealed distinct chromatin accessibility of these marker genes (Fig. 5c). Based on our analysis, POSTN was a characteristic marker for endocardial EC which we validated by smFISH (Extended Data Fig. 11c). Analysis of cell proportion among the myogenic-enriched, ischemic-enriched, and fibrotic enriched samples revealed a reduction of capillary endothelial cells in the ischemic samples associated with a concordant increase in venous endothelial cells (Fig. 5d; Extend Data Fig. 11d-e). Furthermore, we observed that lymphatic endothelial cells were overall less abundant than the other populations, as expected, but were significantly increased in the IZ, suggesting increased abundance of lymphatics modulating the immune response following cardiac injury[52] (Fig. 5d)."[...]

Analysis of our novel integrated data now demonstrates a significant increase in lymphatic endothelial cells in the ischemic zone samples, thus highlighting neo-lymphangiogenesis in this specific tissue zone following acute myocardial infarction (New Fig. 5d). It has previously been shown that the immune-cell lymphatic endothelial cell cross-talk impacts cardiac recovery after myocardial infarction (Haussari et al., Atheriosclerosis, Thrombosis and Vascular Biology 2020, PMID:32404007). Spatial modelling of lymphatic endothelial cells in our human data indeed recovered a dependency with myeloid cells, but additionally with fibroblasts (Extended Data Fig. 11f).

4) Zonation of endothelial cells subtypes across patients

"[...] *We modelled the association of the different endothelial cell subtypes with the abundances of the other major cell-types in spatial transcriptomics. We observed that the markers of arterial endothelial cells were best predicted by vSMCs within a spot and in the local neighbourhood (radius of 15 spots) reflecting the anatomy of arterioles in the heart (Fig.*

5e-f; Extended Data Fig. 11f). Moreover, the expression of markers of capillary endothelial cells were best predicted by the presence of pericytes in the tissue in line with the known presence and role of pericytes in direct contact to capillary endothelium[53](Fig. 5g)."

5) Identification of differential compositions of SPP1+ macrophages in ischemic samples (New Fig. 6n, New Extended Data 14h+n) and colocalization events with myofibroblasts (Fib2) from spatial transcriptomics. Validations of the zonation events were performed in a collection of 137 images from 26 patients using in-situ hybridisation, which recapitulated the defined cell states and their spatial interaction in ischemic zones of human myocardial infarction.

Overall we have performed comparative efforts between different patient groups using spatial data and orthogonal, experimental validations to describe zonations of cell-types associated with myocardial remodelling.

11. Integration efforts of the snATAC data to extract regulatory and mechanistic information is very sparse and anecdotal. Here, the authors should make an effort to better explain regulatory activity responsible for the tissue remodelling.

We appreciate the suggestion and agree with the reviewer that the efforts of integrating snATAC-seq and snRNA-seq were sparse in our previous manuscript.

We have now revised our computational analysis for snATAC-seq data as follows:

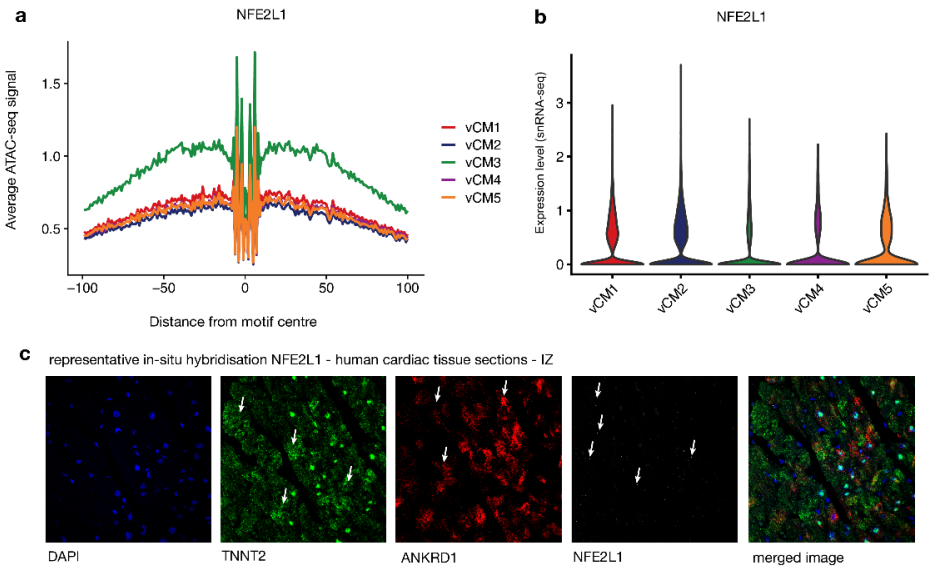
- We integrated and clustered the snATAC-seq data from all samples (New Extended Data Fig. 5a). Clusters were annotated based on gene activity score and eight major cell types were identified (New Fig. 1f-g).
- We performed differential footprinting analysis using HINT-ATAC (Li, *et al*., Nat Com 2019. PMID: 30808370) based on cell-type-specific pseudo-bulk ATAC-seq profiles to compare the TF binding activity between major cell types (New Extended Data Fig. 5f).
- We associated genetic variants with cell types by enrichment analysis using cardiomyopathy-related SNPs obtained from GWAS analysis (Pirruccello JP, et al. 2020. PMID: 32382064) and mapped this information to spatial transcriptomics (New Extended Data Fig. 5i-j) as suggested by the Reviewer #4.
- We integrated the snRNA-seq and snATAC-seq data for cardiomyocytes, endothelial cells, fibroblasts and myeloid cells, and performed sub-clustering (New Extended Data Fig. 9a; New Extended Data Fig. 11a; New Extended Data Fig. 12a; New Extended Data Fig. 14a). This analysis uncovered a distinct number of cell states and subtypes for the above-mentioned major cell types (New Fig. 4a; New Fig. 5a; New Fig. 6a; New Fig. 6j). The statistical results regarding sub-clustering are provided in Supplementary Table 16.

To better identify regulatory changes responsible for tissue remodelling, as also suggested by reviewer #1, we built enhancer-based gene regulatory (eGRN) networks for cardiomyocytes and fibroblasts based on the integrated snRNA-seq and snATAC-seq by combining information from TF binding activity from snATAC-seq, TF expression from snRNA-seq, enhancer-to-promoter links as predicted by snRNA-seq and snATAC-seq, and target gene expression from snRNA-seq data (New Figs. 4h-j, 6g-h). A schematic of the enhancer-based gene regulatory network is provided in New Extended Data Fig. 10b. For the CM the eGRN analysis revealed several interesting factors which regulate different CM-state. The GRN

analysis revealed 3 modules corresponding to the three CM-states included in the analysis (vCM1-3) (New Extended Data Fig. 10e). Among the transcription factors regulating the CM-states were NR3C2, also called mineralocorticoid receptor, MEF2D and ATF3, both stressed induced factors responding to cardiac injury (Kim et al., JCI 2007, PMID: 18079970; Hai et al., Gene Expr. 199, PMID 10440233; Kalfon, Cardiovasc Res. 2017, PMID 28082453). Our analysis of the binding activity of this TF along pseudotime revealed a decrease in binding activity with a concordant decline in gene expression of NR3C2 itself. Mapping of the NR3C2 target gene expression in our spatial transcriptomic slides demonstrated gradual changes of target gene expression of the identified factors e.g. in the broderzone of spatial transcriptomics slides (Fig. 4j). Additionally, for fibroblast we built quantitative enhancer-based gene regulatory networks (eGRN)(New Fig. 6g-h). We identified important putative regulators (i.e., TFs) based on our network analysis of fibroblast to myofibroblast differentiation and highlighted the putative role of KLF4, TEAD3, GLI2 and RUNX2 (New Fig. 6 g-i, New Extended Data Fig. 13a-d). In summary we identified important putative regulators (i.e., TFs) using network analysis. We were able to map predicted TFs of distinct cardiomyocyte and fibroblast cell states and target genes into space to support their role in cardiac remodelling and fibrogenesis.

12. The central regulatory role of NFE2L1 in cardiomyocytes 1 is intriguing, but entirely based on correlation analysis and prediction based on binding motifs. Since the authors highlight NFE2L1 as potential key regulator to derive disease-specific cardiomyocyte subpopulations, an experimental validation of the regulatory role of NFE2L1 in cardiomyocytes would be appreciated.

We thank the reviewer for this important point. We previously selected NEF2L1 (Nrf1) by first predicting the transcription factor binding sites (TFBSs) and then comparing the chromatin accessibility of these TFBSs between different cell types as measured by the snATAC-seq data. This analysis identified NFE2L1 as an important regulator for one of the cardiomyocyte sub-populations (Old Fig. 4e-f). Now, by using our extended dataset, we again observed that this TF showed higher chromatin accessibility around the binding sites in the newly defined “stressed”-CM vCM3 state compared to other cardiomyocytes, consistent with our previous results (Reply Letter Fig. 4a).



Reply Letter Fig. 4 (a). Footprinting profile of NFE2L1 between different cardiomyocyte cell states. The x-axis represents the distance from the NFE2L1 motif centre and the y-axis represents the average ATAC-seq signal around the predicted binding sites of NFE2L1. Colours refer to different cell states. **(b).** Violin plot showing the gene expression of NFE2L1 between different cardiomyocyte cell states based on the snRNA-seq data. **(c)**. Representative In-situ hybridisation of NFE2L1 on human cardiac tissue sections. Note that in stress-marker positive (ANKRD1)-tissue areas NFE2L1 was undetected.

As this approach was solely based on chromatin accessibility data and motif prediction, it can generate false positives. For example, we observed that NFE2L1 showed no clear differences of gene expression between different cardiomyocyte states (Reply Letter Fig. 4b). We additionally performed in-situ hybridisation of NFE2L1 (RNAscope) on human myocardial tissues after acute myocardial infarction but were unable to detect NFE2L1 (Reply Letter Fig. 4c). Moreover, in a recent study (published after the initial submission of our work), the role of NFE2L1 was thoroughly studied and the authors demonstrated that NFE2L1 encodes for the stress-responsive transcription factor Nrf1 in cardiomyocytes and that NFE2L1 overexpression protected the adult mouse heart from ischemia-reperfusion injury (Cui et al., Nature Commun 2021 PMID:34489413).

Therefore, in our revised manuscript, we have improved our computational analysis to build a quantitative gene regulatory network (GRN) by including TF binding activity from snATAC-seq and TF expression from snRNA-seq data (New Extended Data Fig. 10b). Moreover, we also incorporated the expression of target genes which were predicted by using enhancer-to-promoter links and motif binding sites. Our current GRN analysis of human cardiomyocytes pinpointed several interesting regulators (New Fig. 4h). For example, we identified the mineralocorticoid receptor (MR, NR3C2), a main target of common heart failure treatment, as a major regulator of the vCM1 state (New Fig. 4h). For this reason, we now removed NFE2L1 from the revised version of our manuscript.

Minor:
This statement is too broad for the related section: “In summary, the data indicated distinct spatial gene regulation in response to the ischemia associated cell-death with gene regulation driving the acute cardiac injury response.”

We agree with the reviewer and apologise for this general statement. We have now removed this statement from the text entirely. Since we now performed integrative analysis of three defined patient groups, we have additionally rewritten the previous section entirely headed “Demarcation of the ischemic zone visualised by distinct gene expression and regulation”.

Referee #3:
This study by Kuppe et al. investigates the cellular and local gene expression changes in response to ischemic injury in human heart samples at different time points after injury. They do so by performing single cell sequencing, ATAC sequencing and spatial transcriptomic profiling and different bioinformatic approaches to further mine the data.

While these are all state-of-the art methods and bioinformatic approaches that provide insightful data there are currently several issues with the experimental design and the presented data that make it difficult to determine the value and generalizability of the findings.

We thank the reviewer and appreciate the overall positive comments. In order to increase the generalizability of the findings and the value of this multiomic atlas of human myocardial infarction we have now increased the number of tissue samples in this study from initially 8 to 28 from 20 patients over four different cardiac tissue regions, as well as control samples (control = 4, ischemic region = 12, border zone = 3, remote zone = 6, fibrotic zone = 6) and additionally 3 snRNA-Seq datasets from human acute MI for cross-validation. 28 of these samples have a multi-omic profiling with single nuclei (sn) RNA-seq, snATAC-seq, and spatial transcriptomics.

To address the reviewer's concerns, we have now added a thorough description of sampling location and annotations from a cardiac pathologist and performed multiple in-situ hybridizations and quantification for validation of our findings (New Extended Data Fig. 1-3, New Supplementary Table 1). In addition, we also included scRNA-seq from a myocardial infarction time-course lineage tracing experiment of PDGFRβ-reporter mice at distinct time points after MI which was used in our trajectory analysis for myofibroblasts differentiation (New Extended Data Fig. 12i-l).

We systematically compared our data with previously published datasets (heart cell atlas) and an external dataset of human myocardial infarction (snRNA-seq) which both demonstrated high correlations (New Extended Fig. 4e-g).

To further increase the value we have now made all data publically accessible (HCA and cellxgene), which represents a high contribution to the cardiovascular research field. (https://cellxgene.cziscience.com/collections/8191c283-0816-424b-9b61-c3e1d6258a77/private)

Major comments

- A major concern is the reproducibility and validity of the data. The authors have an n=1 for the different conditions (including control), except for the time point more immediate after myocardial infarction. Here they have tissue from both 2 and 5 days after infarction that represent the more immediate response after injury, however these samples seem to vary a lot from each other in cellular composition, local remodeling and gene expression changes and are also representing different regions.

We thank the reviewer for this important point. For this reason, we have increased the number of tissue samples in this study from initially 8 to now 28 from 20 patients over four different cardiac tissue regions, as well as control samples (control = 4, ischemic region = 12, border zone = 3, remote zone = 6, fibrotic zone = 6) and additionally 3 snRNA-Seq datasets from human acute MI IZ sections for cross-validation. 28 of these samples have a multi-omic profiling with single nuclei (sn) RNA-seq, snATAC-seq, and spatial transcriptomics (New Fig. 1a; New Extended Data Fig. 1-3). In addition, we also included scRNA-seq from a myocardial infarction time-course lineage tracing experiment of Pdgfrβ-reporter mice at distinct time points after MI which was used in our trajectory analysis for myofibroblasts differentiation (New Extended Data Fig. 12i-l).

This extended dataset, with multiple biological replicates per condition, allowed us to describe heart remodelling processes by performing quantitative comparisons of the different sample groups at the compositional, molecular, and spatial levels, as the reviewer suggested. We have included these results in two new sections: “Spatial and compositional variation of histological human cardiac tissue classifications” (New Fig. 2h-o) and “Molecular variation of human cardiac tissue following acute myocardial infarction” (New Fig. 3e-i). Additionally, we have redefined the sub-clusters for several major cell-types including cardiomyocytes (New Fig. 4), endothelial cells (New Fig. 5), fibroblasts, and myeloid cells (New Fig. 6) by integrating snRNA-seq and snATAC-seq from all samples and associated them with the different disease stages.

We also mapped the different functional states of cell-types to spatial transcriptomics to investigate their spatial relationship with other major cell types and their changes in different patient groups (New Figs. 4k-o, 5e, 6l). The top 10 differentially expressed genes of the newly described cardiomyocytes-states vCM1-5 are shown in New Extended Data Figure 9b and the complete results are provided in New Supplementary Table 10. We included these results in three new sections: “Identification of disease specific cardiomyocyte states ” (New Fig. 4), “Analysis of endothelial cell heterogeneity at spatial resolution ” (New Fig. 5), and “Spatial organisation of fibro-myeloid cell states in cardiac remodelling” (New Fig. 6).

- Is the sample for spatial transcriptomic take transmurally? The manuscript would benefit from a better description of the exact sample collection procedure.

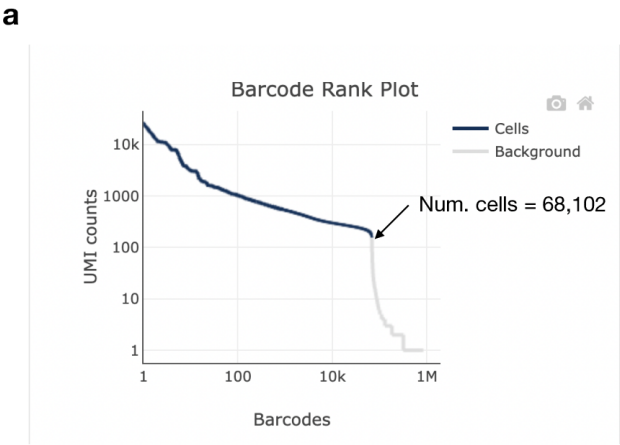
We agree with the reviewer and have significantly improved the sample location description and annotation of the cardiac regions in the new version of the manuscript. Overall, 20 of these datasets were generated from the left ventricular free-wall and 11 from LV-apex. The visium target area is 6.5x6.5mm and the left ventricular wall was for many specimens thicker than 6.5mm, therefore we could not get transmural data. Furthermore, many of the specimens used were biobanked for years and while the region of the left ventricle was known the directionality of the sample within the cryovial of the biobank was not always clear. We performed an assessment by an experienced and blinded cardiac pathologist (after H&E staining) to confirm the specimen type (ischemic, border zone etc.) and target area.

We added this description in detail now to the Methods section (“Human tissue processing and screening”. We have however now included a diagram from where the samples have been taken as suggested by the reviewer and additional macroscopic and microscopic images of all samples from which these were available (New Ext. Data Fig. 1). Additionally, we provide all available clinical covariate data including location of the biobank, tissue localization, infarct location and a detailed pathologists’ description of the tissue annotation in Supplementary Table 1 of all H&E stainings of the previous and added spatial transcriptomics datasets.

- Where are the data for the ischemic zone from patient 3 in Extended Data Figure 1 and 3d?

We apologise that this was not clear enough in our previous manuscript. For this sample, the CellRanger pipeline initially reported 68,102 cells from the snRNA-seq library, likely due to ambient RNA from necrotic tissue areas (Reply Letter Fig. 5a). We therefore decided to not include it in our previous analysis given that this number (68,102) is usually unexpected.

In our current manuscript, we have revised our data quality control strategy to filter cells rather than samples. Specifically, for each sample we removed nuclei (i) in the top 1% in terms of the number of genes, (ii), with less than 300 genes and less than 500 UMIs, (iii) or more than 5% of mitochondrial gene expression, and (iv) doublets as estimated using *scDblFinder* (Germain, P. et al., F1000Research 2021, doi: 10.12688/f1000research.73600.1) with default parameters. After filtering, we obtained 581 cells for this sample (Supplementary Table 2). We next used these cells as input for data integration and clustering (New Extended Data Fig. 4b). Annotation of the clusters revealed that this sample included similar cell types as other ischemic samples, demonstrating the power of the integrative analysis (New Extended Data Fig. 4d).



Reply Letter Fig. 5 (a). Barcode rank plot generated by CellRanger pipeline showing cell calling results for the snRNA-seq data of the ischemic zone from patient 3. The y-axis is the number of UMI counts mapped to each barcode and the x-axis is the number of barcodes below that value.

- It is unclear how well the different datasets for patient 2 and 3 overlap. This should be clarified more as these are the only 2 patients from the same condition/group. Are the same cell cluster found when comparing these 2 patients and are the same gene expression correlations found between the 2? It would be better for clarity to compare the same type of analyses for the different samples.

We appreciate these comments on patient comparisons in our heart atlas. We have now increased the number of specimens as outlined above and changed the computational strategy to allow for comparisons between different patients. Overall, our atlas now contains 12 ischemic zone samples, compared to 2 in the previous version. Our new integrated data now comprises several biological replicates in all major patient groups (New Fig. 2h). This allowed us to perform comparative analysis between the specimens in one condition and across conditions (New Fig. 2i-o, New Fig. 3-6). The characterization of ischemic samples revealed the following:

1) Despite reduced recovery of nuclei and low gene counts in ischemic samples (New Extended Data Fig. 2c), we were able to identify in most of the samples 11 major cell-types in snRNA-seq and the eight most abundant major cell-types in snATAC-seq (New Extended Data Fig. 5e).

2) Ischemic samples contained a reduced proportion of cardiomyocytes, and an increased proportion of lymphatic, myeloid and cell-cycling cells based on the estimation of cell-type compositions across omic modalities (New Fig. 2i). Similar cell-type compositions were observed in an external validation dataset of three ischemic patients (New Extended Data Fig. 4f-g).

3) Increased compositions of an inflammatory cell-type niche in spatial transcriptomics (groups of spots with an increased proportion of immune cells colocalized with fibroblasts) compared to the other patient groups. Additionally, we observed reduced compositions of myogenic-related niches (groups of spots with increased cardiomyocyte compositions) (New Fig. 2l-o).

4) In spatial transcriptomics, we observed an increased colocalization of immune cells within the spot and in direct neighbours across the tissue, compared to myogenic-enriched or fibrotic-enriched samples, where the distribution of these cells was less structured (New Extended Data Fig. 7h)

5) We observed an increased enrichment of a “stressed” cardiomyocyte cell-state compared to myogenic-enriched or fibrotic-enriched samples. These observations were recapitulated in an independent cohort of 17 patients, using in-situ hybridisation (number of images = 98) (New Fig. 4c).

6) We observed a reduction in the proportions of capillary endothelial cells and an increment of lymphatic endothelial cells in ischemic samples compared to the other patient groups (New Fig. 5h). Within capillary endothelial cells, we observed increased activities of the hypoxia and TGFβ signalling pathways, compared to myogenic enriched samples (New Fig. 5j).

7) Increased compositions of a myofibroblast cell-state (Fib2) (New Fig. 6d) and SPP1+ macrophages (New Extended Data Fig. 14f) were observed in ischemic samples compared to the other patient groups. Validation of SPP1+ macrophage enrichment in ischemic samples was done with an orthogonal data set of 26 patients using in-situ hybridization (137 images) (New Fig. 6n, New Extended Data Fig. 15g). Moreover we observed a spatial gradient of the colocalization of myofibroblasts and SPP1+ macrophages in spatial transcriptomics (New Fig. 6m, New Extended Data Fig. 15b).

Overall, our increased data set and revised computational analysis allowed us to describe cardiac remodelling events observed in ischemic samples that covered changes in cell-type composition and tissue organisation.

- It currently is unclear which portion of the control heart was taken for the analyses.

We thank the reviewer for this important point. We revised our sample description and now included all annotations in Supplementary Table 1. We included macroscopic and microscopic images of all samples in New Extended Data Fig. 1-3. Overall, 20 of these datasets were generated from the LV free-wall and 11 from LV-apex as depicted in New Extended Data Fig. 1a. The control specimens of non transplanted donor hearts were taken from the LV free-wall (New Supplementary Table 1)

- The authors combine spatial transcriptomics on a 10um section with snRNA seq and ATAC seq from an adjacent portion of the heart. However, the results of these three give very different cellular compositions and gene expression profiles. The spatial transcriptomic for example shows an overrepresentation of cardiomyocytes (likely because of their size), while the cellular composition based on snRNA seq versus ATAC seq also gives a very different image (which can for example be seen in Extended Data Fig 3d). What would the Extended Data Fig 3d look like for ATAC seq only? If the dataset differ so much it is hard to generate an integrative molecular map as the authors indicate in the abstract.

We thank the reviewer for the comment. We agree that in order to provide an integrative molecular map of human myocardial infarction, consistent cellular compositions should be observed across technologies. In the revised manuscript we have improved our computational analysis strategy to define major cell types and sub-clusters across all samples and modalities, to ensure consistency in the provided atlas. Together with the increased sample size, we now performed a more thorough comparison of the cell-type compositions detected across modalities and conditions.

We demonstrated that overall, we could identify in most of the samples and across conditions 11 major cell-types in the snRNA-seq data. From these cell-types, the eight most abundant ones were also recovered in the snATAC-seq (New Extended Data Fig. 5e).

To provide evidence that cell type compositions are stable between technologies in the analysed samples, we leveraged cell type quantification provided by the different technologies used (snRNA-seq and snATAC-seq, and deconvolution of spatial transcriptomics data). Median Spearman correlation of cell-type compositions estimated from sn-RNA-seq and sn-ATAC-seq data was 0.92, while median Spearman correlation between snRNA-seq and snATAC-seq with spatial transcriptomics was 0.83 and 0.75, respectively. In general we observed stability between the cell-type compositions of each sample across technologies, with exceptions in some ischemic samples (Reply Letter Figure 3, page 12). Mean pathway expression of BioCarta's "Death Pathway" and Reactome's "Regulated cell death Pathway" revealed a significant increase in ischemic samples (New Extended Data Fig. 2d).

To compare the identified cell types between snRNA-seq and snATAC-seq, we performed label transfer and observed a high agreement as measured by adjusted Rand index (ARI = 0.98) (New Extended Data Fig. 5b). To demonstrate that the patient groups contained similar expression profiles, we created pseudo-bulk transcriptional profiles of the spatial transcriptomics slides and performed hierarchical clustering (New Extended Data Fig. 7f). Overall, we observed that samples grouped by their histomorphological region: border zones, remote zones, and controls formed one cluster, while ischemic and fibrotic samples formed two different clusters.

Our extended analysis provides evidence to suggest consistency across omics modalities of the analysed samples and within patient groups.

- How many cells are roughly represented per spot in the spatial transcriptomics?

We thank the reviewer for this important question. We have now quantified the number of nuclei (cells) leveraging the H&E stainings of the spatial transcriptomic sections. We used VistoSeq, a matlab pipeline which allowed us to determine the exact nuclei count per visium

spatial detection spot. Our quantification revealed a mean value of 4 nuclei per spot (New Extended Data Fig. 2c).

- The snRNA seq and ATAC seq data generated on samples taken 2 or 5 days after ischemic injury appear to be strongly influenced by the presence of cell death and only 4 cell types can be distinguished. However, in looking at extended Data Figure 6e all 4 cell types appear to be expressing TNNT2, a cardiomyocyte marker.

We thank the reviewer for this important observation. Indeed, ischemic samples (in the previous and current sample collection) are influenced by cell death and necrotic processes as expected (New Extended Data Fig. 2d), hence their lower recovery of nuclei and gene counts (New Extended Data Fig. 2c). Cell-death leads to autolysis with leaky nuclei and thus high ambient RNA, particularly from the dead cells of this area which are many cardiomyocytes. This could explain why we observed background expression of TNNT2 in these patients.

In our current computational strategy we decided to annotate single nuclei of snRNA-seq by integrating and clustering single nuclei after a first round of quality control for each data set separately (New Extended Data Fig. 4a). We ensured that the highly variable genes used to cluster the integrated data sets were representative of most of the samples, under the assumption that this collection of genes will represent cell-type variability rather than technical variability (including background signal). With this strategy applied to the extended dataset we were able to identify in most of the samples 11 major cell-types in snRNA-Seq. (New Extended Data Fig. 5e). Moreover, for the definition of cell-states we performed multimodal integration of snRNA-seq and snATAC-seq data of cell-types of interest and performed stricter quality controls of nuclei to prioritise biological variability over technical variability. This, together with the fact that cell-states were present in all patient groups in different compositions demonstrate that although the quality of ischemic samples is different from the other patient groups because of cell-death, our current annotation is capturing biological variability that is consistent in multiple patients.

Based on this question we have added the following sentence to the discussion of the revised manuscript :

“[...] Of note, in the ischemic samples we observed high levels of cell death, as expected, and thus also higher levels of ambient RNA which could introduce a bias in our analyses. [...]”

- While the validation studies in heart failure samples and functional follow-up studies support the relevance of the sequencing data, Runx1 has already been linked to TGFb signaling and myofibroblast differentiation and fibrosis.

We thank the reviewer for pointing this out. Indeed the link of RUNX1 as being involved in myofibroblast differentiation and TGFβ-signalling has been shown before, e.g. in zebrafish heart (Koth et al., Development 2020, PMID: 32341028) or in the cell-culture using bone marrow-derived mesenchymal stem cells (Kim et al., PNAS 2014, PMID: 25313057). We also recently identified and validated a role of RUNX1 in fibroblast to myofibroblast differentiation in the mouse kidney (Li et al. Nat. Com. 2021, PMID: 34737275). We therefore agree that this finding is not particularly exciting or novel. However, we have provided the first evidence that

in human heart fibrosis (Extended Data Fig. 12b) and in human myofibroblast differentiation (Extended Data Fig 12d) Runx1 is critically involved. We have now removed the RUNX1 panels from the main figure (Old Figure 6k,l+n, New Fig. 6) and included the Old Figure Panel 6n to the New Extended Data 12d).

We have adapted the section on RUNX1 in the manuscript including the two suggested references of the reviewer:

"[...] Notably, Fib2 also showed an up-regulation of RUNX1 which we have recently reported as being involved in kidney myofibroblast differentiation[58]. A role of RUNX1 in myofibroblast differentiation has been suggested in-vitro using mesenchymal stem cells[59] and in-vivo in zebrafish heart injury[60]. Overexpression of RUNX1 in human heart PDGFRb cells led to increased myofibroblast differentiation and matrix expression (Extended Data Fig. 12d). [...]"

- How do these snRNA seq data compare to previously published studies on human heart tissue?

We thank the reviewer for this important question. We now have compared the cell-type compositions of our myogenic-enriched samples (remote zones, border zones, and controls) with the reported cell-type compositions of the single nuclei samples in the human cell atlas by Litvinuková et al. (Nature 2020, PMID: 32971526) (New Extended Data Fig. 4e). We observed a Pearson correlation of 0.88 between the median cell-type composition across patients of the two datasets (New Extended Data Fig. 4e, right panel). We observed a significant overlap between the marker genes inferred from these two datasets for the same cell type (New Extended Data Fig. 4e, left panel). Additionally, we compared the cell-type compositions of the ischemic samples with a validation set of three human ischemic samples from an independent cohort. Pearson correlation between the median cell-type compositions across patients of the two atlases was 0.79 (New Extended Data Fig. 4f-g). Similarly, we observed a significant overlap between the marker genes of these two data sets (New Extended Data Fig. 4g left panel). These results suggested that strong biases could not be detected in our data compared to already published healthy human heart single cell atlas data and samples from acute myocardial infarction patients. Overall our comparative analysis revealed a high degree of correlation to previously published heart atlases.

Referee #4:
Tanevski et al report the results of deploying cutting edge spatial and single cell genomics analyses on precious, rare human ex vivo cardiac samples derived from individuals post myocardial infarction. Certainly the rarity of this sample set, and the promise of the technological tools used, intrigued and excited me about the possible biological insights offered. However, after reading the paper carefully I remain unsure what the authors have actually learned from all of this data generation. Is this paper reporting a biological discovery? I do not study heart biology, so I am genuinely uncertain, but the organization of the paper, and the details of the claims made (see specific points below) do not suggest that a clear, novel insight about myocardial infarction has emerged from this work. And if this biological insight has not been gained, and this manuscript is being presented as a reference atlas of MI, then greater effort needs to be made in codifying, organizing and releasing these data to

the scientific public for consumption (for example, the generation of a web-based tool for plotting genes, performing individual analyses, etc).

We thank the reviewer for the positive impression of our work and for acknowledging the rarity of the sample included in the study. We also appreciate the reviewer's valuable suggestions and criticisms. With this in mind, we sought to address each of the reviewer's points below as thoroughly as possible. Our goal now is to present a spatial multi-omic map of human myocardial infarction, the number one cause of mortality in the world, to the scientific community. In order to do that we have increased the number of tissue samples in this atlas from initially 8 to 28 from 20 patients over four different cardiac tissue regions, as well as control samples (control = 4, ischemic region = 12, border zone = 3, remote zone = 6, fibrotic zone = 6) and additionally 3 snRNA-seq datasets from human acute MI and 4 scRNA-seq datasets from mice at different time points after MI for cross-validation. As in our previous manuscript, now 28 of these samples have a multi-omic profiling with single nuclei (sn) RNA-seq, snATAC-seq, and spatial transcriptomics. This extended dataset allowed us to describe heart remodelling processes by performing quantitative comparisons of the different sample groups at the molecular, compositional and spatial level.

In detail we have now included these results in two new sections: "Spatial and compositional variation of histological human cardiac tissue classifications" and "Molecular variation of human cardiac tissue following acute myocardial infarction". We revised the snRNA-seq and snATAC-seq data integration to identify major cell types using state-of-the-art methods including Harmony and ArchR (as suggested by the reviewers). Instead of annotating the cell types sample by sample as in our previous analysis, for each modality we integrated the data from all samples and then clustered the cells. Clusters were annotated independently for snRNA-seq and snATAC-seq. To validate major cell-type annotations we performed cross-modality comparison between snRNA-seq and snATAC-seq data within our samples, and between the human heart cell atlas that profiled healthy samples and an external reference snRNA-seq data set of ischemic samples (mentioned above) (New Extended Data Fig. 4a,e-g). We have significantly improved the description of the metadata involving a cardiac pathologist who annotated the samples (New Extended Data Fig. 1-3, New Supplementary Table 1).

We additionally performed a transgenic mouse experiment using PdgfrβCreER;tdTomato inducible fate tracing of all mesenchymal cells in myocardial infarction (LAD ligation) with subsequent scRNA-sequencing of tdTomato sorted (FACS) cells at different timepoints (day 0, 4, 7, 14) to compare (New Extended Data Fig. 12i-l).

We built quantitative enhancer-based gene regulatory networks (eGRN) using the integrated snRNA-seq and snATAC-seq data by combining information from transcription factor (TF) binding activity, TF expression, enhancer-to-promoter links, and target gene expression for cardiomyocytes (New Fig. 4h-j) and fibroblasts (New Fig. 6g-h). We identified important putative regulators (i.e., TFs) using network analysis. We could map predicted TFs and target genes into space to support their role in cardiac remodelling and fibrogenesis.

We have reorganised the manuscript so that the results' presentation not only highlights the richness of the data but also shows how the different combination of omics technologies and resolutions can help to describe myocardial remodelling and show the relations between cell-type location, organisation and function.

To make our atlas accessible and the data explorable even for non-computational trained researcher we have made the data freely accessible via the HCA and explorable in the

cellxgene data portal: (https://cellxgene.cziscience.com/collections/8191c283-0816-424b-9b61-c3e1d6258a77/private).

In general, a lot of analyses presented are suggestive, but not clearly demonstrative, of the claims. In particular, there is a logical leap made by the authors between the results of MISTy and the conclusion that the genes identified are somehow directly mediating interactions between cell types. The algorithm is uncovering spatial correlations at different length scales, with certainly could suggest a causal interaction, but may also simply by a correlative effect due to, for example, the developmental patterning of the tissue (certain cells get positioned closer to others, but do not necessarily interact in a causal fashion). Although certainly many of the nominated interactions are intriguing, it's really hard to know what the reader should do with these specific examples, without some sort of perturbation in an animal or organoid model to provide support for causal mechanisms.

We thank the reviewer for this observation. We agree that MISTy's objective is to identify spatial interactions between groups of markers (e.g. gene expression) in different scales and the results do not necessarily imply any type of causality. In our previous version of the manuscript we used MISTy to predict the expression of marker genes of cell-states of interest with a collection of cytokines and extracellular matrix related ligands, under the assumption that these genes could provide potential mechanistic knowledge on cell communication. We agree that MISTy suggests spatial relationships between these collections of genes, but do not provide evidence of the causal effect of the predictors in the expression of the cell-state marker genes. Given these observations, the limitations of inferring cell communication events from transcriptomics data (Dimitrov et al., bioRxiv, DOI: 10.1101/2021.05.21.445160), and the difficulties of disentangling patterning events from communication events in spatial transcriptomics, in this version of the manuscript we instead used MISTy to exclusively study the spatial organisation of cell-types and cell-states in the extended spatial transcriptomics slides.

In our revised analysis, we used MISTy to:

1) Estimate the importance of the abundance of each major cell-type in explaining the abundance of the other major cell-types in different spatial contexts (New Fig. 2e, New Extended Fig. 6g). This analysis explicitly focused on tissue patterning and allowed for the identification of differential cell-type dependencies between patient groups (New. Fig 2j-k, New Extended Fig. 7h).

2) Estimate the relationship between the tissue structure and cell functions, encoded as signalling pathway activities per spot estimated from gene expression (New. Fig 2f-g, New Extended Fig. 7a-c). In this model, we predicted spatial relationships between pathway activities and the importances of cell-type abundances in predicting the patterns of pathway activities.

   "[...]To link tissue organisation to function, we analysed spatial dependencies between signalling pathways and cell-types. Model importances in all spatial contexts captured relationships between PI3K and p53 signalling, which showed a mutually exclusive spatial distribution. Both pathways were related to the abundance of cardiomyocytes

(Fig. 2f, Extended Data Fig. 7a-c). PI3K signalling in cardiomyocytes controls the hypertrophic response to preserve cardiac functions[24], while p53 is known to act as a master regulator in cardiac homeostasis[25]. Spatial segregation of these CM-related pathways points towards functional CM heterogeneity. We observed colocalized and extended neighbourhood relationships of known key pathways in fibrosis including TGFβ and NFkB predicted by fibroblasts, and JAK-STAT and NFkB predicted by immune cells (Extended Data Fig. 7b-e). Overall, CMs were the best predictor cell-types of the activities of the estimated pathways. Hypoxia and WNT pathway showed a colocalization to CMs in ischemic specimens (Fig. 2g, Extended Data Fig. 7b-e) highly consistent with the cardiomyocyte differentiation events occurring after myocardial infarction[26,27]. Our results compiled tissue organisation principles of the human heart that relate to coordinated cellular processes and provide a basis for comparative analysis."[...]

3) Estimate associations between the tissue organisation and the spatial distribution of failing cardiomyocytes and the different endothelial and fibroblast cell-states. We hypothesised that the distribution of specific cell-states in the spatial transcriptomics slides could be modelled by the cell-type composition or cell-state presence of individual spots and their neighbourhood. With these models we were able to capture known co-localizations of venous endothelial cells, vascular smooth muscle cells, and capillary endothelial cells and pericytes (New Fig. 5e-g). Moreover, we describe differential spatial correlations of the transcriptional signature of the "stressed" cardiomyocyte state vCM3 and the abundance of other cell-types between different patient groups. This analysis suggested that the location of this cardiomyocyte state can occur in different tissue regions depending on the time point and region after MI (New Fig. 4k-o). In relation to fibroblast and myeloid cell-state populations, we observed a spatial gradient of myofibroblasts that aligned with a gradient of SPP1+ macrophages. We validated this tissue patterning with orthogonal in-situ hybridisations in a collection of 137 images from an independent cohort of 26 patients (New Fig. 6l-n, New Extended Data Fig. 15g).

With these analyses we aimed to provide to the community a general description of the cell-type organisation of cardiac tissue and its changes after myocardial infarction. Additionally, we demonstrate that cardiac cell-states are associated with their local microenvironment. At the same time, we acknowledge that the reported "interactions" are hypotheses that need to be tested in follow-up studies to clearly relate the spatial correlations to cell-type communications.

We have adapted the methods section to explicitly state the limitations of the reported spatial interactions.

"[...] The aggregated estimated importances (e.g. median) of each view of all slides were interpreted as cell-type dependencies in different spatial contexts, such as colocalization or mutual exclusion. Nevertheless, the reported interactions don't imply any causal relation."

In the discussion of our revised manuscript we have also included these limitations:

"[...] The combination of spatial technologies with single cell data represented an opportunity to study how cardiac cell states are influenced by their tissue microenvironment. The

nominated interactions between cell-types largely reflect the spatial organisation of the tissue and, while many other factors are involved, these interactions provide hypotheses for further analysis."

Here are some specific points where I identified novelty, but where additional follow up work might help to bolster it to a solid mechanistic hypothesis:

One very intriguing insight the authors report is the presence of additional molecular/cell state heterogeneity in acute MI tissue beyond what can be seen histologically (Fig 4). This seems like a great example of how especially the spatial data was able to nominate additional molecular processes beyond what can be observed by conventional pathology, but the analysis ends at suggestive correlations. Can the authors take this a step further, and perhaps nominate a small number of markers for these heterogeneous states, and perform immunohistochemistry on a larger set of cardiac MI samples (for example FFPE samples), more deeply explore whether there are diagnostic, prognostic, or other clinically relevant implications for these different identified cell states?

We thank the reviewer for the positive comments and the important suggestions. We agree that our previous data delivered unprecedented insights into remodelling of the human myocardium after acute MI beyond what can be observed from histology (Old Fig. 4a+d). We agree that a direct transfer of molecular insight of our data is exciting and could lead to novel diagnostic, prognostic or therapeutic approaches in the clinics. However, one major issue here is that our acute MI specimens are solely from total artificial heart programs or heart transplant recipients and thus the clinical outcome of the patients is not associated with the actual data we generated, since the patients have received either an artificial heart with subsequent transplantation or directly a donor heart. Therefore we can not correlate the data in a meaningful way with clinical outcome data directly since the patients included are either already deceased or have received an unrelated donor heart by transplantation.

As suggested by the reviewer, we maximised our effort and collaborated with two of the largest heart biobanks in Europe to increase our sample number. This increased dataset combined with our multiomic approach revealed several novel insights and a better molecular understanding of the processes involving human cardiac tissue remodelling after MI, which cannot be gained by analysing the histology or single-cell RNA-seq data alone. Our aim is to provide a map of human myocardial infarction to the scientific community which is accessible through novel online tools (cellxgene and HCA).

Our novel data includes the following major findings with diagnostic, prognostic and clinical implications:

1) Identification of a stressed cardiomyocyte cell-state that was overrepresented in ischemic samples (New Fig. 4a-d). Validation of differential state compositions was performed in a collection of 98 images from 17 patients with in-situ hybridisation (New Fig. 4c)

"[...] To further investigate distinct CM-states, we aimed to understand the molecular heterogeneity of cardiomyocytes after myocardial infarction. We co-embedded the snRNA-seq and snATAC-seq data from cardiomyocytes into a common low-dimensional space and clustered the cells (Extended Data Fig. 9a). This uncovered five cell-states of cardiomyocytes

(vCM1-vCM5), spanning multiple samples, regions, and modalities (Fig. 4a). [...]"

"[...] Cellular composition comparison between sample groups revealed that vCM1 was associated with myogenic-enriched samples and vCM3 was significantly associated with ischemic-enriched samples. This was validated in an independent cohort using in-situ hybridisation, suggesting that these CM states represent distinct cellular stress states within the acute myocardial infarction phase. (i.e., vCM1; "non-stressed", vCM2 "pre-stressed" and "stressed" vCM3) (Fig. 4c-d; Extended Data Fig. 9h-i). [...]"

2) Differential zonation of the stressed cardiomyocyte state vCM3 across patient groups (New Fig. 4k-o).

"[...] We next estimated the cell dependencies of the stressed cardiomyocyte state -vCM3 with other cell types within each spatial spot and its local neighbourhood (radius of 5 spots) between sample groups (Fig. 4k-o). We observed that the importance of vSMCs in predicting vCM3 within a spot was higher in myogenic and ischemic samples (Fig. 4k), while the importance of fibroblasts and myeloid cells increased in fibrotic samples (Fig. 4k). The local neighbourhood modelling of vCM3 revealed that the abundance of fibroblasts better explained vCM3 in myogenic enriched samples compared to fibrotic samples (Fig. 4l, Extended data Fig. 10i). To gain further insight, we visualised the dependencies of vSMCs and fibroblasts with vCM3 in myogenic enriched samples, and observed that their co-localization occurred in the perivascular niches (Fig. 4n). Overall this demonstrates that the "stressed" CM-state vCM3 occurs in the perivascular niche of larger blood vessels, highlighting the interaction of mesenchymal cells[49] of the perivascular niche with stressed cardiomyocytes in this tissue area. Furthermore we noticed that when comparing RZ with control samples, stressed vCM3s are best predicted by myeloid cells (Fig. 4o). This underlines the importance of immune-CM interactions that could additionally explain the increased arrhythmia susceptibility in the remote regions of the post-infarct heart, since it has been shown that cardiac macrophages influence normal and aberrant cardiac conduction[50,51]. Our results showed that the "stressed"-CM-vCM3 can be found in distinct spatial cell-type neighbourhoods enriched by different compositions of vSMCs, fibroblasts, adipocytes or myeloid cells. [...]"

3) Cardiomyocyte gene-regulatory network analysis (New Fig. 4h-j)

"[...] To infer an enhancer-based gene regulatory network (eGRN) we leveraged our multi-omics data to further investigate molecular mechanisms differentiating the relevant cardiomyocyte states (i.e., vCM1-vCM3) (see Methods). To this end, we paired the cells between snATAC-seq and snRNA-seq data and studied gene-regulatory changes along the cellular continuum from vCM1 to vCM3 (Extended Data Fig. 10a). Next, we estimated an enhancer mediated TF target network by considering TF activity (snATAC-seq), expression of TF and target genes (snRNA-seq), and motif supported peak-to-gene links (Extended Data Fig. 10b-d). Clustering of these TFs to the target network revealed three major modules with each corresponding to a distinct cardiomyocyte state (Extended Data Fig. 10e). [...]"

4) Fibroblast trajectory and gene-regulatory network analysis (New Fig. 13a-d)

"[...] To understand the regulatory mechanisms of these stromal cell differentiation processes we inferred a fibroblast eGRN (Fig. 6g, Extended Data Fig. 13 a-b). Clustering resolved two eGRN modules while each corresponded to a distinct fibroblast state (Extended Data Fig. 13c) and identified potential regulators of myofibroblast differentiation (Fig. 6g).

Among the transcription factors regulating the Fib1 module was KLF4, which regulates diverse cellular functions[61] including cellular growth arrest[62], and is also one of the original reprogramming factors of induced pluripotent stem cells[63]. Our network analysis highlighted the role of KLF4 regulating SCARA5 and PCOLCE2 expression in Fib1, while it also targets MBLN1, an important regulator of cardiac wound healing[64] and fibroblast to myofibroblast transition[65]. Concordantly, we observed a reduced KLF4 binding activity and reduced SCARA5 expression in our pseudotime analysis (Fig. 6h), highlighting the role of KLF4 as a putative inhibitor of fibroblast activation. Among the transcription factors identified in the Fib2 module were TEAD3 (an effector of the Hippo pathway), GLI2 (hedgehog pathway) and RUNX2, which have been previously identified as regulators of myofibroblast differentiation[66,67] (Fig. 6h, Extended Data Fig. 13d-e). [...]"

5) Analysis of myeloid-cell heterogeneity and spatial modelling of fibroblast-myeloid interactions and validation using in-situ hybridization on FFPE sections (New Fig. 6g-i, New Ext. Data Fig. 13a-d)

"[...] We observed that the presence of SPP1+ macrophages better predicted all fibroblasts states compared to other myeloid cell states, with a higher importance for myofibroblasts within a spot and in the local neighbourhood (Fig. 6l, Extended Data Fig. 15a). Myofibroblasts marker expression aligned with a gradient of expression of the markers of SPP1+ macrophages (Fig. 6m). This pattern was also recovered by our cell-type niche definition, where the inflammatory niche 5 was surrounded by the fibrotic-rich niche 4 (Extended Data Fig. 15b), which we could confirm by a higher expression of SPP1+ macrophages and myofibroblast marker genes in niche 5 compared to niche 4 (Extended Data Fig. 15c).[...]"

6) We are releasing our atlas data via the Human Cell Atlas Initiative (HCA, www.humancellatlas.org) and via the explorable data portal of the Chan Zuckerberg Initiative (cellxgene, https://cellxgene.cziscience.com/).

The analysis presented does not specifically point towards a graded, pseudotime-ordered progression in the fibroblasts. In the UMAP in Fig 6a, I see an unclear trajectory structure, perhaps with branching occurring, but it is also possible the pseudotime model is overfitting and the actual gene expression landscape is far more complex (multiple distinct populations, or multiple transitions occurring within multiple populations). The heat map presented would also appear to support this—I do not see a whole lot of marker overlap, but rather what appears to be three rather distinct populations with few examples of cells in transition. What additional evidence is there to support that these fibroblasts are all progressing along a single trajectory?

We thank the reviewer for these important points and acknowledge that there are many possible trajectories. Here we focused on myofibroblasts, for which we leveraged our prior insights obtained from human and mouse kidney (Kuppe et al., Nature 2021, PMID: 33176333; Li et al., Nat Com 2021, PMID: 34737275).

To increase the robustness of our findings as much as possible, we have increased the overall dataset to obtain data from different human disease groups (myogenic group comprising control specimens, remote zone and border zone as well as ischemic and fibrotic specimens)

and time points which allowed us to incorporating temporal information by statistically comparing the cellular composition dynamics. For example, we observed that the *SCARA5*+ fibroblasts had the highest proportion in the myogenic samples while POSTN+ myofibroblasts were significantly enriched in the ischemic group.
Furthermore, we have improved our computational strategy in the revised manuscript for this analysis as follows:

- Instead of analysing the snRNA-seq and snATAC-seq data separately (Old Fig. 6a and Old Fig. 6h), we integrated the snRNA-seq and snATAC-seq for fibroblasts from all samples and performed clustering analysis (New Extended Data Fig. 12a). This approach allowed us to define the same sub-clusters (i.e., Fib1-4) shared by multiple samples and modalities (New Fig. 6a).
- To build a trajectory for the differentiation process, we here only included the relevant sub-clusters of fibroblasts (i.e., *SCARA5*+ fibroblasts and *POSTN*+ myofibroblasts) based on our previous works on human and mouse kidney myofibroblasts origins (Kuppe et al., Nature 2021, PMID: 33176333; Li et al., Nat Com 2021, PMID: 34737275) rather than using all of the fibroblasts.
- We inferred the trajectory using the method from ArchR (Granja et al., Nat Gen 2020, PMID: 33633365) based on a diffusion map (Haghverdi et al., Nat Methods 2016, PMID: 27571553) (New Fig. 6e).

To further understand the differentiation of fibroblasts and as suggested by the Referee #1, we performed a transgenic mouse experiment using PdgfrβCreER;tdTomato inducible fate tracing of all mesenchymal cells in myocardial infarction (LAD ligation) with subsequent scRNA-sequencing of tdTomato sorted (FACS) cells at different timepoints (day 0, 4, 7, 14) (New Extended Data Fig. 12i-l). We integrated and clustered the cells from all timepoints. To annotate the clusters, we also integrated the mouse and human fibroblasts and performed label transfer using Seurat, uncovering three sub-clusters in mouse data (i.e., Fib1-3). (New Extended Data Fig. 12m-n). We observed that the Fib1 (*SCARA5*+) population decreased over time while the Fib2 (*POSTN*+) population increased (New Extended Data Fig. 12o-p).

Based on these observations, we inferred a pseudotime trajectory from Fib1 to Fib2 (myofibroblasts) in the human samples which was further supported by an increased enrichment of extracellular matrix (ECM) score and of ECM-related biological Gene Ontology (GO) processes (New Fig. 6e-f; New Extended Data Fig. 12q). Our revised heatmaps of TF binding activity, TF expression and gene expression showed a clear transition pattern of molecular programs (New Extended Data Fig. 13a-b).

We have adjusted the paper accordingly and the methods section to explain in detail our updated trajectory approach which now shows a more graded pseudotime ordered change in cell-states of fibroblasts.

"[...] Next, we produced a diffusion map (DM) and created trajectories in this space using the function addTrajectory from ArchR (v1.0.1) [...]"

The final insight of the paper, for which there is one in vitro validation experiment performed—is the nomination of RUNX1 as a key effector of myofibroblast differentiation. Again, I am not in this field, but a quick google search for RUNX1 and myofibroblast did reveal literature that

suggests this insight is not entirely novel. For example, how does the observation made by the authors meaningfully differ from the work reported in PMID 32341028 and PMID 25313057?

We agree that the link of RUNX1 being involved in myofibroblast differentiation has been made before, in zebrafish heart (Koth et al., Development 2020, PMID: 32341028) or in the cell-culture using bone marrow-derived mesenchymal stem cells (Kim et al., PNAS 20214, PMID: 25313057) but not in mammalian myocardial infarction. We also recently identified and validated a role of RUNX1 in fibroblast to myofibroblast differentiation in the mouse kidney (Li et al. Nat. Com. 2021, PMID: 34737275). We therefore agree that this finding is not particularly exciting or novel. However, we have provided the first evidence that RUNX1 is critically involved in human heart fibrosis (Extended Data Fig. 12b) and in human myofibroblast differentiation (Extended Data Fig 12d). We have now removed the RUNX1 panels from the main figure (Old Figure 6k,l+n, New Fig. 6) and included the Old Figure Panel 6n to the New Extended Data 12d).

We adapted the section on RUNX1 in the manuscript including the two suggested references of the reviewer:

"[...] Notably, Fib2 also showed an up-regulation of RUNX1 which we have recently reported as being involved in kidney myofibroblast differentiation[58]. A role of RUNX1 in myofibroblast differentiation has been suggested in-vitro using mesenchymal stem cells[59] and in-vivo in zebrafish heart injury[60]. Overexpression of RUNX1 in human heart PDGFRb cells led to increased myofibroblast differentiation and matrix expression (Extended Data Fig. 12d) [...]"

Smaller points:

- Some cardiomyocytes are multinucleated. Can the authors comment on how snRNA-seq and snATAC-seq might be affected by the (likely) additional correlation between nuclei derived from the same cell?

We agree with the reviewer that this is a very interesting question. A study found that about 25% of all human adult cardiomyocytes are binucleated and about 1% trinucleated (Bergmann et al., Cell 2015, PMID 26073943). Previous studies have indicated multiple biological roles of polyploid cardiomyocytes including distinct physiological roles like regeneration, higher resistance to stress and apoptosis (reviewed in Derk et al. Circulation Research 2020, PMID: 32078450). However, Yekelchyk et al. (Basic Research Cardiology 2019, PMID: 31399804) have performed a single cell study that demonstrated that mono- or multinucleated ventricular cardiomyocytes are transcriptionally homogenous.

Given this transcriptional homogeneity, these multinucleated cardiomyocytes would not affect the cell-type and cell-state annotation. One clear effect that multinucleated cardiomyocytes will have in the analysis is in the overestimation of the absolute number of cardiomyocytes. Nevertheless, all of our compositional analyses are based on proportions, which are relative to the analysed sample.

We added the following text to the discussion in our revised manuscript:

"*[...]* Furthermore, we cannot exclude an overestimation of cardiomyocytes in our cell type proportion analysis, since 25% of adult human cardiomyocytes are binucleated[87], however reported to be transcriptionally homogenous[88]. *[...]*"

- Have the authors tried localizing cardiac-associated GWAS signals to their clusters and spatial data? This might be an interesting way of leveraging the human data in away that is unique, since it may be much harder to make credible conclusions about this from existing mouse datasets.

We appreciate this suggestion and we have now included this analysis. We downloaded GWAS summary statistics for 4 MRI based left ventricle (LV) function parameters (Pirruccello et al., Nat Commun 2020, PMID: 32382064) from the Cardiovascular Disease Knowledge Portal (https://cvd.hugeamp.org/) as we anticipated that SNPs relevant to LV function might provide the most biologically relevant information for the cellular composition of myocardial tissue. For each phenotype, GWAS summary statistics were clumped with Plink (Chang et al., GigaScience 2015, PMID: 25722852) to identify index SNPs using the European samples from 1000Genomes as a reference population.

The cell-type-specific GWAS signal enrichment was performed using gchromVAR (Ulirsch, Jacob C., et al. 2019, PMID: 32888494) and enrichment scores were normalised to z-scores. Our analysis revealed that SNPs associated with stroke volume (SV) and left ventricular end-diastolic volume were associated with endothelial cells (New Extended Data Fig. 5i). This result is consistent with endothelial cell function in cardiac relaxation and dilation (Ruetten et al., Cardiovasc Res. 2005, PMID: 15914105). SNPs associated with left-ventricular end-systolic volume and left ventricle ejection fraction (EF) enriched in cardiomyocytes instead, supporting the relationship between contraction and these LV measures. As a next step we associated the genetic variants with spatial data by mapping the information to each spot and observed spatially distributed GWAS signals, which is in turn associated with the location of these cells in space (New Extended Data Fig. 5j).

We added the following text to our revised manuscript:

"[...] To test the association of genetic variants with cell- types, we performed enrichment analysis based on cell-type-specific pseudo-bulk ATAC-seq profiles and cardiomyopathy-related single nucleotide polymorphisms (SNPs) obtained from genome wide association analysis (GWAS) studies[19]. We focussed on SNPs relevant forto left ventricular function since we hypothesised that these might provide the most biologically relevant information for the cellular composition of myocardial tissue. This analysis revealed that SNPs associated with stroke volume (SV) and left ventricular end-diastolic volume were enriched in endothelial cells (Extended Data Fig. 5i), consistent with the role of the endothelial cells in cardiac relaxation and dilation[20]. SNPs associated with left ventricular end-systolic volume and left ventricular ejection fraction were enriched in cardiomyocytes, supporting the relationship between contraction and these left ventricular measures. We also visualised the spatial distribution of GWAS signals by mapping SNPs associated with left ventricular ejection fraction to each spot from spatial transcriptomics (Extended Data Fig. 5j). In summary, our integrated spatial atlas allowed us to map cell-type abundance, signalling pathway activities, TF binding activity, and GWAS signals across the complete spectrum of cardiac tissue zonations providing an in-depth view at tissue remodelling processes following myocardial infarction in humans. [...]"

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have extensively revised their manuscript and now have an excellent resource with careful analysis to share.

Referee #2 (Remarks to the Author):

The authors made an excellent effort to address all my comment. I have no further concerns and support publication of the manuscript.

Referee #3 (Remarks to the Author):

By adding more samples and expanding on the depth of the analyses and clarification the authors we able to improve the manuscript significantly. Although the biological follow up remains a bit limited, the study will likely serve as a relevant data source for many other investigators.
In this updated version the authors were able to answer most issues that were raised by the reviewers and there are no further comments

Referee #4 (Remarks to the Author):

The authors have submitted an extensively revised manuscript, more than doubling their sample size, adding additional data modalities, and refining a lot of analyses that in the original submission were not imperfect or incomplete. They have also made their data available on a site that enables interaction with the data. They have satisfied my concerns and I see this as likely to be a very important and informative resource for the cardiovascular biology community.