nature research

Peer Review Information

Journal: Nature Genetics
Manuscript Title: Cell-specific chromatin landscape of human coronary artery resolves regulatory mechanisms of disease risk
Corresponding author name(s): Dr Clint Miller

Reviewer Comments & Decisions:

Decision Letter, initial version:

13th Jul 2021

Dear Clint,

Your Article entitled "Cell-specific chromatin landscape of human coronary artery resolves regulatory mechanisms of disease risk" has now been seen by 3 referees, whose comments are attached. While they find your work of potential interest, they have raised serious concerns which in our view are sufficiently important that they preclude publication of the work in Nature Genetics, at least in its present form.

While the referees find your work of some interest, they raise concerns about the strength of the novel conclusions that can be drawn at this stage.

In brief, two referees are supportive of your study, but the other has raised an important issue.

Reviewer #1's report regards the number of cells analysed as much too low. They think that this means that the downstream analyses will be underpowered.

Reviewers #2 and #3, on the other hand, are supportive, saying this is an important dataset for the CAD field. Reviewer #2's comments seem largely minor. Reviewer #3, however, thinks that the final part of the study - on PRDM16 and TBX2 - is preliminary and should be expanded.

In our reading of these reports, we find Reviewer #1's comments on cell number to be particularly concerning, and would have to be fully addressed in a revision. We also noted that some of Reviewer #1's specific comments on how your dataset is underpowered (e.g. the caQTL analysis, cross-disease stage comparisons) may not apply to all of the results presented. The ideal solution would be to

analyse more cells, but we acknowledge that patient samples are difficult to collect. In that case, we believe that a detailed description of your analysis, including power calculations, may be sufficient to assuage this concern.

We also thought that Reviewer #1's comments for an integrative analysis with your past RNAseq dataset, and Reviewer #2's request for further work on PRDM16 and TBX2, are important and would also greatly improve the impact of your work.

Should further experimental data allow you to fully address these criticisms we would be willing to consider an appeal of our decision (unless, of course, something similar has by then been accepted at Nature Genetics or appeared elsewhere). This includes submission or publication of a portion of this work someplace else.

The required new experiments and data include, but are not limited to those detailed here. We hope you understand that until we have read the revised manuscript in its entirety we cannot promise that it will be sent back for peer review.

If you are interested in attempting to revise this manuscript for submission to Nature Genetics in the future, please contact me to discuss a potential appeal. Otherwise, we hope that you find our referees' comments helpful when preparing your manuscript for resubmission elsewhere.

Sincerely,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

Referee expertise:

Referee #1: single-cell genomics, including scATAC.

Referee #2: cardiovascular disease, genetics.

Referee #3: cardiovascular disease, gene regulation.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
In this article, the authors have applied single-cell ATAC-seq (scATAC-seq) to profile 28,316 cells across coronary artery segments from 41 patients with varying stages of Coronary artery disease (CAD). They identified 14 distinct cellular clusters and mapped ~320,000 accessible sites across these cells. The authors also identified cell type-specific elements, few transcription factors, and attempted to do chromatin accessibility quantitative trait locus (caQTL) mapping. They also used cis-regulatory elements (CREs) to link genes to disease-associated transcription factors such as PRDM16 and TBX2.

The authors claim to provide a single cell atlas for coronary artery and help interpreting cis-regulatory mechanisms in CAD.

A) It's technically a single-nucleus ATAC-seq as they cannot isolate cells from frozen samples to begin with. In the manuscript it should be relabeled as snATAC-seq

B) The most critical aspect of the paper is the number of nuclei obtained after snATAC-seq - 28,316 cells. While superficially it may look like a decent number of nuclei for all the analysis, the fact that those cells are coming from 41 samples underscores the fact that the dataset is underpowered to dissect out inter-sample group-level differences and more importantly caQTLs. A closer look at the supplementary data (Supplementary Figure 3b-c) reveals that 50% of samples (about 19 samples) had less than 500 captured cells , with at least 8 samples less than 250 cells. More striking is the fact that only 8 samples (out of 41 samples -- 19%) had more than 1000 captured cells, and none over 1500 cells. The difference between different biological groups is even more striking with 87% of category 3 (atherosclerotic) samples having less than 500 cells while significantly more cells were obtained from category 1 samples (controls). As such the group differences become confounded with huge cell-type capture-rate differences and it is bound to affect data analysis downstream.

C) The inherent noise in snTAC-seq data, compounded by the fact that the authors only captured less than 750 cells from majority of samples which are now sub-clustered into 14 clusters suggest that even cluster differences for small-to -medium sized clusters is grossly underpowered. Case-in-point is subclustering of smooth muscle clusters 6 and 7 which are less than 2000 cells each. While the authors do not provide how many cells from each sample contribute to clusters 6 and 7, the average number of cells for cluster 6 and 7 from each sample will be less than 50 cells/sample (2000cell/41samples = 48 cells/sample). With that low resolution of cells, I will be hard-pressed to find any meaningful analysis for these clusters.

The situation is not that different in the 2 largest clusters (Smooth muscle cells, cluster 4 and 5) with about 6000 cells in each cluster and average contribution from each samples will be 150 cells/sample

D) Can the authors integrate their data with their own snRNAseq data, now in bioRxiv (https://doi.org/10.1101/2020.10.27.357715) ? More specifically, do the CREs line up caQTLs and eQTLS from the snRNAseq data? What is the overlap of open chromatin accessibility at the promoter of the first exon of a gene (snATAC-seq) compared to its expression (from snRNA-seq)?

E) The authors should read this paper (https://doi.org/10.1038/s41467-020-19365-w) on effective design of single-cell sequencing experiments for cell-type-specific QTL analysis, applicable for both eQTLs and caQTLs. The number of cells captured is far more critical than the number of samples used or read-depth of each cell.

Reviewer #2:
Remarks to the Author:
This manuscript describes a tour de force employing single-cell ATAC-seq to 28,316 cells in 3

epicardial coronary artery segments from patients with and without coronary atherosclerosis to identify cis-regulatory mechanisms in VSMC transition states. They further extend the findings of recent GWAS using an integrative approach to identify plausible molecular mechanisms whereby a risk locus confers an effect in a cell specific manner. Further they identify PRDM16 and TBX2 as transcriptional regulators in SMCs.

The unique value of this work stands as the first single-cell atlas of human coronary artery chromatin accessibility that is a unique resource of value to the scientific community.

Comments.
Figure 1 a. Are the reasons for transplant rejection of donor hearts known? Were calcified samples excluded? Of note, SMC can also differentiate towards osteoblastic cells, again a unique SMC phenotype.

Figure 1 e. Although 3 arterial segments were harvested from 41 individuals, data are shown for 44 samples. Is this because of less than optimal sample quality or processing? Might the unknown cells represent adipocytes?
It is of interest that the atherosclerosis samples with or without adventitia are generally similar in cell type.

Figure 3. These are very nice data that define at a single cell level, marker genes, TF enrichment and differential promoter peaks that extend our understanding of differentiated vs fibromyocytic SMCs. Apropos of sample calcification, in Fig 3f – Does RUNX include RUNX2, a marker of osteogenic differentiation?

Figure 4. It is of interest and perhaps not unexpected that the majority of recently reported CAD GWAS variants demonstrate functionality in vascular SMCs. 4d. Given the ongoing interest in immune function and CAD, the authors may wish to highlight the variants with strong macrophage peaks.

Figure 6. Relevant to the PRDM16 findings, beige (not brown) adipocytes are known to express SMC markers and can be derived from VSMCs under the direction of PRDM16. If present they could be identified by UCP1 expression. Here the authors should review the work of Spiegelman (2014).

Reviewer #3:
Remarks to the Author:
This manuscript by Turner et al utilizes single cell ATAC-seq (scATAC-seq) data from coronary arteries from 41 patients to characterize cis-regulatory regions that are linked to CAD loci. This analysis reveals cell-specific accessible chromatin and potential binding sites at CAD risk loci and elucidates potential mechanisms for smooth muscle cell phenotypic. scATAC-seq studies were published recently on human carotid endarterectomy samples by Depuydt et al (Circ Res, 2020) and Ord et al (Circ Res, 2021), which removes some novelty of the study by Turner. However, the Ord study had data from only 3 patients (all with advanced disease) and ~7000 cells total. The currently manuscript reports data on human right coronary arteries, LAD arteries and left circumflex arteries from 41 patients with various stages of disease and >28,000 individual cells. Because of the inclusion of multiple patients, Turner et al were able to perform chromatin accessibility QTL analyses, which greatly adds to the study. This is also the first scATAC-seq dataset from coronary arteries. Overall, this is an excellent

study and will provide an important resource for the community. The data and analysis are of high quality. There are several questions that remain.

Major comments:

1) The introduction section does not mention the other scATAC-seq datasets on human carotid artery atherosclerotic plaque, but instead refers to studies on cultured cells. The Ord study does not appear to be cited. The omission of these references in the introduction oversells the novelty somewhat.
2) It isn't clear whether the staging of atherosclerosis severity is according to the Stary criteria with four stages (Circulation, 1995). This is standard for the field and would provide more detailed information about patient plaque characteristics. Was pathology performed on plaque samples? While scATAC-seq was performed on various stages of disease, it seems that all of the data is combined rather than analyzed according to disease stage. Is the accessibility of CAD loci altered by disease stage?
3) Are the gene score differences between cell clusters in Fig. 3B statistically significant?
4) Would analysis of 'super enhancers' or clustered ATAC-seq peaks provide any additional information regarding potential functional regulatory regions?
5) Further functional information on PRDM16 and TBX2 in atherosclerosis would be helpful. For the immunofluorescence in Fig. 6E, it would be helpful to include healthy tissue and advanced lesions for comparison to better understand the expression of PRDM16 in atherosclerosis. The methods indicate that healthy and sub-clinical atherosclerosis samples were used, but the healthy controls are not included in the manuscript. Is there eQTL or caQTL data for PRDM16 or TBX2? The data presented for PRDM16 and TBX2 are not entirely convincing and appear to be preliminary.

Minor comments:
1) The heatmap in Fig. 3D could use additional annotation to indicate where the trajectory begins and ends.
2) The legend for the heatmap in Fig. 4D is confusing. Are 'zero' peaks white? It seems that 'zero' peaks are black on the legend, but this doesn't seem to be correct.

Author Rebuttal to Initial comments

We first thank all of the reviewers for their helpful comments and constructive feedback and the opportunity to consider this work for publication in *Nature Genetics*. We have provided extensive new material and revisions to the text that we feel has markedly improved the quality of the manuscript. We have added power calculations and generated new bulk ATAC-seq caQTL results to support our primary results. Importantly, the caQTLs we highlight (for example rs13324341 (*MRAS*) in smooth muscle cells) were replicated in bulk coronary artery ATAC-seq libraries (50,000 nuclei per individual). We have also included histology characterization of these coronary artery samples in healthy and atherosclerotic contexts and more comprehensive immunostaining of PRDM16 in the vessel wall. Below we provide point-by-point responses (in blue) to each comment/remark from the three reviewers. At the end we also detail specific additions/changes we made to the manuscript. Specific changes to the enclosed manuscript text are also highlighted in blue for clarity.

Reviewer Comments:

Reviewer #1:
Remarks to the Author:
In this article, the authors have applied single-cell ATAC-seq (scATAC-seq) to profile 28,316 cells across coronary artery segments from 41 patients with varying stages of Coronary artery disease (CAD). They identified 14 distinct cellular clusters and mapped ~320,000 accessible sites across these cells. The authors also identified cell type-specific elements, few transcription factors, and attempted to do chromatin accessibility quantitative trait locus (caQTL) mapping. They also used cis-regulatory elements (CREs) to link genes to disease-associated transcription factors such as PRDM16 and TBX2. The authors claim to provide a single cell atlas for coronary artery and help interpreting cis-regulatory mechanisms in CAD.

A) It's technically a single-nucleus ATAC-seq as they cannot isolate cells from frozen samples to begin with. In the manuscript it should be relabeled as snATAC-seq

> Thank you, we agree with this comment and we have now changed “scATAC-seq” to “snATAC-seq” throughout the manuscript to more accurately reflect our single nuclei isolations from the frozen samples.

B) The most critical aspect of the paper is the number of nuclei obtained after snATAC-seq - 28,316 cells. While superficially it may look like a decent number of nuclei for all the analysis, the fact that those cells are coming from 41 samples underscores the fact that the dataset is underpowered to dissect out inter-sample group-level differences and more importantly caQTLs. A closer look at the supplementary data (Supplementary Figure 3b-c) reveals that 50% of samples (about 19 samples) had less than 500 captured cells , with at least 8 samples less than 250 cells. More striking is the fact that only 8 samples (out of 41 samples -- 19%) had more than 1000 captured cells, and none over 1500 cells. The difference between different biological groups is even more striking with 87% of category 3 (atherosclerotic) samples having less than

500 cells while significantly more cells were obtained from category 1 samples (controls). As such the group differences become confounded with huge cell-type capture-rate differences and it is bound to affect data analysis downstream.

We appreciate Reviewer 1's comment in this regard. Here, we respond with a few points to clarify our rationale and support our results:

First, we actually conducted two steps of quality control (QC) for nuclei that were captured and sequenced in the snATAC-seq experiments. We captured and sequenced 94,432 total nuclei (ranging from 113 (minimum) and 20,050 (maximum) with an average of 2,146 nuclei/sample and 17,433 unique fragments/nuclei) using the default QC thresholds from the 10x Genomics Cell Ranger ATAC pipeline (**Supplementary Table 2**). We then subjected these nuclei to more stringent QC thresholds (transcription start site (TSS) enrichment ≥7 and ≥10,000 unique fragments per nuclei) (**Supplementary Table 2**). As expected, the high number of unique fragments per cell dictates how informative each individual cell is for downstream analyses and also contributes towards the power to detect *cis*-regulatory elements (Mandric et al., 2020). Although we did not obtain high nuclei numbers in some individuals, this is not unexpected based on the more stringent QC thresholds and the atherosclerotic disease nature of many of the samples profiled. By pooling all individuals together and obtaining 28,316 high quality nuclei, we were able to overcome the fact that we had fewer post-QC nuclei per individual. Importantly, cells belonging to the various clusters were evenly distributed across all of the individuals (**Rebuttal Figure 1**). Thus, when we are comparing clusters and cell types, the differences are unlikely to be driven by a select few individuals.

Second, we feel our study is adequately powered to dissect out differences between cell types and subtypes (for example, differentiated smooth muscle cells vs. fibromyocytes in **Figure 3**). Recently developed power calculators for single-cell RNA-seq analysis (e.g. SCOPIT (Davis et al., 2019)) demonstrate that we are able to detect cell type differences (assuming a cluster of at least 50 cells) for our rarest cell type (mast cells at 0.7% frequency) by sequencing > 9,129 cells. Given that we focused our analysis on the main cell types in the vessel wall, we are well above this threshold to identify cell type differences in chromatin accessibility. To support this we also developed a script to perform a per gene differential accessibility power analysis (**Supplementary Data 9**). Based on these results we had more than enough cells to detect differential accessibility between SMC and fibromyocytes for all of the significant loci. Reviewer 1 makes a valid point that we have fewer cells per sample in category 3 (advanced atherosclerosis) compared to category 1 (healthy controls). This may contribute to the limited number of differentially accessible genes/peaks identified between healthy and diseased individuals (**Rebuttal Figure 2**). However this may also be driven by the continuous rather than dichotomous nature of atherosclerotic plaques, which have a high degree of heterogeneity and plasticity of cell types (van Kuijk et al., 2019). In fact, we have observed consistent results in group level differences in bulk ATAC-seq profiled samples

from this set of patients (unpublished observations). Nonetheless, we acknowledge this as a known limitation in the Discussion and mention the benefit of performing more single nuclei profiling of advanced disease (category 3) samples to better address this question in future studies. Given the difficulty in procuring these precious samples from heart transplants (especially during the pandemic), we have been unable to obtain more unique individual samples for this study.

Third, we believe we suitably capture differences between individuals with respect to the caQTL analysis. To begin to estimate our power for caQTL analyses, we used the powerEQTL package for single-cell RNA-seq based eQTL studies, which assumes a standard linear regression model. Assuming an average of 500 cells captured per individual across 41 individuals, we have 80% power to detect SNP associations for common variants (MAF=0.1) with moderate effect size (beta = 0.3). However, these calculations grossly underestimate the improved power derived from an allele-specific based method such as RASQUAL (Kumasaka et al., 2016), which is designed to detect molecular trait associations from small sample sizes. In fact, we previously demonstrated a 7-fold increase in significant eQTLs discovered with this approach compared to FastQTL in a cohort of 52 human coronary artery SMC (Liu et al., 2018). We used RASQUAL for our caQTL mapping and also limited these analyses to the two major cell types present in each of the samples (SMC and macrophages). Using a more stringent inclusion criteria for the individual samples, we re-ran RASQUAL for these cell types, and discovered 1,984 and 1,210 cis-caQTLs in SMC and macrophages, respectively, at 5% FDR. This is similar to our original results and we have now updated **Figure 5** and **Supplementary Data 6**.

To strengthen our single cell caQTL findings, we attempted bulk ATAC-seq on the remaining frozen coronary arteries for all 41 of these matched patients. 35 of these bulk ATAC libraries were of sufficient quality and used for downstream bulk caQTL analysis. We added methods on the bulk ATAC-seq to the manuscript and added results in **Supplementary Data 7**. Since we did not need to perform single cell capture for these bulk reactions, each bulk ATAC-seq library contained ~50,000 nuclei per individual. Our results show many of our smooth muscle and macrophage single cell caQTLs are also detected as significant bulk caQTLs, with the direction of effects primarily consistent (**Rebuttal Figure 3** and **Supplementary Figure 11**). Most importantly, the single cell caQTLs that we highlight in the manuscript (such as *MRAS*, *MEF2D*, *FCHO1*, *SMAD3*) were also significant bulk caQTLs with consistent direction of effects. For instance, for rs13324341 at the *MRAS* locus, the peak containing this variant is accessible in smooth muscle cells. The T allele associates with greater accessibility in both our smooth muscle caQTL dataset as well as in the bulk coronary caQTL dataset. Furthermore, as we show in **Figure 5B** and **Figure 5C**, many of these caQTLs are significant eQTLs (5% FDR) in GTEx arterial tissues. We acknowledge in the Discussion that certain caQTLs are likely cell-type/state specific whereas other caQTLs are likely to be shared across several cell types and tissues. Also, we acknowledge that larger cohorts are likely needed to fully address this question.

Lastly, to compensate for the reduced power to detect cell type caQTLs (especially for less abundant cell types), we performed machine learning based prediction of CAD regulatory variant function using three different methods (**Methods**). Using the gapped k-mer support vector machine (gkm-SVM), gkm-Explain, and deltaSVM trained on accessible peaks from SMC, macrophages, fibroblasts, endothelial cells, pericytes, T/NK or Mast cells we identified single nucleotide variants predicted to alter chromatin accessibility and potentially transcription factor binding from the reference sequence (**Supplementary Data 8**). We identified a number of top candidate regulatory variants at CAD loci (e.g. *LIPA* and *SMAD3*) that are predicted to function through cell-specific regulatory mechanisms. This powerful and complementary approach could be extended to deep learning based models to enable more exhaustive prediction of variants that may be difficult to prioritize from QTL based methods alone.

It is worth noting that despite following the Omni-ATAC protocol (Corces et al., 2017) that is optimized for frozen tissue and careful optimization, we were unable to capture high numbers of nuclei from these advanced coronary artery plaques. These advanced, category 3 samples had high amounts of extracellular matrix, calcification, and necrosis, and our group in addition to others have had difficulty capturing high numbers of cells/nuclei per sample from fibrous/calcified human tissues, especially atherosclerotic plaques. For example, a recent study from Depuydt et al. in Circulation Research (Depuydt et al., 2020) performed single cell RNA-seq from 18 patients, all of which were human carotid artery atherosclerotic plaques. Across all 18 patients, this study obtained a total of 3,282 cells, which is approximately 234 cells per individual. This study also performed snATAC-seq for 6 individuals (fibro-atheromatous carotid artery plaques) but only obtained a few hundred suitable nuclei per individual. As is the case for many other human atherosclerosis single-cell datasets, only 3-4 individual samples are typically available (refer to our data portal at PlaqView.com), further demonstrating the difficulty in obtaining and analyzing these samples.

Overall, we acknowledge the reviewer's important comments. However, given that this is the first study to perform snATAC-seq of both healthy and diseased human coronary arteries, which are extremely difficult to procure, we feel that 28,316 analyzed high-quality nuclei, post-QC still represents a very valuable resource for the community. *More importantly, based on the detailed explanations above, we feel that analyzing more nuclei per individual would not affect the major conclusions of this manuscript.*

C) The inherent noise in snATAC-seq data, compounded by the fact that the authors only captured less than 750 cells from majority of samples which are now sub-clustered into 14 clusters suggest that even cluster differences for small-to -medium sized clusters is grossly underpowered. Case-in-point is subclustering of smooth muscle clusters 6 and 7 which are less than 2000 cells each. While the authors do not provide how many cells from each sample contribute to clusters 6 and 7, the average number of cells for cluster 6 and 7 from each sample

will be less than 50 cells/sample (2000cell/41samples = 48 cells/sample). With that low resolution of cells, I will be hard-pressed to find any meaningful analysis for these clusters.

The situation is not that different in the 2 largest clusters (Smooth muscle cells, cluster 4 and 5) with about 6000 cells in each cluster and average contribution from each samples will be 150 cells/sample

We acknowledge the reviewer's concern. Overall, we believe our snATAC data can identify biologically meaningful differences between main clusters and in the case of clusters 4-7 (smooth muscle cells), differences between cell sub-clusters (as shown in **Figure 3**). To further clarify this first point, we now provide individual UMAPs separated by each sample to visualize the number of cells contributing to each main cluster and sub-cluster (**Rebuttal Figure 1**). We also include the numbers of cells analyzed per cluster per sample in **Supplementary Table 3**.

Secondly, to address the reviewer's concern for power, we performed a power calculation of detecting differentially expressed genes between main clusters and sub-clusters using the recently developed Hierarchicell R package (Zimmerman and Langefeld, 2021). As noted by the authors of this package, in order to identify at least 1.2 fold changes as statistically significant (power > 0.80), "researchers will need a minimum of 40 samples..and 100 cells per sample." This is assuming a traditional case/control study design. Given that atherosclerosis is a continuous trait, we already achieve maximal power with 100 cells/sample. There are actually only minor differences in power when sampling 100, 250, 500, or 1,000 cells per sample. In fact, the authors demonstrate that as the overall sample size increases (e.g. 20 to 100) there is a drop in power gains by sampling more cells/sample. This is consistent with the per gene sample size power calculation for our differential analysis of the integrated data, as described above (**Supplementary Data 9**).

We did recognize the intrinsic noise in snATAC when performing these sub-cluster differential analyses as shown in **Figure 3**. Instead of comparing cluster 6 vs. cluster 7 from the snATAC data alone, we actually compared cells labelled as 'smooth muscle cells' (SMCs) against cells labelled as 'fibromyocytes' upon integrating human coronary artery (HCA) scRNA-seq data from Wirka et al., Nature Medicine 2019 (Wirka et al., 2019). Integration of scRNA-seq data into snATAC-data has been shown to enhance clustering resolution and enable identification of cell subpopulations not possible with snATAC-seq data alone. Using the integrated data, we were able to annotate 6,518 contractile SMCs and 2,512 fibromyocytes for the downstream differential accessibility analysis. We believe the integration of the snATAC-seq with coronary artery scRNA-seq data strengthens the cell type annotations and overcomes the inherent noise in the gene activity scores derived from snATAC accessibility data. As mentioned above, each individual in the study contains cells belonging to each cluster with nearly equivalent cell compositions across individuals. Further, the lower numbers of captured cells in certain individuals is further mitigated since we pooled all cells together for the smooth muscle

cell vs. fibromyocyte analyses. We think this is a valid approach since the main goal of this differential analysis was to compare contractile and modulated SMCs as opposed to disease vs non-disease status. Thus it remains unlikely that specific individuals are driving the observed differences between clusters.

Nonetheless, we agree that increased sample sizes and/or nuclei/sample may still be necessary to perform sub-cluster comparisons between smaller clusters representing less abundant cell types (e.g. T cells) and more context-specific mechanisms. Since we did not focus on these less abundant cells in this manuscript, this would be an opportunity for follow-up studies. We have added this note in the revised Discussion.

D) Can the authors integrate their data with their own snRNAseq data, now in bioRxiv (https://doi.org/10.1101/2020.10.27.357715) ? More specifically, do the CREs line up caQTLs and eQTLS from the snRNAseq data? What is the overlap of open chromatin accessibility at the promoter of the first exon of a gene (snATAC-seq) compared to its expression (from snRNA-seq)?

We apologize for any confusion. The coronary artery scRNA-seq data analyzed in the bioRxiv paper by Ma et al. was from GSE131778 (Wirka et al., Nature Medicine 2019), which is the same published scRNA-seq dataset used in this manuscript for integration with snATAC-seq. We are co-authors on this original paper and it is currently the only publicly available human coronary artery scRNA-seq/snRNA-seq dataset. We have highlighted the agreement between our snATACseq and scRNAseq from the integration analysis with geneScore (snATACseq) vs. expression (scRNAseq) in **Figure 1D**.

Regarding the alignment of caQTLs and eQTLs through similar CREs, this question is not yet possible as single-cell eQTLs have not yet been mapped. This will be an interesting question to pursue in future multimodal (snATAC/snRNA) analyses of these tissues. We did perform pseudo-bulk level comparisons between snATACseq and scRNAseq by comparing ATAC-seq peak occupancy at the promoters (e.g., 3 kb of TSS) of expressed genes (detected by scRNAseq pseudo-bulk level, e.g., RPKM >=1) for the same cell type. The results of these analyses are now reported in **Supplementary Figure 6C**, with an overall good correlation (Pearson r~0.55). The "gene score" model implemented in ArchR (Granja et al., 2021) that accounts for enhancer activities provides higher performance for estimation of gene expression.

E) The authors should read this paper (https://doi.org/10.1038/s41467-020-19365-w) on effective design of single-cell sequencing experiments for cell-type-specific QTL analysis, applicable for both eQTLs and caQTLs. The number of cells captured is far more critical than the number of samples used or read-depth of each cell.

We thank Reviewer 1 for sharing this paper (Mandric et al., 2020) and agree that this paper provides very good advice for planning experiments for cost. We agree that a high number of cells captured is important for identification of cell type QTLs, but believe this

paper also emphasizes that a high number of individuals is very important. Both of these factors seem to be more important than the read depth per cell. Thus, we feel that the number of nuclei we captured is largely offset by increasing the number of individual samples, as described in the above responses. We have now expanded our Discussion section to emphasize the need for both increased numbers of captured cells and/or numbers of samples for improved cell type QTL discovery and cited the referenced paper.

For the snATAC experiments we aimed for both a high number of individuals and a high number of nuclei/sample. We had sufficient nuclei concentrations for most samples and aimed to capture ~5,000 nuclei/per sample. However, our capture rate from the 10x Chromium Controller was lower than we anticipated. As we previously mentioned we obtained 94,432 total nuclei from Cell Ranger ATAC but filtered more stringently for downstream analyses.

One key difference between our study and this paper is that we used RASQUAL (Kumasaka et al., 2016) to detect cell-type QTLs whereas this paper used Matrix eQTL. Matrix eQTL (Shabalin, 2012) uses linear regression to compare read counts between individuals across genotypes. Typical power calculators for single cell QTL studies (e.g. powerEQTL described above) rely on similar linear models. However, RASQUAL maximizes association detection by combining both between-individual and within-individual allelic effects. Thus RASQUAL is much better equipped to discover QTLs in smaller sample sizes compared to Matrix eQTL. Although the samples analyzed in the RASQUAL paper were from bulk ATAC libraries, RASQUAL was able to detect thousands of caQTLs from only 24 individuals.

Reviewer #2:
Remarks to the Author:
This manuscript describes a tour de force employing single-cell ATAC-seq to 28,316 cells in 3 epicardial coronary artery segments from patients with and without coronary atherosclerosis to identify cis-regulatory mechanisms in VSMC transition states. They further extend the findings of recent GWAS using an integrative approach to identify plausible molecular mechanisms whereby a risk locus confers an effect in a cell specific manner. Further they identify PRDM16 and TBX2 as transcriptional regulators in SMCs.

The unique value of this work stands as the first single-cell atlas of human coronary artery chromatin accessibility that is a unique resource of value to the scientific community.

Comments.
Figure 1 a. Are the reasons for transplant rejection of donor hearts known? Were calcified samples excluded? Of note, SMC can also differentiate towards osteoblastic cells, again a unique SMC phenotype.

> We should clarify that these are not “rejected” hearts in the typical nomenclature (e.g. rejected organs after transplantation). These “rejected” donor hearts are actually candidate donor hearts that were turned down by the transplant surgical team prior to being transplanted. We have now modified the terminology in the text to prevent any confusion. Regardless, while we have detailed information on the individual donor hearts from the United Network for Organ Sharing (UNOS), we do not usually have a specific reason for “rejection” beyond the surgical teams’ discretion. Some common reasons include: donor-recipient incompatibility such as size mismatches between donor and recipient, cerebrovascular incidents, and drug use.
>
> We included calcified coronary artery samples (all in category 3 (**Figure 1**)), yet these samples provided poorer capture rates and lower number of nuclei that passed our quality control thresholds. While the Omni-ATAC protocol generally worked well for frozen coronary samples, nuclei isolation was substantially more difficult for the highly calcified samples. The highly calcified/advanced plaque samples were difficult to break into small pieces and resulted in more debris in the nuclear preparations. As a result, we did not capture as many *bona fide* osteoblastic cells that are present in advanced atherosclerotic lesions. We suspect that this snATAC dataset does include osteochondrogenic cells that are at earlier stages of SMC transdifferentiation towards this cell type. This is based on preliminary integrative analyses with lineage-traced mouse scRNA-seq data, where we identified integrated expression of the chondrocyte *Sox9* marker in SMC-derived cells. Nonetheless, this would be an interesting follow-up study to identify other regulators of osteoblast differentiation by integrating mouse SMC lineage traced scRNA-seq datasets.

Figure 1 e. Although 3 arterial segments were harvested from 41 individuals, data are shown for 44 samples. Is this because of less than optimal sample quality or processing? Might the unknown cells represent adipocytes?
It is of interest that the atherosclerosis samples with or without adventitia are generally similar in cell type.

> We have 44 samples from 41 individuals due to less than optimal sample processing from some extremely diseased/highly calcified samples for 3 individuals on the first attempt. For these samples we obtained a low number of nuclei that passed the ArchR QC thresholds, but still kept the high-quality nuclei for downstream analyses. We then repeated snATAC-seq library preparation and analyses for adjacent segments from these same 3 individuals.
>
> This is an interesting question regarding the potential assignment of the unknown cluster as adipocytes. While perivascular adipocytes and adipocyte signaling play important roles in inflammation and atherosclerosis, we do not think this unknown cluster (cluster 9) represents adipocytes. Prior to sample freezing we carefully dissected the perivascular fat from the outer layer of the coronary arteries. From our experience, other studies do not trim the fat before profiling arterial tissue, which was done here to enrich vascular wall cell types. When we looked up snATAC gene scores for traditional adipocyte marker genes (e.g. *UCP1*, *CITED1*, and *ZIC1*) there was no noticeable enrichment for these genes compared to other cell clusters (**Rebuttal Figure 4**). Lastly, based on the snATAC marker gene scores, snRNA-seq integrated expression and the confusion matrix shown in **Figure 1D**, the 'unknown' cells closely resemble the 'Macrophage' and 'Mast' labels from scRNA-seq. Top marker genes for this cluster include *CD163* (reported to be an M2 macrophage marker) and *CD300LB* (expressed on myeloid cells). Nonetheless, we acknowledge that there is considerable heterogeneity in myeloid/immune cells in atherosclerosis (Lin et al., 2019, Fernandez et al., 2019).

Figure 3. These are very nice data that define at a single cell level, marker genes, TF enrichment and differential promoter peaks that extend our understanding of differentiated vs fibromyocytic SMCs. Apropos of sample calcification, in Fig 3f – Does RUNX include RUNX2, a marker of osteogenic differentiation?

> Thank you, this is an interesting question. For the position weight matrices displayed in **Figure 3f**, the RUNX motif includes RUNX2 (smooth muscle cell calcification marker (Lin et al., 2015, 2016) in addition to the very similar RUNX1 and RUNX3 motifs. We have now added a sentence in the main text in this regard. This would potentially suggest the fibromyocyte SMCs also encompass some osteochondrogenic cells and agree with recent single cell studies highlighting that SMCs can transition to osteogenic or chondrocytic phenotypes. This also agrees with the identification of *TNFRSF11B* (Osteoprotegerin) and *POSTN* (periostin) as top marker genes in fibromyocytes.

Figure 4. It is of interest and perhaps not unexpected that the majority of recently reported CAD GWAS variants demonstrate functionality in vascular SMCs. 4d. Given the ongoing interest in immune function and CAD, the authors may wish to highlight the variants with strong macrophage peaks.

> We agree that macrophages and other immune cells play key roles in atherosclerosis and these cell types have been intensely studied in the CAD field. After smooth muscle cells, macrophages represented the second most abundant cell type in our snATAC dataset. We have now highlighted additional CAD associated variants residing in strong macrophage peaks (e.g. rs7296737 at *SCARB1* and rs17680741 at *TSPAN14*) in the main text. We also highlighted a top regulatory variant in *LIPA* in **Figure 5,** which is predicted to alter macrophage-specific TF binding sites, as identified through our machine learning analysis.

Figure 6. Relevant to the PRDM16 findings, beige (not brown) adipocytes are known to express SMC markers and can be derived from VSMCs under the direction of PRDM16. If present they could be identified by UCP1 expression. Here the authors should review the work of Spiegelman (2014).

> We thank Reviewer 2 for sharing this important paper from Spiegelman's lab (Long et al., 2014). As mentioned, ectopic expression of PRDM16 *in vitro* can convert VSMCs to beige adipocytes (UCP1 being a thermogenic beige adipocyte marker). Additionally, this study emphasizes how VSMC-like cells display similarities to beige adipocytes rather than traditional brown adipocytes. We did attempt to correlate PRDM16 and UCP1 based on imputed gene scores in SMCs, however these were modestly negatively correlated (Pearson r = -0.28).

Reviewer #3:
Remarks to the Author:
This manuscript by Turner et al utilizes single cell ATAC-seq (scATAC-seq) data from coronary arteries from 41 patients to characterize cis-regulatory regions that are linked to CAD loci. This analysis reveals cell-specific accessible chromatin and potential binding sites at CAD risk loci and elucidates potential mechanisms for smooth muscle cell phenotypic. scATAC-seq studies were published recently on human carotid endarterectomy samples by Depuydt et al (Circ Res, 2020) and Ord et al (Circ Res, 2021), which removes some novelty of the study by Turner. However, the Ord study had data from only 3 patients (all with advanced disease) and ~7000 cells total. The current manuscript reports data on human right coronary arteries, LAD arteries and left circumflex arteries from 41 patients with various stages of disease and >28,000 individual cells. Because of the inclusion of multiple patients, Turner et al were able to perform chromatin accessibility QTL analyses, which greatly
adds to the study. This is also the first scATAC-seq dataset from coronary arteries. Overall, this is an excellent study and will provide an important resource for the community. The data and analysis are of high quality. There are several questions that remain.

Major comments:

1) The introduction section does not mention the other scATAC-seq datasets on human carotid artery atherosclerotic plaque, but instead refers to studies on cultured cells. The Ord study does not appear to be cited. The omission of these references in the introduction oversells the novelty somewhat.

> We fully agree with this comment and have now mentioned and cited the new Ord et al. study in Circulation Research (Örd et al., 2021). It came online a few days before we initially submitted. We also cited the original scATAC-seq study on this carotid artery dataset by Depuydt et al (Depuydt et al., 2020).

2) It isn't clear whether the staging of atherosclerosis severity is according to the Stary criteria with four stages (Circulation, 1995). This is standard for the field and would provide more detailed information about patient plaque characteristics. Was pathology performed on plaque samples? While scATAC-seq was performed on various stages of disease, it seems that all of the data is combined rather than analyzed according to disease stage. Is the accessibility of CAD loci altered by disease stage?

> Thank you, we agree that the Stary criteria is the standard in the field to grade the severity of atherosclerosis. We have performed histology analyses on most of these coronary artery samples, which include healthy samples with minimal intimal thickening, early/intermediate atheromas as well as fibro-fatty plaques with calcification. Importantly, whenever possible, we used samples for snATAC-seq that were adjacent to regions of the coronary artery that were used for histology analysis. Representative histology images and quantitation of a subset of samples per disease stage is now included in **Supplementary Figure 4 and Rebuttal Figure 5**. Using Oil Red O (ORO) staining of

lipids we observed an accumulation of lipid laden cells in the subintimal layer of the early atheroma (category 2) and fibroatheroma plaques (category 3). Sirius red and H&E staining also demonstrates increased intimal hyperplasia, collagen type I/III accumulation and decreased lumen diameter of the early atheroma and advanced lesion segments relative to the healthy control segments (category 1). In general, our results are consistent with the Stary (Stary et al., 1995) classification stages in that category 1 represents type I/II lesions with adaptive intimal thickening and initial fatty streak/foam activation, category 2 represents type III/IV lesions (intermediate/advanced atheroma) with more intimal thickening and accumulation of lipid, and category 3 represents type V/VI lesions (advanced fibrocalcific atheroma) with evidence of a lipid core, fibrous cap, and/or calcification.

Reviewer 3 is correct that most of our analyses use data combined for 28,316 nuclei across all individuals. In terms of comparing accessibility at CAD loci according to disease stage, we did not observe many genome-wide significant differences in chromatin accessibility between disease stages. This is expected given the continuous rather dichotomous nature of atherosclerosis progression. The loci we did observe make sense biologically in terms of the implicated genes, with marker genes having higher accessibility in Category 3 linked to inflammation and immune processes (e.g. *CD5*, *CD84*, *CCL4L2,* and *ICAM1*) (**Rebuttal Figure 2**). In contrast SMC marker genes related to contractile function (e.g. *CNN1*, *KCNA5*) were more accessible in Category 1 samples. Unfortunately due to difficulty obtaining high numbers of nuclei from more advanced diseased/plaque samples (Category 3) these samples have fewer analyzed nuclei compared to the other categories, as noted in response to Reviewer 1. Finally, dissecting differences between disease stages remains a challenging task since coronary lesions are very heterogeneous and this type of analysis may be more suitable for spatial genomic approaches to compare omic profiles *in situ*.

3) Are the gene score differences between cell clusters in Fig. 3B statistically significant?

This is a good point, which we have now clarified in a revised **Figure 3B**. The differences in gene scores comparing Cluster 4 + Cluster 7 vs. Cluster 5 + Cluster 6 are statistically significant and now reflected with p values shown in the **Figure 3B** panel.

4) Would analysis of 'super enhancers' or clustered ATAC-seq peaks provide any additional information regarding potential functional regulatory regions?

We agree that this is a very interesting question. In cultured human coronary artery smooth muscle cells ((Miller et al., 2016), GSE72696), H3K27ac ChIP-seq marks (established feature of super enhancers) have previously been used to identify SMC super enhancers. We re-analyzed these datasets using the SICER (Zang et al., 2009) package, which is optimized for broad peak calling from histone modifications and identified super enhancers from H3K27ac peaks >10 kb, as previously described (Wang et al., 2019). These cultured human coronary artery SMC super enhancers were

enriched at SMC marker genes in our snATAC-seq data compared to all other cell types (**Supplementary Figure 7**).

Here, we provide overlaps of accessible chromatin regions in each cell type with long stretches of H3K27ac marks (**Supplementary Table 5**). We find that ATAC-seq peak clusters (ATAC-seq peaks longer than 10 kb) in smooth muscle cells showed the highest association with super enhancers. We further observed that the SMC super enhancers showed significantly higher regulatory potential for the identified SMC marker genes compared to the marker genes from all other cell types (**Supplementary Figure 7**), supporting the additional functional insights gained from this analysis. We highlight an example at the *LMOD1* locus (SMC marker and CAD gene), which harbors two SMC superhancers (**Supplementary Figure 7**). Enhancers at this gene have already been validated experimentally, however other top candidate super enhancers deserve validation using in vitro models.

5) Further functional information on PRDM16 and TBX2 in atherosclerosis would be helpful. For the immunofluorescence in Fig. 6E, it would be helpful to include healthy tissue and advanced lesions for comparison to better understand the expression of PRDM16 in atherosclerosis. The methods indicate that healthy and sub-clinical atherosclerosis samples were used, but the healthy controls are not included in the manuscript. Is there eQTL or caQTL data for PRDM16 or TBX2? The data presented for PRDM16 and TBX2 are not entirely convincing and appear to be preliminary.

We agree with Reviewer 3 in that more functional information would help support our novel findings prioritizing PRDM16 and TBX2 at their respective CAD loci. For the revision we focused primarily on PRDM16 but also included some additional information for TBX2. For PRDM16 we conducted comprehensive immunofluorescence staining in both healthy (n=4) and diseased (n=8) coronary arteries along with whole slide confocal scanning and quantification (**Figure 6**, **Supplementary Figure 15 and Rebuttal Figure 6**). We leveraged the CVPath biorepository of atherosclerotic samples to carry out these more comprehensive analyses in lesions at well defined disease stages. We observed PRDM16 staining colocalized in ACTA2 positive smooth muscle cells in the medial layer of healthy and early atheroma samples, which was reduced in the advanced fibroatheroma samples (both thin cap fibroatheroma and thick cap fibroatheroma). Interestingly, we observed high PRDM16 staining in the vasa vasorum, marking arterioles that are positive for ACTA2. We also observed staining in a few weakly ACTA2 positive arterioles, which could represent pericytes and/or endothelial cells (EC). This is consistent with a recent murine study demonstrating a role for both EC and SMC expressed Prdm16 in regulating flow recovery in post-ischemia PAD models (Craps et al., 2021). While we identified PRDM16 as a SMC-specific marker based on our snATAC-seq data, further studies are needed to dissect the functional interplay in EC and SMCs.

In addition to immunofluorescence, we leveraged data from our lab's bulk RNA-seq data, publicly available arterial scRNA-seq data, and performed additional *in vitro* experiments. While neither of these genes were differentially expressed in bulk tissues from our coronary artery dataset (n=150), PRDM16 and TBX2 were both significantly upregulated in perivascular adipose tissue from diseased coronary arteries (n=44) (Numaguchi et al., 2019) (**Supplementary Table 6 and 7**). By querying human and mouse atherosclerosis datasets we confirmed the SMC and pericyte (and limited EC) expression for PRDM16 and TBX2. (**Supplementary Figure 14**). To gain more functional insight into the mapped regulatory elements for PRDM16 we cloned 4 candidate PRDM16 enhancer sequences overlapping CAD SNPs upstream of a minimal promoter and compared these to two different promoter sequences in standard luciferase reporter assays (**Rebuttal Figure 7**). Candidate enhancer #3 had ~5 fold increase in normalized luciferase activity compared to the empty vector control, consistent with the activity of the PRDM16 promoter. These results in HEK 293T were consistent with assays in immortalized human coronary artery smooth muscle cells, despite the lower magnitude of activation. We observed that PRDM16 expression is greatly reduced in cultured SMCs compared to intact tissue, making systematic validation challenging. Due to the dozens of candidate PRDM16 SNPs highly associated with CAD, overall locus complexity, and difficulty transfecting SMCs, we feel this would be better suited for future in-depth investigations.

In terms of eQTLs for *PRDM16* and *TBX2* in CAD-relevant tissues, we looked up signals at these CAD loci in the GTEx (GTEx Consortium et al., 2017), STARNET (Franzén et al., 2016) and Cardiogenic QTLizer (Munz et al., 2020) databases (**Supplementary Data 10**). We observed nominally significant eQTLs in STARNET and GTEx artery tissues for both genes, however there were modest numbers of eQTLs at both the *PRDM16* and *BCAS3/TBX2* loci in all of the studies. This is not completely surprising as many biologically important genes at GWAS loci often do not harbor many significant eQTLs due to constraint by negative selection (O'Connor et al., 2019; Wang and Goldstein, 2020). Indeed, both PRDM16 and TBX2 have high loss-of-function intolerance probabilities (pLI = 0.999 and 0.964, respectively) supporting their constraint and haploinsufficiency.

With respect to caQTLs, we observed a CAD-associated signal at the *PRDM16* locus (rs10797377, *ACTRT2*, chr1:3012242-3012643) that associates with peak accessibility in SMCs (RASQUAL q value = 1.86×10^{-4}). Moreover, this regulatory element has a highly significant Peak2Gene link with PRDM16 and PRDM16-DT (**Supplementary Data 5**) in which chromatin accessibility strongly correlates with gene expression. Notably, using the activity-by-contact (ABC) based enhancer-promoter linking method (Nasser et al., 2021) in ENCODE human coronary artery, we also identified *PRDM16* and *TBX2* as the target genes from CAD associated SNPs (**Supplementary Table 9**). While there are numerous linked CAD-associated candidate SNPs at these two loci, we believe the lack of additional caQTLs could reflect context-specific effects or buffering effects of allelic variation within enhancers at these loci. In the case of PRDM16, another

independent association signal, rs2493292, is a missense variant, which is predicted to be tolerated/benign (SIFT, PolyPhen, MVP).

Taken together, along with our gene regulatory analyses identifying these two genes as key driver genes in STARNET, these additional functional data demonstrate that PRDM16 and TBX2 are indeed the target genes for their respective CAD loci and may play critical roles in SMC during disease.

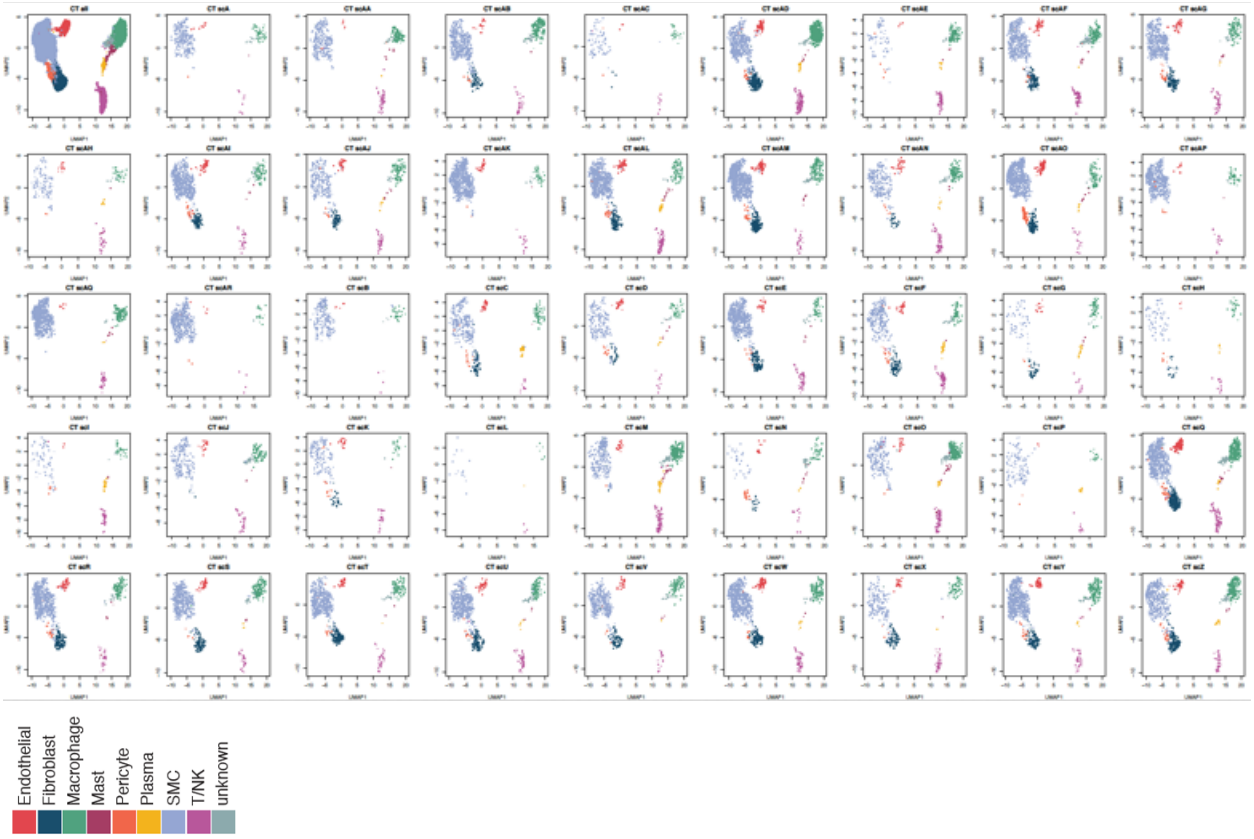
Minor comments:

1) The heatmap in Fig. 3D could use additional annotation to indicate where the trajectory begins and ends.

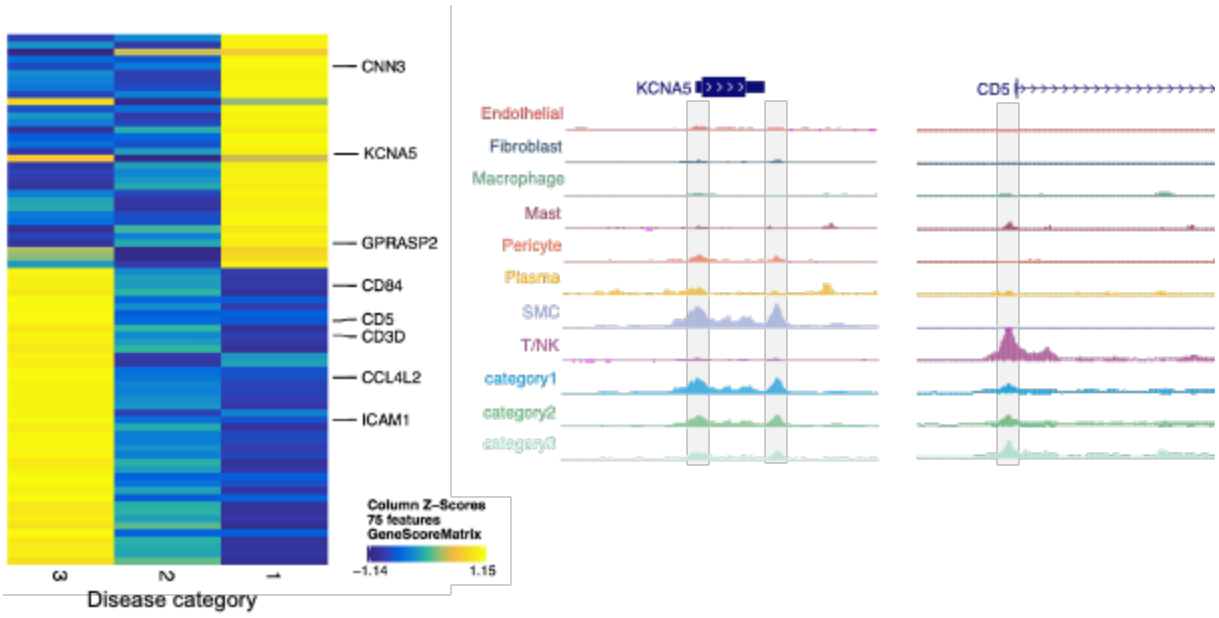
We agree that providing additional annotation to the heatmap improves interpretation of this figure. We have now added labels with arrows to the heatmap to indicate both where the trajectory starts and ends.

2) The legend for the heatmap in Fig. 4D is confusing. Are 'zero' peaks white? It seems that 'zero' peaks are black on the legend, but this doesn't seem to be correct.

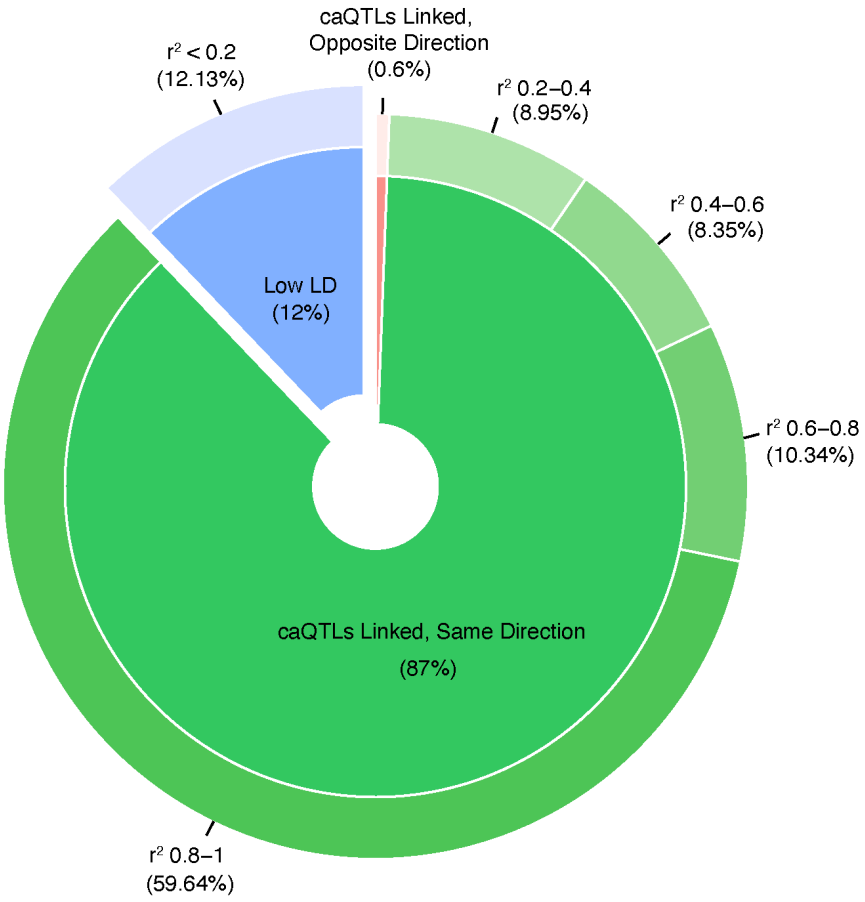
We thank Reviewer 3 for pointing this out. Zero overlaps should be white and the scale should start at 1. We have corrected this in **Figure 4D**.



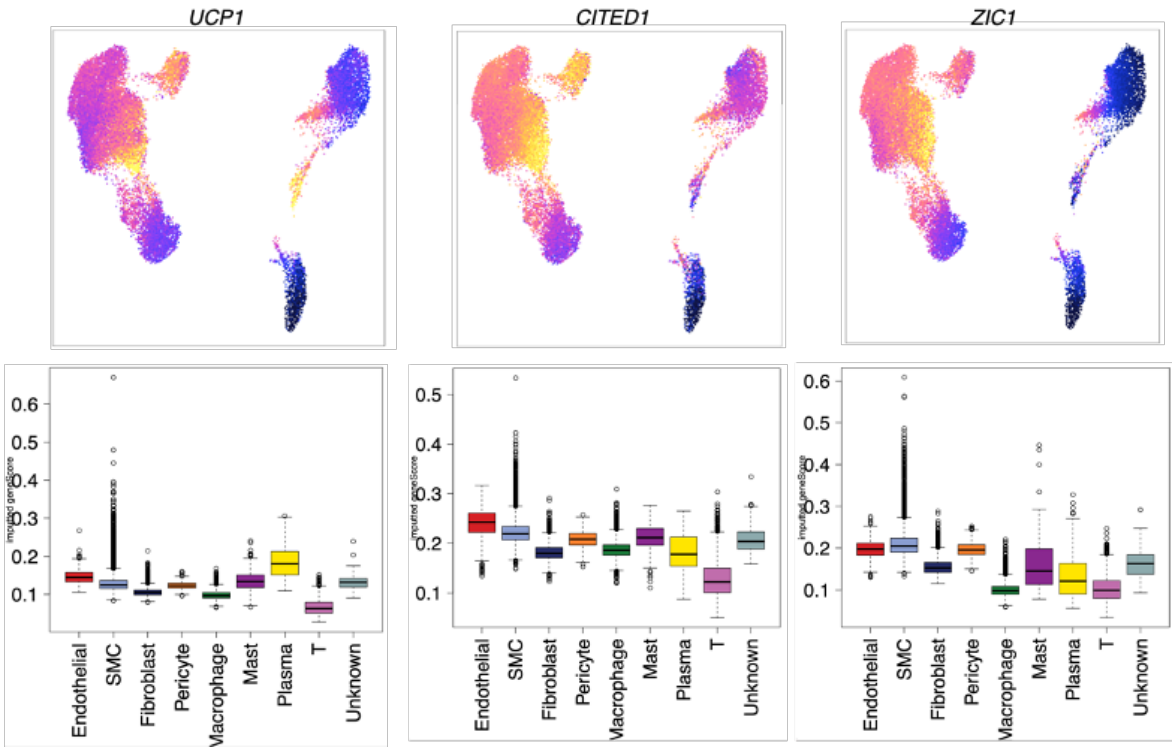
Rebuttal Figure 1. UMAP plots separated by each donor sample showing cells for each main cell cluster are represented from individual snATAC-seq libraries. Note: samples lacking an adventitial layer do not have cells in the fibroblast cluster (dark blue). Also, sample scL was excluded from caQTL mapping due to overall low cell numbers.



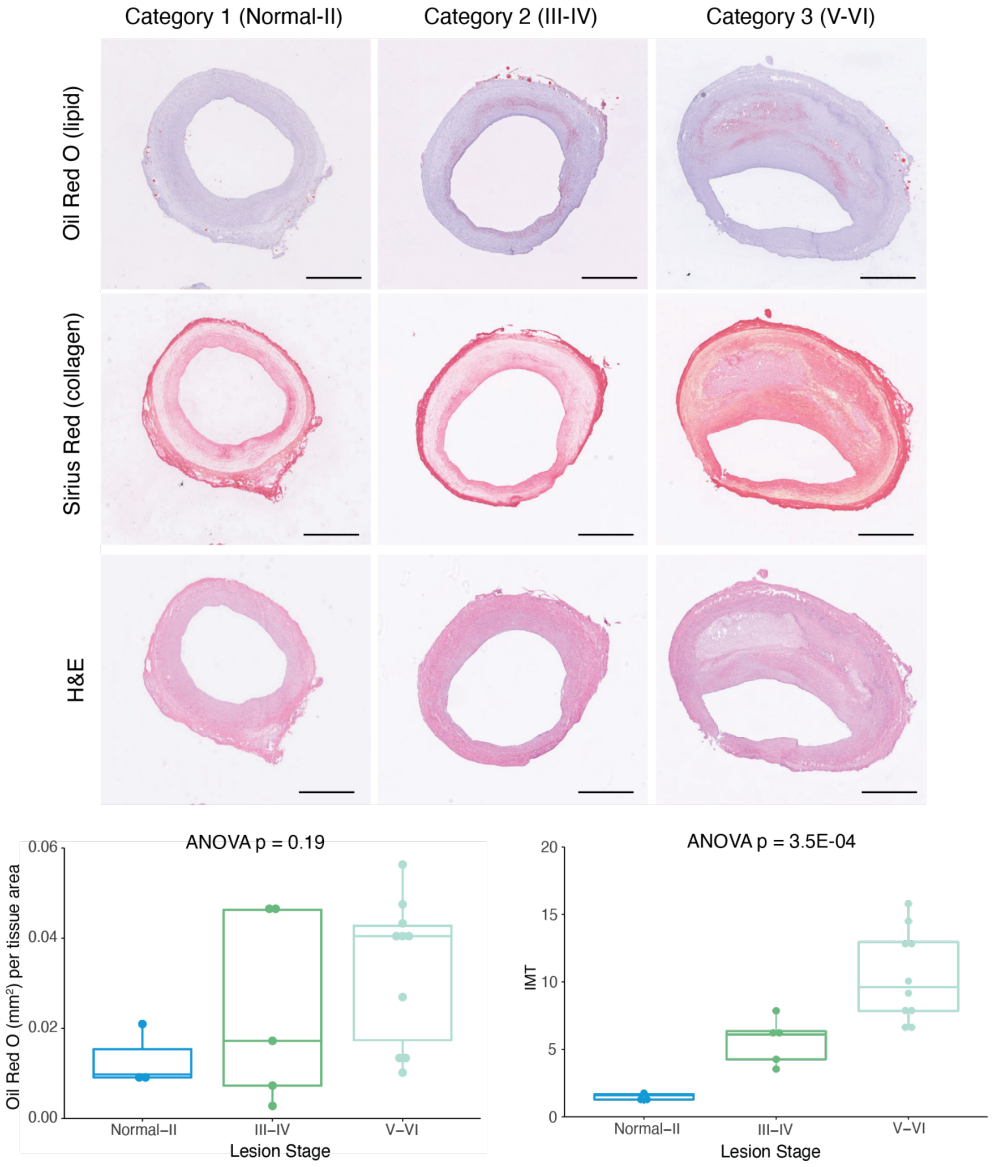
Rebuttal Figure 2. Differentially accessible marker genes between different disease categories. Highlighted genes are immune genes that harbor more accessible peaks in advanced diseased coronary segments (category 3) relative to healthy segments (category 1). Also contractile SMC marker genes are more accessible in healthy artery segments relative to diseased segments. Differential marker genes were detected at FDR <= 0.1 and log2FC > log2(1.5).



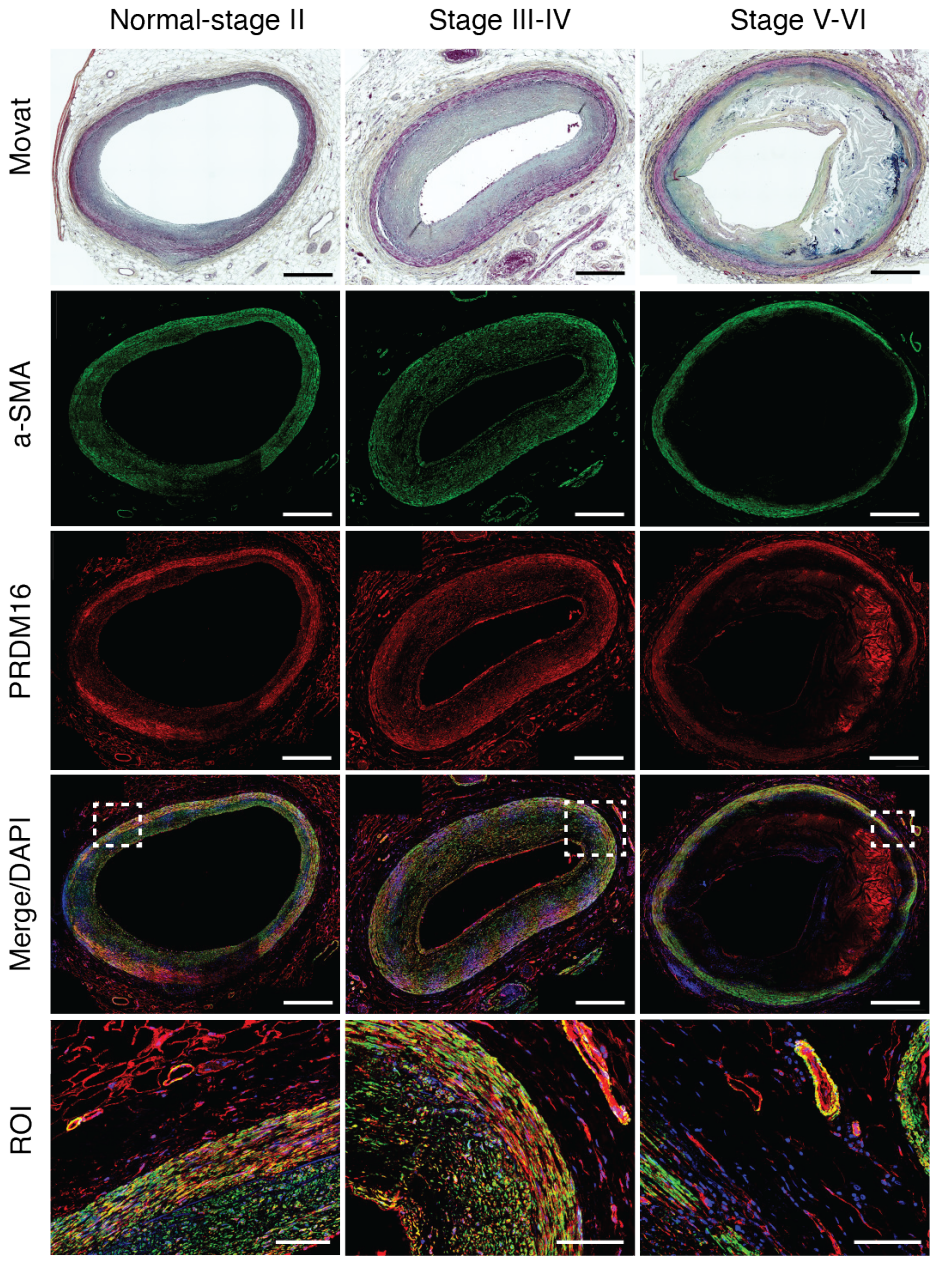
Rebuttal Figure 3. Comparison of effect size directions between smooth muscle cell caQTLs (5% FDR) and bulk coronary artery caQTLs (5% FDR). For this analysis, 503 caQTL peaks are shared between both datasets (peaks with a corresponding significant caQTL variant). The rsID reported in the SMC caQTL results (n=40 individuals) was compared with the rsID reported in the bulk caQTL results (n=35 individuals). Two variants were considered to be in linkage disequilibrium (LD) if the r2 value between them was between 0.2 and 1 (in EUR population). If variants had an r2 value < 0.2 (in EUR population), the variants were considered to be in low LD (blue portion of pie). For the caQTL effect size direction, we considered the RASQUAL Pi statistic. The RASQUAL Pi statistic can range from 0-1, where Pi < 0.5 reflects lower peak accessibility for the alternative allele and Pi > 0.5 reflects higher accessibility for the alternative allele. The effect sizes for linked variants go in the same direction (green portion of pie) if the Pi values in SMCs and bulk coronary artery are both < 0.5 or both > 0.5.



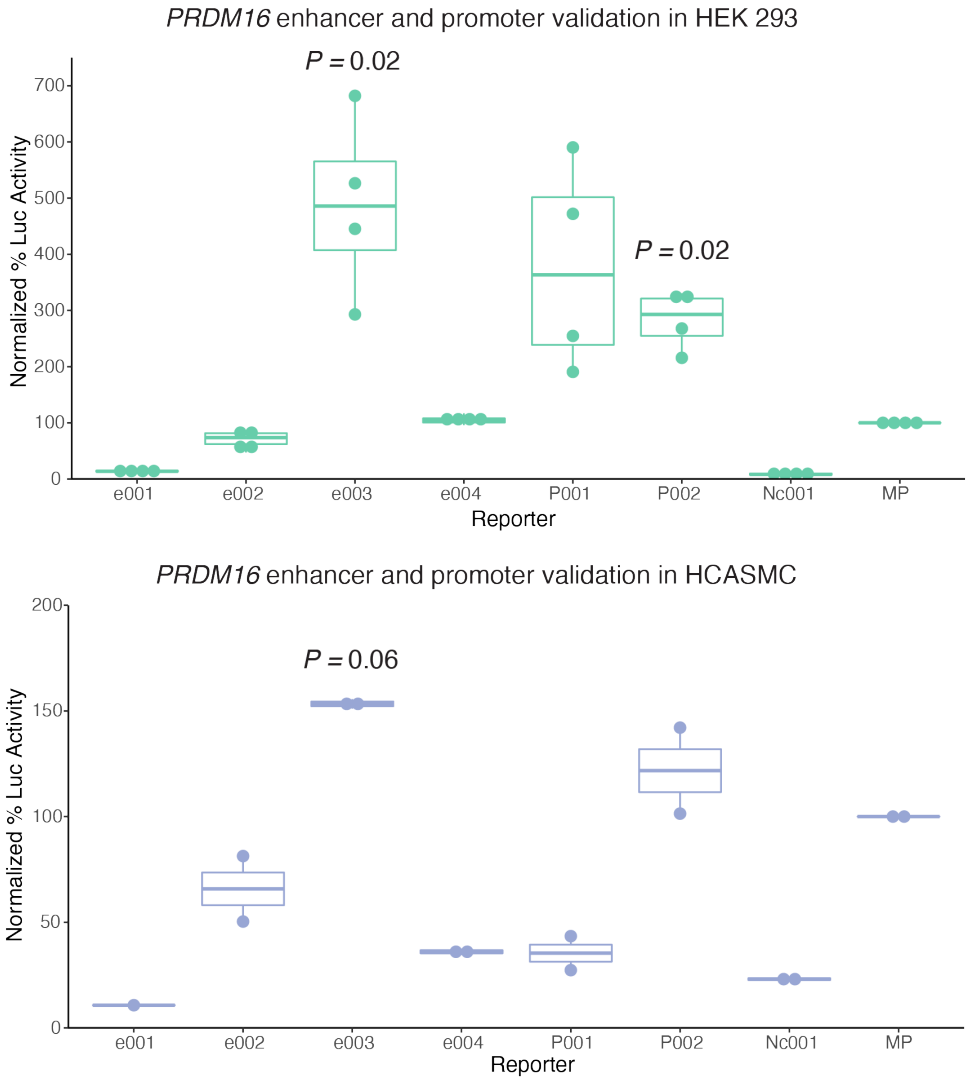
Rebuttal Figure 4. UMAP and box plots of imputed geneScores for known adipocyte marker genes (*UCP1*, *CITED1*, and *ZIC1*) demonstrating the absence of adipocytes in our coronary artery snATAC dataset. This is consistent with the fact that perivascular adipose was trimmed prior to archiving these coronary artery specimens.



Rebuttal Figure 5. Representative histology staining of adjacent frozen human coronary artery sections at different disease categories used for snATAC profiling (n=4 donors per category). Category 1 reflects normal to Stary atherosclerosis stage I/II lesions with adaptive intimal thickening and early lipid (Oil Red O (ORO)) and collagen (Sirius Red) accumulation in the subintimal layer. Category 2 reflects Stary stage III/IV early/intermediate atheroma lesions with increased lipid and collagen accumulation and proliferation (Hematoxylin & Eosin (H&E)). Category 3 reflects Stary stage V/VI advanced fibroatheroma or complex lesions with more severe lipid and collagen deposition as well as lipid core and thin media layer. (Below) Whole slide quantitative results of ORO area (mm^2) normalized to overall tissue area and Sirius Red based quantitation of intima-media thickness (IMT) with maximum intima and average media width captured from >6 automatically defined measurements (Methods). n=3, n=5, and n=10 donors per lesion stage, respectively. ANOVA p-values shown for comparisons across lesion stages. Scale bar = 1mm.



Rebuttal Figure 6. Movat pentachrome staining and PRDM16 (red) and alpha-smooth muscle actin (a-SMA) (green) immunofluorescence staining of atherosclerotic human coronary artery segments - left anterior descending (LAD) from normal-Stage I, Stage III-IV, and Stage V-VI lesions based on Stary classification stages. Whole slide images captured from 20x confocal microscopy stitched tiles. PRDM16/a-SMA co-staining (see arrows) depicted in yellow from merged images. DAPI (blue) marks nuclei. n = 4 per group. Scale bar = 1 mm, except for region of interest (ROI): scale bar = 100 μm.



Rebuttal Figure 7. (Top) Luciferase reporter assays evaluating the transcriptional activity of 4 candidate enhancer regulatory regions overlapping CAD variants in lentiviral transduced HEK 293 cells. Candidate regions were cloned into the pLS-mP-Luc plasmid (Addgene plasmid # 106253). Dots represent mean of triplicates from n=4 biological replicate experiments. Boxes and whiskers represent the median with upper and lower quartiles. (Bottom) Luciferase reporter assays evaluating the transcriptional activity of 4 candidate enhancer regulatory regions overlapping CAD variants in lentiviral transduced human coronary artery smooth muscle cells. Dots represent mean of triplicates from n=2 biological replicate experiments. Box and whiskers represent the median with upper and lower quartiles. e001-e004 are CAD enhancer regions, P001 and P002 are endogenous PRDM16/PRDM16-DT promoter sequences, Nc001 is a negative control enhancer sequence, and MP is the minimal promoter and empty vector from pMCS-Luc. Values are presented as % normalized luciferase of MP reporter. Statistical significance determined from % change of test reporters versus MP reporter using Wilcoxon Rank Test with continuity correction.

Rebuttal Table 1. Candidate regulatory regions at the PRDM16 locus used for luciferase assays in cultured cells.

Enhancer candidate name	**Region (hg38)**	**PRDM16 CAD-associated SNPs**
e001	chr1:2998925-2999935	rs72629460, rs12239064, rs12240128, rs67927838, rs59653178, rs67142023, rs72629462
e002	chr1:3072552-3073451	rs2981890
e003	chr1:3074233-3074650	rs35397508
e004	chr1:3075236-3075530	No SNPs but highly conserved
p001 (*promoter)	chr1:3067941-3069685	PRDM16 promoter
p002 (*promoter)	chr1:3067041-3068642	PRDM16-DT promoter rs2297829
nc001	chr1:3230824-3231620	Negative control sequence

Summary of changes

Figures

Figure	Change/Addition
Figure 3	Added a legend to the trajectory heatmap (Figure 3d)
Figure 4	Changed the start value of the scale from 0 to 1 in Figure 4d
Figure 5	Updated caQTL panels after removing individuals with low numbers of cells
Figure 6	Added additional panels for PRDM16

Supplementary Figures

Supplementary Figure	Change/Addition
4	Added representative histology images for various stages of CAD
6	Added plots showing correlation of snATAC promoter accessibility with integrated scRNA-seq expression
7	Added plots for super enhancer analyses
11	Added donut plot showing the correspondence of effect size directions between smooth muscle caQTLs and bulk coronary artery caQTLs
14	Added UMAP feature plots of PRDM16 and TBX2 in human and mouse atherosclerosis scRNA-seq datasets
15	Added additional representative whole slide histology and confocal immunofluorescence images for PRDM16, a-SMA, and LMOD1

Supplementary Tables/Supplementary Data

We have now moved many of the prior Supplementary Tables to separate Supplementary Data files below. We feel this repackaging makes the supplemental information much more organized than before.

Supplementary Data File	Details

1	Top snATAC marker genes in each coronary artery cell type
2	Consensus set of 323,767 coronary artery peaks across all cell types
3	List of enriched transcription factor motifs within coronary artery cell type peaks
4	Differential peak and promoter analysis results between differentiated smooth muscle cells and fibromyocytes
5	Overlap of CAD GWAS SNPs with coronary artery cell type peaks and Peak2Gene link coordinates. For the Peak2Gene links, the peak coordinates now match the peak set used in the rest of the manuscript
6	Chromatin accessibility QTLs within individual cell types calculated using RASQUAL (shown are SNPs passing 5% FDR threshold). For the updated caQTL analysis we excluded individuals from the analysis if they contained <20 nuclei belonging to the respective cell type
7	Chromatin accessibility QTLs from bulk coronary artery samples calculated using RASQUAL (shown are SNPs passing 5% FDR threshold)
8	Significant results for functional CAD variant prediction using gkm-SVM, gkmExplain, and DeltaSVM
9	Sample size estimations for top fibromyocyte genes (comparing smooth muscle cells and fibromyocytes)
10	eQTLs for PRDM16 and TBX2 in STARNET and GTEx artery tissues

In the new version of the Supplementary Table file we have changed or added the following:

Table	**Change/Addition**
ST2	Added the number of nuclei in each cell type captured and detected using 10x Genomics Cell Ranger ATAC QC
ST3	Added the number of nuclei analyzed in each sample per cell type (after stringent ArchR QC filtering)
ST5	Summary of snATAC peaks overlapping coronary artery smooth muscle cell super enhancers (SE)

Bibliography

Chiou, J., Geusz, R.J., Okino, M.-L., Han, J.Y., Miller, M., Melton, R., Beebe, E., Benaglio, P., Huang, S., Korgaonkar, K., et al. (2021). Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature *594*, 398–402.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Craps, S., Van Wauwe, J., De Moudt, S., De Munck, D., Leloup, A.J., Boeckx, B., Vervliet, T., Dheedene, W., Criem, N., Geeroms, C., et al. (2021). Prdm16 supports arterial flow recovery by maintaining endothelial function. Circ. Res.

Davis, A., Gao, R., and Navin, N.E. (2019). SCOPIT: sample size calculations for single-cell sequencing experiments. BMC Bioinformatics *20*, 566.

Depuydt, M.A.C., Prange, K.H.M., Slenders, L., Örd, T., Elbersen, D., Boltjes, A., de Jager, S.C.A., Asselbergs, F.W., de Borst, G.J., Aavik, E., et al. (2020). Microanatomy of the Human Atherosclerotic Plaque by Single-Cell Transcriptomics. Circ. Res. *127*, 1437–1455.

Fernandez, D.M., Rahman, A.H., Fernandez, N.F., Chudnovskiy, A., Amir, E.-A.D., Amadori, L., Khan, N.S., Wong, C.K., Shamailova, R., Hill, C.A., et al. (2019). Single-cell immune landscape of human atherosclerotic plaques. Nat. Med. *25*, 1576–1588.

Franzén, O., Ermel, R., Cohain, A., Akers, N.K., Di Narzo, A., Talukdar, H.A., Foroughi-Asl, H., Giambartolomei, C., Fullard, J.F., Sukhavasi, K., et al. (2016). Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science *353*, 827–830.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

van Kuijk, K., Kuppe, C., Betsholtz, C., Vanlandewijck, M., Kramann, R., and Sluimer, J.C. (2019). Heterogeneity and plasticity in healthy and atherosclerotic vasculature explored by single-cell sequencing. Cardiovasc. Res. *115*, 1705–1715.

Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat. Genet. *48*, 206–213.

Lin, J.-D., Nishi, H., Poles, J., Niu, X., Mccauley, C., Rahman, K., Brown, E.J., Yeung, S.T., Vozhilla, N., Weinstock, A., et al. (2019). Single-cell analysis of fate-mapped macrophages reveals heterogeneity, including stem-like properties, during atherosclerosis progression and regression. JCI Insight *4*.

Lin, M.-E., Chen, T., Leaf, E.M., Speer, M.Y., and Giachelli, C.M. (2015). Runx2 expression in smooth muscle cells is required for arterial medial calcification in mice. Am. J. Pathol. *185*, 1958–1969.

Lin, M.-E., Chen, T.M., Wallingford, M.C., Nguyen, N.B., Yamada, S., Sawangmake, C., Zhang, J., Speer, M.Y., and Giachelli, C.M. (2016). Runx2 deletion in smooth muscle cells inhibits vascular osteochondrogenesis and calcification but not atherosclerotic lesion formation. Cardiovasc. Res. *112*, 606–616.

Liu, B., Pjanic, M., Wang, T., Nguyen, T., Gloudemans, M., Rao, A., Castano, V.G., Nurnberg, S., Rader, D.J., Elwyn, S., et al. (2018). Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci. Am. J. Hum. Genet. *103*, 377–388.

Long, J.Z., Svensson, K.J., Tsai, L., Zeng, X., Roh, H.C., Kong, X., Rao, R.R., Lou, J., Lokurkar, I., Baur, W., et al. (2014). A smooth muscle-like origin for beige adipocytes. Cell Metab. *19*, 810–820.

Mandric, I., Schwarz, T., Majumdar, A., Hou, K., Briscoe, L., Perez, R., Subramaniam, M., Hafemeister, C., Satija, R., Ye, C.J., et al. (2020). Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. Nat. Commun. *11*, 5504.

Miller, C.L., Pjanic, M., Wang, T., Nguyen, T., Cohain, A., Lee, J.D., Perisic, L., Hedin, U., Kundu, R.K., Majmudar, D., et al. (2016). Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. Nat. Commun. *7*, 12092.

Munz, M., Wohlers, I., Simon, E., Reinberger, T., Busch, H., Schaefer, A.S., and Erdmann, J. (2020). Qtlizer: comprehensive QTL annotation of GWAS results. Sci. Rep. *10*, 20417.

Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238–243.

Numaguchi, R., Furuhashi, M., Matsumoto, M., Sato, H., Yanase, Y., Kuroda, Y., Harada, R., Ito, T., Higashiura, Y., Koyama, M., et al. (2019). Differential phenotypes in perivascular adipose tissue surrounding the internal thoracic artery and diseased coronary artery. J. Am. Heart Assoc. *8*, e011147.

O'Connor, L.J., Schoech, A.P., Hormozdiari, F., Gazal, S., Patterson, N., and Price, A.L. (2019). Extreme polygenicity of complex traits is explained by negative selection. Am. J. Hum. Genet. *105*, 456–476.

Örd, T., Õunap, K., Stolze, L.K., Aherrahrou, R., Nurminen, V., Toropainen, A., Selvarajan, I., Lönnberg, T., Aavik, E., Ylä-Herttuala, S., et al. (2021). Single-Cell Epigenomics and Functional Fine-Mapping of Atherosclerosis GWAS Loci. Circ. Res. *129*, 240–258.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358.

Stary, H.C., Chandler, A.B., Dinsmore, R.E., Fuster, V., Glagov, S., Insull, W., Rosenfeld, M.E., Schwartz, C.J., Wagner, W.D., and Wissler, R.W. (1995). A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association.

Circulation *92*, 1355–1374.

Wang, X., and Goldstein, D.B. (2020). Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. Am. J. Hum. Genet. *106*, 215–233.

Wang, X., Cairns, M.J., and Yan, J. (2019). Super-enhancers in transcriptional regulation and genome organization. Nucleic Acids Res. *47*, 11481–11496.

Wirka, R.C., Wagh, D., Paik, D.T., Pjanic, M., Nguyen, T., Miller, C.L., Kundu, R., Nagao, M., Coller, J., Koyano, T.K., et al. (2019). Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. Nat. Med. *25*, 1280–1289.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952–1958.

Zimmerman, K.D., and Langefeld, C.D. (2021). Hierarchicell: an R-package for estimating power for tests of differential expression with single-cell data. BMC Genomics *22*, 319.

Decision Letter, appeal:

IMPORTANT: Please note the reference number: NG-A57752R-Z Miller. This number must be quoted whenever you communicate with us regarding this paper.

19th Nov 2021

Dear Clint,

Thank you for your message of 19th Nov 2021, asking us to reconsider our decision on your manuscript "Cell-specific chromatin landscape of human coronary artery resolves regulatory mechanisms of disease risk". I have now discussed the points of your letter with my colleagues, and we think that your revision has satisfactorily addressed the concerns raised at the past round of review. We therefore invite you to submit the revision of your manuscript.

It might be useful to also include more details (e.g. statistics) on what further analysis has been performed, in response to Reviewer #1's comments on statistical power, in your point-by-point response; currently this section is very qualitative and a more rigorously quantitative presentation may well be more persuasive.

When preparing a revision, please ensure that it fully complies with our editorial requirements for format and style; details can be found in the Guide to Authors on our website (http://www.nature.com/ng/).

Please be sure that your manuscript is accompanied by a separate letter detailing the changes you have made and your response to the points raised. At this stage we will need you to upload:
1) a copy of the manuscript in MS Word .docx format.
2) The Editorial Policy Checklist:
https://www.nature.com/documents/nr-editorial-policy-checklist.pdf
3) The Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf
(Here you can read about the role of the Reporting Summary in reproducible science:
https://www.nature.com/news/announcement-towards-greater-reproducibility-for-life-sciences-research-in-nature-1.22062)

Please use the link below to be taken directly to the site and view and revise your manuscript:

[REDACTED]

With kind wishes,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

Author Rebuttal, first revision:

Point-by-point response to Nature Genetics, 2021, Turner and Hu et al.

We first thank all of the reviewers for their helpful comments and constructive feedback and the opportunity to consider this work for publication in *Nature Genetics*. We have provided extensive new material and revisions to the text that we feel has markedly improved the quality of the manuscript. We have added power calculations and generated new bulk ATAC-seq caQTL results to support our primary results. Importantly, the caQTLs we highlight (for example rs13324341 (*MRAS*) in smooth muscle cells) were replicated in bulk coronary artery ATAC-seq libraries (50,000 nuclei per individual). We have also included histology characterization of these coronary artery samples in healthy and atherosclerotic contexts and more comprehensive immunostaining of PRDM16 in the vessel wall. Below we provide point-by-point responses (in blue) to each comment/remark from the three reviewers. At the end we also detail specific additions/changes we made to the manuscript. Specific changes to the enclosed manuscript text are also highlighted in blue for clarity.

Reviewer Comments:

Reviewer #1:
Remarks to the Author:
In this article, the authors have applied single-cell ATAC-seq (scATAC-seq) to profile 28,316 cells across coronary artery segments from 41 patients with varying stages of Coronary artery disease (CAD). They identified 14 distinct cellular clusters and mapped ~320,000 accessible sites across these cells. The authors also identified cell type-specific elements, few transcription factors, and attempted to do chromatin accessibility quantitative trait locus (caQTL) mapping. They also used cis-regulatory elements (CREs) to link genes to disease-associated transcription factors such as PRDM16 and TBX2. The authors claim to provide a single cell atlas for coronary artery and help interpreting cis-regulatory mechanisms in CAD.

A) It's technically a single-nucleus ATAC-seq as they cannot isolate cells from frozen samples to begin with. In the manuscript it should be relabeled as snATAC-seq

Thank you, we agree with this comment and we have now changed “scATAC-seq” to “snATAC-seq” throughout the manuscript to more accurately reflect our single nuclei isolations from the frozen samples.

B) The most critical aspect of the paper is the number of nuclei obtained after snATAC-seq - 28,316 cells. While superficially it may look like a decent number of nuclei for all the analysis, the fact that those cells are coming from 41 samples underscores the fact that the dataset is underpowered to dissect out inter-sample group-level differences and more importantly caQTLs. A closer look at the supplementary data (Supplementary Figure 3b-c) reveals that 50% of samples (about 19 samples) had less than 500 captured cells , with at least 8 samples less than 250 cells. More striking is the fact that only 8 samples (out of 41 samples -- 19%) had more than 1000 captured cells, and none over 1500 cells. The difference between different biological groups is even more striking with 87% of category 3 (atherosclerotic) samples having less than 500 cells while significantly more cells were obtained from category 1 samples (controls). As such the group differences become confounded with huge cell-type capture-rate differences and it is bound to affect data analysis downstream.

We appreciate Reviewer 1's comment in this regard. Here, we respond with a few points to clarify our rationale and support our results:

First, we actually conducted two steps of quality control (QC) for nuclei that were captured and sequenced in the snATAC-seq experiments. We captured and sequenced 94,432 total nuclei (ranging from 113 (minimum) and 20,050 (maximum) with an average of 2,146 nuclei/sample and 17,433 unique fragments/nuclei) using the default QC thresholds from the 10x Genomics Cell Ranger ATAC pipeline (**Supplementary Table 2**). We then subjected these nuclei to more stringent QC thresholds (transcription start site (TSS) enrichment ≥7 and ≥10,000 unique fragments per nuclei) (**Supplementary Table 2**). As expected, the high number of unique fragments per cell dictates how informative each individual cell is for downstream analyses and also contributes towards the power to detect *cis*-regulatory elements (Mandric et al., 2020). Although we did not obtain high nuclei numbers in some individuals, this is not unexpected based on the more stringent QC thresholds and the atherosclerotic disease nature of many of the samples profiled. By pooling all individuals together and obtaining 28,316 high quality nuclei, we were able to overcome the fact that we had fewer post-QC nuclei per individual. Importantly, cells belonging to the various clusters were evenly distributed across all of the individuals (**Rebuttal Figure 1**). Thus, when we are comparing clusters and cell types, the differences are unlikely to be driven by a select few individuals.

Second, we feel our study is adequately powered to dissect out differences between cell types and subtypes (for example, differentiated smooth muscle cells vs. fibromyocytes in **Figure 3**). Recently developed power calculators for single-cell RNA-seq analysis (e.g. SCOPIT v1.1.4 (Davis et al., 2019)) demonstrate that we are able to detect cell type

differences (Retrospective mode, assuming a cluster of at least 50 cells and set the required probability of sequencing the cells from each subpopulation to 0.95) for our rarest cell type (Mast cells at 0.7% frequency) by sequencing >9,129 cells (**Rebuttal Figure 2a and Supplementary Figure 16a**). Given that we focused our analysis on the main cell types in the vessel wall, we are well above this threshold to identify cell type differences in chromatin accessibility. To confirm that we had the appropriate sample size needed for SMC and fibromyocytes differential accessibility comparisons, we performed a similar canonical 2-sample t-test power analysis as described above (**Rebuttal Figure 2b and Supplementary Figure 16b**). Upon extracting ArchR differential accessibility-based gene scores using the same parameters as in the differential peak analysis (see above methods section), we calculated effect sizes for each gene using the gene score means and pooled standard deviations across cells from the SMC and fibromyocyte groups. To calculate the sample sizes required to reject the null hypothesis (no significant difference in accessibility-based gene scores between the two cell groups) at a given power value, we developed a custom python script using the statsmodels package. As before, we used stringent cutoffs for both the statistical power (0.99) and significance level (type 1 error rate, alpha = 0.01) to show the robustness of the differential accessibility-based analysis. As a result, we found that the required sample size for 93% of differential genes (based on accessibility-based gene scores) including fibromyocyte markers such as *FBLN2, LUM*, *F2R* and *TNFAIP6* was well below the number of cells available for the two groups (SMC n=6518, fibromyocytes n=2512) (**Supplementary Data 9**). We also calculated the mean required sample sizes across all differential genes for a wide range of power values (0.05-0.99) to further confirm that we were well-powered to detect significant accessibility differences between SMCs and fibromyocytes (**Rebuttal Figure 2b and Supplementary Figure 16b)**.

Reviewer 1 makes a valid point that we have fewer cells per sample in category 3 (advanced atherosclerosis) compared to category 1 (healthy controls). This may contribute to the limited number of differentially accessible genes/peaks identified between healthy and diseased individuals (**Rebuttal Figure 3**). However this may also be driven by the continuous rather than dichotomous nature of atherosclerotic plaques, which have a high degree of heterogeneity and plasticity of cell types (van Kuijk et al., 2019). In fact, we have observed consistent results in group level differences in bulk ATAC-seq profiled samples from this set of patients (unpublished observations). Nonetheless, we acknowledge this as a known limitation in the Discussion and mention the benefit of performing more single nuclei profiling of advanced disease (category 3) samples to better address this question in future studies. Given the difficulty in procuring these precious samples from heart transplants (especially during the pandemic), we have been unable to obtain more unique individual samples for this particular study.

Third, we believe we suitably capture differences between individuals with respect to the caQTL analysis. To begin to estimate our power for caQTL analyses, we used the powerEQTL package for single-cell RNA-seq based eQTL studies, which assumes a standard linear regression model (**Rebuttal Figure 2c and Supplementary Figure 16c**) . Assuming an average of 670 SMC captured per individual across 41 individuals, we have 90% power to detect SNP associations for common variants (MAF~5%) with moderate effect size (beta = 0.3). However, these calculations grossly underestimate the improved power (>1.6 fold) derived from an allele-specific based method such as RASQUAL (Kumasaka et al., 2016), which is designed to detect molecular trait associations from small sample sizes. In fact, we previously demonstrated a 7-fold increase in significant eQTLs discovered with this approach compared to FastQTL in a cohort of 52 human coronary artery SMC (Liu et al., 2018). We used RASQUAL for our caQTL mapping and also limited these analyses to the two major cell types present in each of the samples (SMC and macrophages). Using more stringent inclusion criteria for the individual samples, we re-ran RASQUAL for these cell types, and discovered 1,984 and 1,210 cis-caQTLs in SMC and macrophages, respectively, at 5% FDR. This is similar to our original results and we have now updated **Figure 5** and **Supplementary Data 6**.

To strengthen our single cell caQTL findings, we performed bulk ATAC-seq on the remaining frozen coronary arteries for all 41 of these matched patients. 35 of these bulk ATAC libraries were of sufficient quality and used for downstream bulk caQTL analysis. We added methods on the bulk ATAC-seq to the manuscript and added results in **Supplementary Data 7**. Since we did not need to perform single cell capture for these bulk reactions, each bulk ATAC-seq library contained ~50,000 nuclei per individual. Our results show many of our smooth muscle and macrophage single cell caQTLs are also detected as significant bulk caQTLs, with the direction of effects primarily consistent (Pearson r=0.814) (**Rebuttal Figure 4** and **Supplementary Figure 11**). Most importantly, the single cell caQTLs that we highlight in the manuscript (such as *MRAS*, *MEF2D*, *FCHO1*, *SMAD3*) were also significant bulk caQTLs with consistent direction of effects. For instance, for rs13324341 at the *MRAS* locus, the peak containing this variant is accessible in smooth muscle cells. The T allele associates with greater accessibility in both our smooth muscle caQTL dataset as well as in the bulk coronary caQTL dataset. Furthermore, as we show in **Figure 5b** and **Figure 5c**, many of these caQTLs are significant eQTLs (5% FDR) in GTEx arterial tissues. We acknowledge in the Discussion that certain caQTLs are likely cell-type/state specific whereas other caQTLs are likely to be shared across several cell types and tissues. Also, we acknowledge that larger cohorts are likely needed to fully address this question.

Lastly, to compensate for the reduced power to detect cell type caQTLs (especially for less abundant cell types), we performed machine learning based prediction of CAD regulatory variant function using three different methods (**Methods**). Using the gapped k-mer support vector machine (gkm-SVM), gkm-Explain, and deltaSVM trained on accessible peaks from SMC, macrophages, fibroblasts, endothelial cells, pericytes, T/NK or Mast cells we identified single nucleotide variants predicted to alter chromatin accessibility and potentially transcription factor binding from the reference sequence (**Supplementary Data 8**). We identified a number of top candidate regulatory variants at CAD loci (e.g. *LIPA* and *SMAD3*) that are predicted to function through cell-specific regulatory mechanisms. This powerful and complementary approach could be extended to deep learning based models to enable more exhaustive prediction of variants that may be difficult to prioritize from QTL based methods alone.

It is worth noting that despite following the Omni-ATAC protocol (Corces et al., 2017) that is optimized for frozen tissue and careful optimization, we were unable to capture high numbers of nuclei from these advanced coronary artery plaques. These advanced, category 3 samples had high amounts of extracellular matrix, calcification, and necrosis, and our group in addition to others have had difficulty capturing high numbers of cells/nuclei per sample from fibrous/calcified human tissues, especially atherosclerotic plaques. For example, a recent study from Depuydt et al. in *Circulation Research* (Depuydt et al., 2020) performed single cell RNA-seq from 18 patients, all of which were human carotid artery atherosclerotic plaques. Across all 18 patients, this study obtained a total of 3,282 cells, which is approximately 234 cells per individual. This study also performed snATAC-seq for 6 individuals (fibro-atheromatous carotid artery plaques) but only obtained a few hundred suitable nuclei per individual. As is the case for many other human atherosclerosis single-cell datasets, only 3-4 individual samples are typically available (refer to our data portal at PlaqView.com), further demonstrating the difficulty in obtaining and analyzing these samples.

Overall, we acknowledge the reviewer's important comments. However, given that this is the first study to perform snATAC-seq of both healthy and diseased human coronary arteries, which are extremely difficult to procure, we feel that 28,316 analyzed high-quality nuclei, post-QC still represents a very valuable resource for the community. *More importantly, based on the detailed explanations above, we feel that analyzing more nuclei per individual would not affect the major conclusions of this manuscript.*

C) The inherent noise in snATAC-seq data, compounded by the fact that the authors only captured less than 750 cells from majority of samples which are now sub-clustered into 14

clusters suggest that even cluster differences for small-to -medium sized clusters is grossly underpowered. Case-in-point is subclustering of smooth muscle clusters 6 and 7 which are less than 2000 cells each. While the authors do not provide how many cells from each sample contribute to clusters 6 and 7, the average number of cells for cluster 6 and 7 from each sample will be less than 50 cells/sample (2000cell/41samples = 48 cells/sample). With that low resolution of cells, I will be hard-pressed to find any meaningful analysis for these clusters.

The situation is not that different in the 2 largest clusters (Smooth muscle cells, cluster 4 and 5) with about 6000 cells in each cluster and average contribution from each samples will be 150 cells/sample

We acknowledge the reviewer's concern. Overall, we believe our snATAC data can identify biologically meaningful differences between main clusters and in the case of clusters 4-7 (smooth muscle cells), differences between cell sub-clusters (as shown in **Figure 3**). To further clarify this first point, we now provide individual UMAPs separated by each sample to visualize the number of cells contributing to each main cluster and sub-cluster (**Rebuttal Figure 1**). We also include the numbers of cells analyzed per cluster per sample in **Supplementary Table 3**.

Secondly, to address the reviewer's concern for power (also refer to **Rebuttal Figure 2** and **Supplementary Figure 16**), we performed another power calculation of detecting differentially expressed genes between main clusters and sub-clusters using the recently developed Hierarchicell R package (Zimmerman and Langefeld, 2021). As noted by the authors of this package, in order to identify at least 1.2 fold changes as statistically significant (power > 0.80), "researchers will need a minimum of 40 samples and 100 cells per sample." This is assuming a traditional case/control study design. Given that atherosclerosis is a continuous trait, we already achieve maximal power with 100 cells/sample. There are actually only minor differences in power when sampling 100, 250, 500, or 1,000 cells per sample. In fact, the authors demonstrate that as the overall sample size increases (e.g. 20 to 100) there is a drop in power gains by sampling more cells/sample. This is consistent with the per gene sample size power calculation for our differential analysis of the integrated data, as described above (**Rebuttal Figure 2**, **Supplementary Figure 16** and **Supplementary Data 9**).

We did recognize the intrinsic noise in snATAC when performing these sub-cluster differential analyses as shown in **Figure 3**. Instead of comparing cluster 6 vs. cluster 7 from the snATAC data alone, we actually compared cells labelled as 'smooth muscle cells' (SMCs) against cells labelled as 'fibromyocytes' upon integrating human coronary artery (HCA) scRNA-seq data from Wirka et al., *Nature Medicine* 2019 (Wirka et al.,

2019). Integration of scRNA-seq data into snATAC-data has been shown to enhance clustering resolution and enable identification of cell subpopulations not possible with snATAC-seq data alone. Using the integrated data, we were able to annotate 6,518 contractile SMCs and 2,512 fibromyocytes for the downstream differential accessibility analysis. We believe the integration of the snATAC-seq with coronary artery scRNA-seq data strengthens the cell type annotations and overcomes the inherent noise in the gene activity scores derived from snATAC accessibility data. As mentioned above, each individual in the study contains cells belonging to each cluster with nearly equivalent cell compositions across individuals. Further, the lower numbers of captured cells in certain individuals is further mitigated since we pooled all cells together for the smooth muscle cell vs. fibromyocyte analyses. We think this is a valid approach since the main goal of this differential analysis was to compare contractile and modulated SMCs as opposed to disease vs non-disease status. Thus, it remains unlikely that specific individuals are driving the observed differences between clusters.

Nonetheless, we agree that increased sample sizes and/or nuclei/sample may still be necessary to perform sub-cluster comparisons between smaller clusters representing less abundant cell types (e.g. T cells) and more context-specific mechanisms. Since we did not focus on these less abundant cells in this manuscript, this would be an opportunity for follow-up studies. We have added this note in the revised Discussion.

D) Can the authors integrate their data with their own snRNAseq data, now in bioRxiv (https://doi.org/10.1101/2020.10.27.357715) ? More specifically, do the CREs line up caQTLs and eQTLS from the snRNAseq data? What is the overlap of open chromatin accessibility at the promoter of the first exon of a gene (snATAC-seq) compared to its expression (from snRNA-seq)?

We apologize for any confusion. The coronary artery scRNA-seq data analyzed in the bioRxiv paper by Ma et al. was from GSE131778 (Wirka et al., Nature Medicine 2019), which is the same published scRNA-seq dataset used in this manuscript for integration with snATAC-seq. We are co-authors on this original paper and it is currently the only publicly available human coronary artery scRNA-seq/snRNA-seq dataset. We have highlighted the agreement between our snATACseq and scRNAseq from the integration analysis with geneScore (snATACseq) vs. expression (scRNAseq) in **Figure 1D**.

Regarding the alignment of caQTLs and eQTLs through similar CREs, this question is not yet possible as single-cell eQTLs have not yet been mapped. This will be an interesting question to pursue in future multimodal (snATAC/snRNA) analyses of these tissues. We did perform pseudo-bulk level comparisons between snATACseq and

scRNAseq by comparing ATAC-seq peak occupancy at the promoters (e.g., 3 kb of TSS) of expressed genes (detected by scRNAseq pseudo-bulk level, e.g., RPKM >=1) for the same cell type. The results of these analyses are now reported in **Supplementary Figure 6c**, with an overall good correlation (Pearson r~0.55). The "gene score" model implemented in ArchR (Granja et al., 2021) that accounts for enhancer activities provides higher performance for estimation of gene expression.

E) The authors should read this paper (https://doi.org/10.1038/s41467-020-19365-w) on effective design of single-cell sequencing experiments for cell-type-specific QTL analysis, applicable for both eQTLs and caQTLs. The number of cells captured is far more critical than the number of samples used or read-depth of each cell.

We thank Reviewer 1 for sharing this paper (Mandric et al., 2020) and agree that this paper provides very good advice for planning experiments for cost. We agree that a high number of cells captured is important for identification of cell type QTLs, but believe this paper also emphasizes that a high number of individuals is very important. Both of these factors seem to be more important than the read depth per cell. Thus, we feel that the number of nuclei we captured is largely offset by increasing the number of individual samples, as described in the above responses. We have now expanded our Discussion section to emphasize the need for both increased numbers of captured cells and/or numbers of samples for improved cell type QTL discovery and cited the referenced paper.

For the snATAC experiments we aimed for both a high number of individuals and a high number of nuclei/sample. We had sufficient nuclei concentrations for most samples and aimed to capture ~5,000 nuclei/per sample. However, our capture rate from the 10x Chromium Controller was lower than we anticipated. As we previously mentioned we obtained 94,432 total nuclei from Cell Ranger ATAC but filtered more stringently for downstream analyses.

One key difference between our study and this paper is that we used RASQUAL (Kumasaka et al., 2016) to detect cell-type QTLs whereas this paper used Matrix eQTL. Matrix eQTL (Shabalin, 2012) uses linear regression to compare read counts between individuals across genotypes. Typical power calculators for single cell QTL studies (e.g. powerEQTL described above) rely on similar linear models. However, RASQUAL maximizes association detection by combining both between-individual and within-individual allelic effects. Thus RASQUAL is much better equipped to discover QTLs in smaller sample sizes compared to Matrix eQTL. Although the samples analyzed in the

RASQUAL paper were from bulk ATAC libraries, RASQUAL was able to detect thousands of caQTLs from only 24 individuals.

Reviewer #2:

Remarks to the Author:

This manuscript describes a tour de force employing single-cell ATAC-seq to 28,316 cells in 3 epicardial coronary artery segments from patients with and without coronary atherosclerosis to identify cis-regulatory mechanisms in VSMC transition states. They further extend the findings of recent GWAS using an integrative approach to identify plausible molecular mechanisms whereby a risk locus confers an effect in a cell specific manner. Further they identify PRDM16 and TBX2 as transcriptional regulators in SMCs.

The unique value of this work stands as the first single-cell atlas of human coronary artery chromatin accessibility that is a unique resource of value to the scientific community.

Comments.

Figure 1 a. Are the reasons for transplant rejection of donor hearts known? Were calcified samples excluded? Of note, SMC can also differentiate towards osteoblastic cells, again a unique SMC phenotype.

We should clarify that these are not “rejected” hearts in the typical nomenclature (e.g. rejected organs after transplantation). These “rejected” donor hearts are actually candidate donor hearts that were turned down by the transplant surgical team prior to being transplanted. We have now modified the terminology in the text to prevent any confusion. Regardless, while we have detailed information on the individual donor hearts from the United Network for Organ Sharing (UNOS), we do not usually have a specific reason for “rejection” beyond the surgical teams’ discretion. Some common reasons include: donor-recipient incompatibility such as size mismatches between donor and recipient, cerebrovascular incidents, and drug use.

We included calcified coronary artery samples (all in category 3 (**Figure 1**)), yet these samples provided poorer capture rates and lower number of nuclei that passed our quality control thresholds. While the Omni-ATAC protocol generally worked well for frozen coronary samples, nuclei isolation was substantially more difficult for the highly calcified samples. The highly calcified/advanced plaque samples were difficult to break into small pieces and resulted in more debris in the nuclear preparations. As a result, we did not capture as many *bona fide* osteoblastic cells that are present in advanced atherosclerotic lesions. We suspect that this snATAC dataset does include osteochondrogenic cells that are at earlier stages of SMC transdifferentiation towards this cell type. This is based on preliminary integrative analyses with lineage-traced

mouse scRNA-seq data, where we identified integrated expression of the chondrocyte *Sox9* marker in SMC-derived cells. Nonetheless, this would be an interesting follow-up study to identify other regulators of osteoblast differentiation by integrating mouse SMC lineage traced scRNA-seq datasets.

Figure 1 e. Although 3 arterial segments were harvested from 41 individuals, data are shown for 44 samples. Is this because of less than optimal sample quality or processing? Might the unknown cells represent adipocytes?
It is of interest that the atherosclerosis samples with or without adventitia are generally similar in cell type.

We have 44 samples from 41 individuals due to less than optimal sample processing from some extremely diseased/highly calcified samples for 3 individuals on the first attempt. For these samples we obtained a low number of nuclei that passed the ArchR QC thresholds, but still kept the high-quality nuclei for downstream analyses. We then repeated snATAC-seq library preparation and analyses for adjacent segments from these same 3 individuals.

This is an interesting question regarding the potential assignment of the unknown cluster as adipocytes. While perivascular adipocytes and adipocyte signaling play important roles in inflammation and atherosclerosis, we do not think this unknown cluster (cluster 9) represents adipocytes. Prior to sample freezing we carefully dissected the perivascular fat from the outer layer of the coronary arteries. From our experience, other studies do not trim the fat before profiling arterial tissue, which was done here to enrich vascular wall cell types. When we looked up snATAC gene scores for traditional adipocyte marker genes (e.g. *UCP1*, *CITED1*, and *ZIC1*) there was no noticeable enrichment for these genes compared to other cell clusters (**Rebuttal Figure 5**). Lastly, based on the snATAC marker gene scores, snRNA-seq integrated expression and the confusion matrix shown in **Figure 1d**, the 'unknown' cells closely resemble the 'Macrophage' and 'Mast' labels from scRNA-seq. Top marker genes for this cluster include *CD163* (reported to be an M2 macrophage marker) and *CD300LB* (expressed on myeloid cells). Nonetheless, we acknowledge that there is considerable heterogeneity in myeloid/immune cells in atherosclerosis (Lin et al., 2019, Fernandez et al., 2019).

Figure 3. These are very nice data that define at a single cell level, marker genes, TF enrichment and differential promoter peaks that extend our understanding of differentiated vs fibromyocytic SMCs. Apropos of sample calcification, in Fig 3f – Does RUNX include RUNX2, a marker of osteogenic differentiation?

Thank you, this is an interesting question. For the position weight matrices displayed in **Figure 3f**, the RUNX motif includes RUNX2 (smooth muscle cell calcification marker (Lin et al., 2015, 2016) in addition to the very similar RUNX1 and RUNX3 motifs. We have now added a sentence in the main text in this regard. This would potentially suggest the fibromyocyte SMCs also encompass some osteochondrogenic cells and agree with recent single cell studies highlighting that SMCs can transition to osteogenic or chondrocytic phenotypes. This also agrees with the identification of *TNFRSF11B* (Osteoprotegerin) and *POSTN* (periostin) as top marker genes in fibromyocytes.

Figure 4. It is of interest and perhaps not unexpected that the majority of recently reported CAD GWAS variants demonstrate functionality in vascular SMCs. 4d. Given the ongoing interest in immune function and CAD, the authors may wish to highlight the variants with strong macrophage peaks.

We agree that macrophages and other immune cells play key roles in atherosclerosis and these cell types have been intensely studied in the CAD field. After smooth muscle cells, macrophages represented the second most abundant cell type in our snATAC dataset. We have now highlighted additional CAD associated variants residing in strong macrophage peaks (e.g. rs7296737 at *SCARB1* and rs17680741 at *TSPAN14*) in the main text. We also highlighted a top regulatory variant in *LIPA* in **Figure 5,** which is predicted to alter macrophage-specific TF binding sites, as identified through our machine learning analysis.

Figure 6. Relevant to the PRDM16 findings, beige (not brown) adipocytes are known to express SMC markers and can be derived from VSMCs under the direction of PRDM16. If present they could be identified by UCP1 expression. Here the authors should review the work of Spiegelman (2014).

We thank Reviewer 2 for sharing this important paper from Spiegelman's lab (Long et al., 2014). As mentioned, ectopic expression of PRDM16 *in vitro* can convert VSMCs to beige adipocytes (UCP1 being a thermogenic beige adipocyte marker). Additionally, this study emphasizes how VSMC-like cells display similarities to beige adipocytes rather than traditional brown adipocytes. We did attempt to correlate PRDM16 and UCP1 based on imputed gene scores in SMCs, however these were modestly negatively correlated (Pearson r = -0.28).

Reviewer #3:
Remarks to the Author:
This manuscript by Turner et al utilizes single cell ATAC-seq (scATAC-seq) data from coronary arteries from 41 patients to characterize cis-regulatory regions that are linked to CAD loci. This analysis reveals cell-specific accessible chromatin and potential binding sites at CAD risk loci and elucidates potential mechanisms for smooth muscle cell phenotypic. scATAC-seq studies were published recently on human carotid endarterectomy samples by Depuydt et al (Circ Res, 2020) and Ord et al (Circ Res, 2021), which removes some novelty of the study by Turner. However, the Ord study had data from only 3 patients (all with advanced disease) and ~7000 cells total. The current manuscript reports data on human right coronary arteries, LAD arteries and left circumflex arteries from 41 patients with various stages of disease and >28,000 individual cells. Because of the inclusion of multiple patients, Turner et al were able to perform chromatin accessibility QTL analyses, which greatly
adds to the study. This is also the first scATAC-seq dataset from coronary arteries. Overall, this is an excellent study and will provide an important resource for the community. The data and analysis are of high quality. There are several questions that remain.

Major comments:

1) The introduction section does not mention the other scATAC-seq datasets on human carotid artery atherosclerotic plaque, but instead refers to studies on cultured cells. The Ord study does not appear to be cited. The omission of these references in the introduction oversells the novelty somewhat.

We fully agree with this comment and have now mentioned and cited the new Ord et al. study in *Circulation Research* (Örd et al., 2021). It came online a few days before we initially submitted. We also cited the original scATAC-seq study on this carotid artery dataset by Depuydt et al (Depuydt et al., 2020).

2) It isn't clear whether the staging of atherosclerosis severity is according to the Stary criteria with four stages (Circulation, 1995). This is standard for the field and would provide more detailed information about patient plaque characteristics. Was pathology performed on plaque samples? While scATAC-seq was performed on various stages of disease, it seems that all of the data is combined rather than analyzed according to disease stage. Is the accessibility of CAD loci altered by disease stage?

Thank you, we agree that the Stary criteria is the standard in the field to grade the severity of atherosclerosis. We have performed histology analyses on most of these coronary artery samples, which include healthy samples with minimal intimal thickening, early/intermediate atheromas as well as fibro-fatty plaques with calcification. Importantly, whenever possible, we used samples for snATAC-seq that were adjacent to regions of the coronary artery that were used for histology analysis. Representative histology images and quantitation of a subset of samples per disease stage is now included in **Supplementary Figure 4 and Rebuttal Figure 6**. Using Oil Red O (ORO) staining of lipids we observed an accumulation of lipid laden cells in the subintimal layer of the early atheroma (category 2) and fibroatheroma plaques (category 3). Sirius red and H&E staining also demonstrates increased intimal hyperplasia, collagen type I/III accumulation and decreased lumen diameter of the early atheroma and advanced lesion segments relative to the healthy control segments (category 1). In general, our results are consistent with the Stary (Stary et al., 1995) classification stages in that category 1 represents type I/II lesions with adaptive intimal thickening and initial fatty streak/foam activation, category 2 represents type III/IV lesions (intermediate/advanced atheroma) with more intimal thickening and accumulation of lipid, and category 3 represents type V/VI lesions (advanced fibrocalcific atheroma) with evidence of a lipid core, fibrous cap, and/or calcification.

Reviewer 3 is correct that most of our analyses use data combined for 28,316 nuclei across all individuals. In terms of comparing accessibility at CAD loci according to disease stage, we did not observe many genome-wide significant differences in chromatin accessibility between disease stages. This is expected given the continuous rather dichotomous nature of atherosclerosis progression. The loci we did observe make sense biologically in terms of the implicated genes, with marker genes having higher accessibility in Category 3 linked to inflammation and immune processes (e.g. *CD5*,

CD84, *CCL4L2,* and *ICAM1*) (**Rebuttal Figure 3**). In contrast SMC marker genes related to contractile function (e.g. *CNN1*, *KCNA5*) were more accessible in Category 1 samples. Unfortunately due to difficulty obtaining high numbers of nuclei from more advanced diseased/plaque samples (Category 3) these samples have fewer analyzed nuclei compared to the other categories, as noted in response to Reviewer 1. Finally, dissecting differences between disease stages remains a challenging task since coronary lesions are very heterogeneous and this type of analysis may be more suitable for spatial genomic approaches to compare omic profiles *in situ*.

3) Are the gene score differences between cell clusters in Fig. 3B statistically significant?

This is a good point, which we have now clarified in a revised **Figure 3b**. The differences in gene scores comparing Cluster 4 + Cluster 7 vs. Cluster 5 + Cluster 6 are statistically significant and now reflected with p values shown in the **Figure 3b** panel.

4) Would analysis of 'super enhancers' or clustered ATAC-seq peaks provide any additional information regarding potential functional regulatory regions?

We agree that this is a very interesting question. In cultured human coronary artery smooth muscle cells ((Miller et al., 2016), GSE72696), H3K27ac ChIP-seq marks (established feature of super enhancers) have previously been used to identify SMC super enhancers. We re-analyzed these datasets using the SICER (Zang et al., 2009) package, which is optimized for broad peak calling from histone modifications and identified super enhancers from H3K27ac peaks >10 kb, as previously described (Wang et al., 2019). These cultured human coronary artery SMC super enhancers were enriched at SMC marker genes in our snATAC-seq data compared to all other cell types (**Supplementary Figure 7**).

Here, we provide overlaps of accessible chromatin regions in each cell type with long stretches of H3K27ac marks (**Supplementary Table 5**). We find that ATAC-seq peak clusters (ATAC-seq peaks longer than 10 kb) in smooth muscle cells showed the highest association with super enhancers. We further observed that the SMC super enhancers showed significantly higher regulatory potential for the identified SMC marker genes compared to the marker genes from all other cell types (**Supplementary Figure 7**), supporting the additional functional insights gained from this analysis. We highlight an example at the *LMOD1* locus (SMC marker and CAD gene), which harbors two SMC superhancers (**Supplementary Figure 7**). Enhancers at this gene have already been validated experimentally, however other top candidate super enhancers deserve validation using in vitro models.

5) Further functional information on PRDM16 and TBX2 in atherosclerosis would be helpful. For the immunofluorescence in Fig. 6E, it would be helpful to include healthy tissue and advanced lesions for comparison to better understand the expression of PRDM16 in atherosclerosis. The methods indicate that healthy and sub-clinical atherosclerosis samples were used, but the healthy controls are not included in the manuscript. Is there eQTL or caQTL data for PRDM16 or TBX2? The data presented for PRDM16 and TBX2 are not entirely convincing and appear to be preliminary.

We agree with Reviewer 3 in that more functional information would help support our novel findings prioritizing PRDM16 and TBX2 at their respective CAD loci. For the revision we focused primarily on PRDM16 but also included some additional information for TBX2. For PRDM16 we conducted comprehensive immunofluorescence staining in both healthy (n=4) and diseased (n=8) coronary arteries along with whole slide confocal scanning and quantification (**Figure 6**, **Supplementary Figure 15 and Rebuttal Figure 7**). We leveraged the CVPath biorepository of atherosclerotic samples to carry out these more comprehensive analyses in lesions at well-defined disease stages. We observed PRDM16 staining colocalized in ACTA2 positive smooth muscle cells in the medial layer of healthy and early atheroma samples, which was reduced in the advanced fibroatheroma samples (both thin cap fibroatheroma and thick cap fibroatheroma). Interestingly, we observed high PRDM16 staining in the vasa vasorum, marking arterioles that are positive for ACTA2. We also observed staining in a few weakly ACTA2 positive arterioles, which could represent pericytes and/or endothelial cells (EC). This is consistent with a recent murine study demonstrating a role for both EC and SMC expressed Prdm16 in regulating flow recovery in post-ischemia PAD models (Craps et al., 2021). While we identified PRDM16 as a SMC-specific marker based on our snATAC-seq data, further studies are needed to dissect the functional interplay in EC and SMCs.

In addition to immunofluorescence, we leveraged data from our lab's bulk RNA-seq data, publicly available arterial scRNA-seq data, and performed additional *in vitro* experiments. While neither of these genes were differentially expressed in bulk tissues from our coronary artery dataset (n=150), PRDM16 and TBX2 were both significantly upregulated in perivascular adipose tissue from diseased coronary arteries (n=44) (Numaguchi et al., 2019) (**Supplementary Table 6 and 7**). By querying human and mouse atherosclerosis datasets we confirmed the SMC and pericyte (and limited EC) expression for PRDM16 and TBX2. (**Supplementary Figure 14**). To gain more functional insight into the mapped regulatory elements for PRDM16 we cloned 4 candidate PRDM16 enhancer sequences overlapping CAD SNPs upstream of a minimal

promoter and compared these to two different promoter sequences in standard luciferase reporter assays (**Rebuttal Figure 8**). Candidate enhancer #3 had ~5 fold increase in normalized luciferase activity compared to the empty vector control, consistent with the activity of the PRDM16 promoter. These results in HEK 293T were consistent with assays in immortalized human coronary artery smooth muscle cells, despite the lower magnitude of activation. We observed that PRDM16 expression is greatly reduced in cultured SMCs compared to intact tissue, making systematic validation challenging. Due to the dozens of candidate PRDM16 SNPs highly associated with CAD, overall locus complexity, and difficulty transfecting SMCs, we feel this would be better suited for future in-depth investigations.

In terms of eQTLs for *PRDM16* and *TBX2* in CAD-relevant tissues, we looked up signals at these CAD loci in the GTEx (GTEx Consortium et al., 2017), STARNET (Franzén et al., 2016) and Cardiogenic QTLizer (Munz et al., 2020) databases (**Supplementary Data 10**). We observed nominally significant eQTLs in STARNET and GTEx artery tissues for both genes, however there were modest numbers of eQTLs at both the *PRDM16* and *BCAS3/TBX2* loci in all of the studies. This is not completely surprising as many biologically important genes at GWAS loci often do not harbor many significant eQTLs due to constraint by negative selection (O'Connor et al., 2019; Wang and Goldstein, 2020). Indeed, both PRDM16 and TBX2 have high loss-of-function intolerance probabilities (pLI = 0.999 and 0.964, respectively) supporting their constraint and haploinsufficiency.

With respect to caQTLs, we observed a CAD-associated signal at the *PRDM16* locus (rs10797377, *ACTRT2*, chr1:3012242-3012643) that associates with peak accessibility in SMCs (RASQUAL q value = 1.86×10^{-4}). Moreover, this regulatory element has a highly significant Peak2Gene link with PRDM16 and PRDM16-DT (**Supplementary Data 5**) in which chromatin accessibility strongly correlates with gene expression. Notably, using the activity-by-contact (ABC) based enhancer-promoter linking method (Nasser et al., 2021) in ENCODE human coronary artery, we also identified *PRDM16* and *TBX2* as the target genes from CAD associated SNPs (**Supplementary Table 9**). While there are numerous linked CAD-associated candidate SNPs at these two loci, we believe the lack of additional caQTLs could reflect context-specific effects or buffering effects of allelic variation within enhancers at these loci. In the case of PRDM16, another independent association signal, rs2493292, is a missense variant, which is predicted to be tolerated/benign (SIFT, PolyPhen, MVP).

Taken together, along with our gene regulatory analyses identifying these two genes as key driver genes in STARNET, these additional functional data demonstrate that

PRDM16 and TBX2 are indeed the target genes for their respective CAD loci and may play critical roles in SMC during disease.

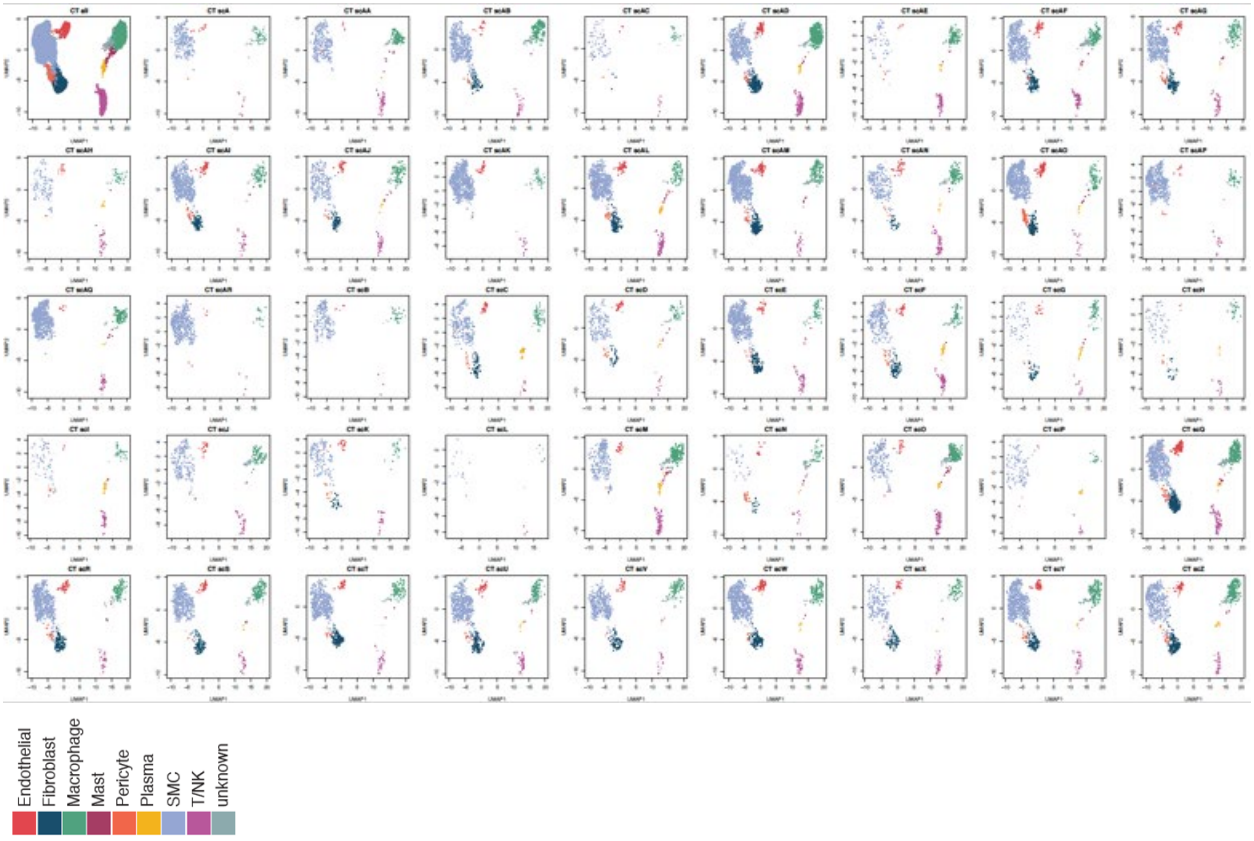
Minor comments:

1) The heatmap in Fig. 3D could use additional annotation to indicate where the trajectory begins and ends.

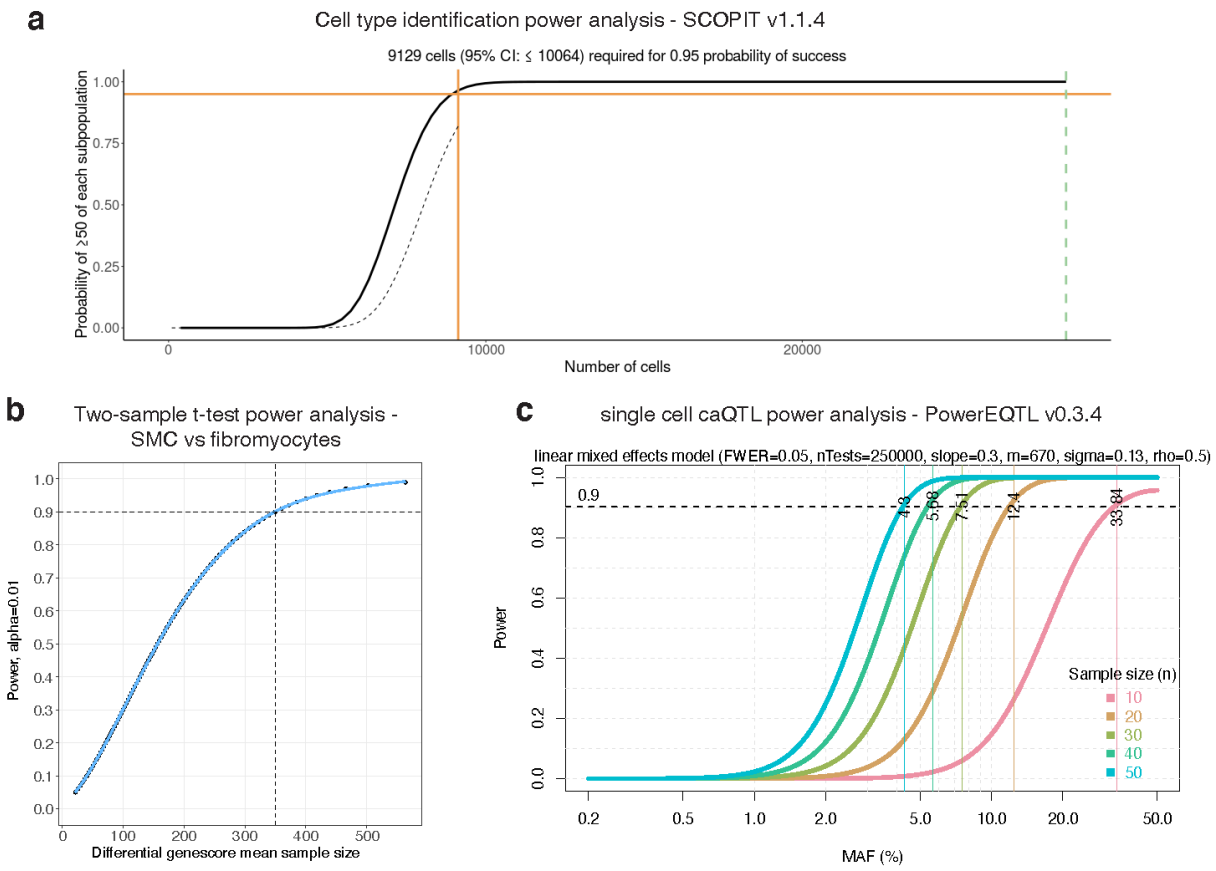
We agree that providing additional annotation to the heatmap improves interpretation of this figure. We have now added labels with arrows to the heatmap to indicate both where the trajectory starts and ends.

2) The legend for the heatmap in Fig. 4D is confusing. Are 'zero' peaks white? It seems that 'zero' peaks are black on the legend, but this doesn't seem to be correct.

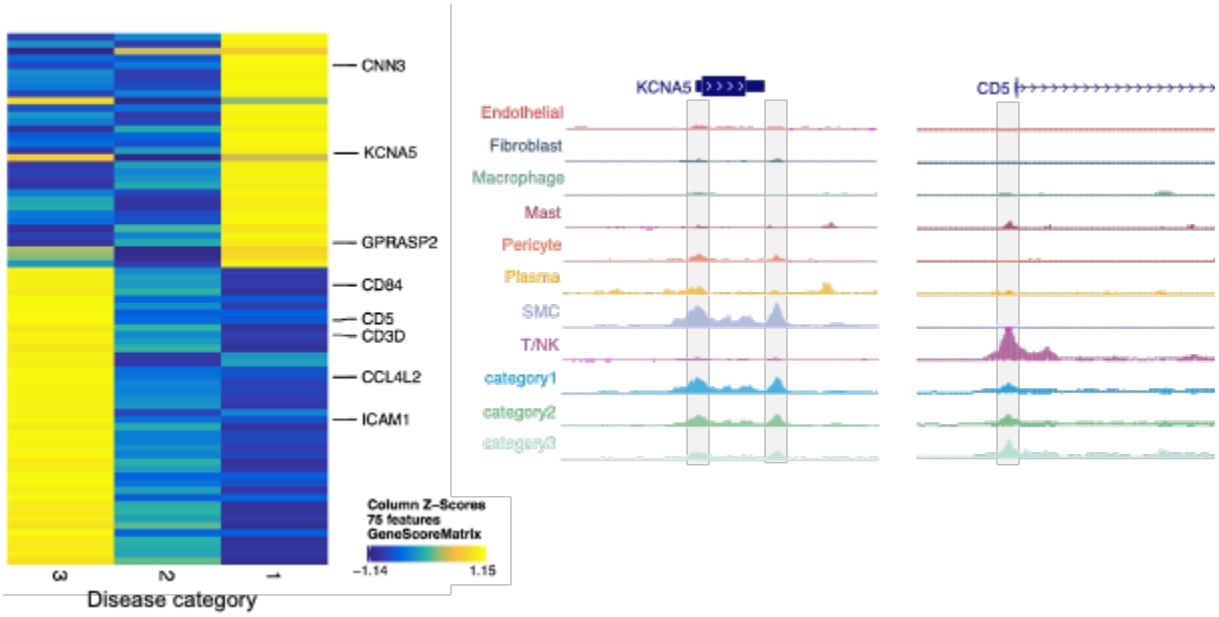
We thank Reviewer 3 for pointing this out. Zero overlaps should be white and the scale should start at 1. We have corrected this in **Figure 4d**.



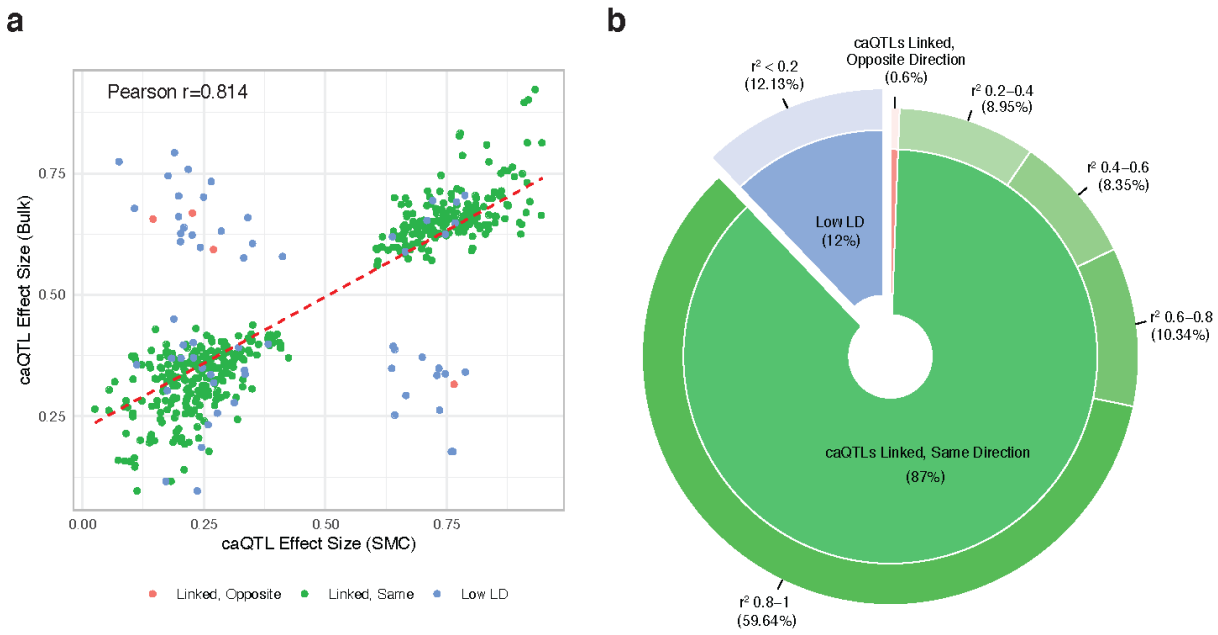
Rebuttal Figure 1. UMAP plots separated by each donor sample showing cells for each main cell cluster are represented from individual snATAC-seq libraries. Note: samples lacking an adventitial layer do not have cells in the fibroblast cluster (dark blue). Also, sample scL was excluded from caQTL mapping due to overall low cell numbers.



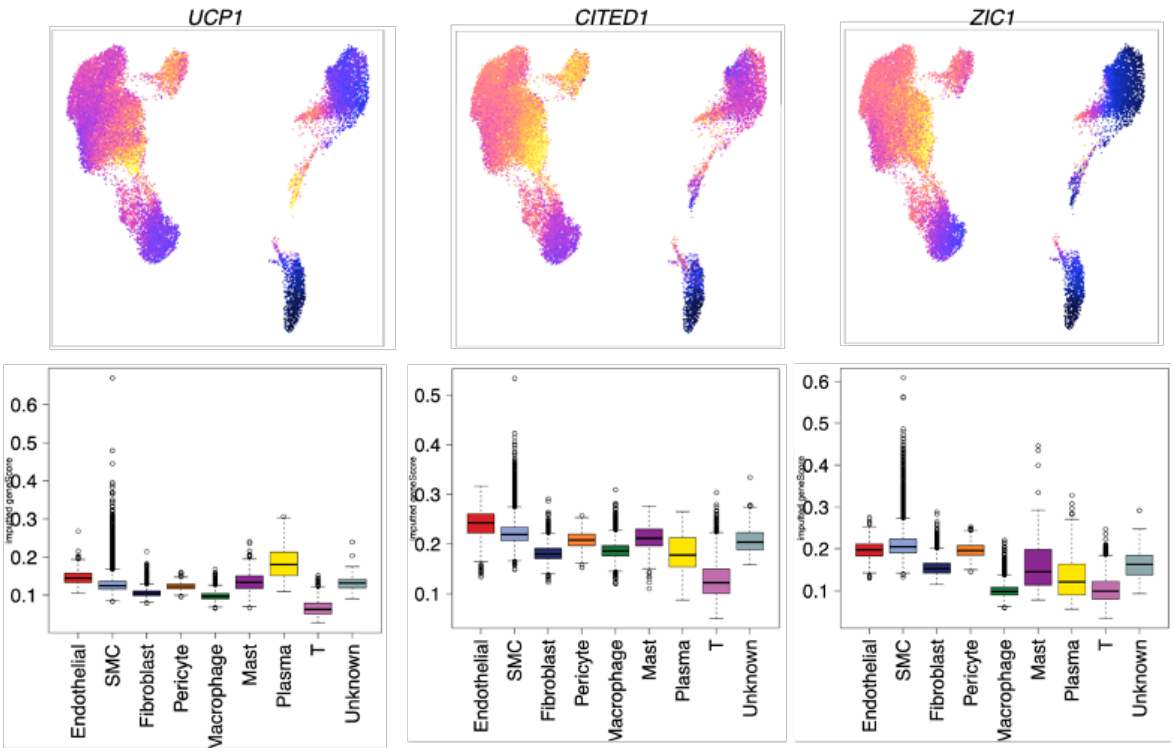
Rebuttal Figure 2. Power analysis results for cell type identification, differential accessibility, and caQTL. (a) SCOPIT v1.1.4 based power curve showing the probability of detecting at least 50 cells of the rarest cell type in crororary artery based on our snATAC data (Mast cell at a frequency of 0.07), assuming alpha =0.05. 99% probability of detection is achieved from a minimum of 9,129 total cells, as shown by the intersecting orange lines. (b) Two-sample t-test based power curve, showing the probability of detecting differential gene scores in smooth muscle cells (SMC) versus modulated SMC (fibromyocytes) as a function of mean sample sizes (calculated from the required sample sizes for all differential genes; see Methods) (c) Single-cell caQTL power analysis based on PowerEQTL v0.3.4, assuming a standard linear mixed effects model, alpha=0.05, nTests = 250,000 overlapping SNPs, minimum effect size=0.3, # cells = 670 (mean SMC/sample). A minimum of 40 samples is required to achieve 90% power to detect caQTL variants at ~5% minor allele frequency (MAF).



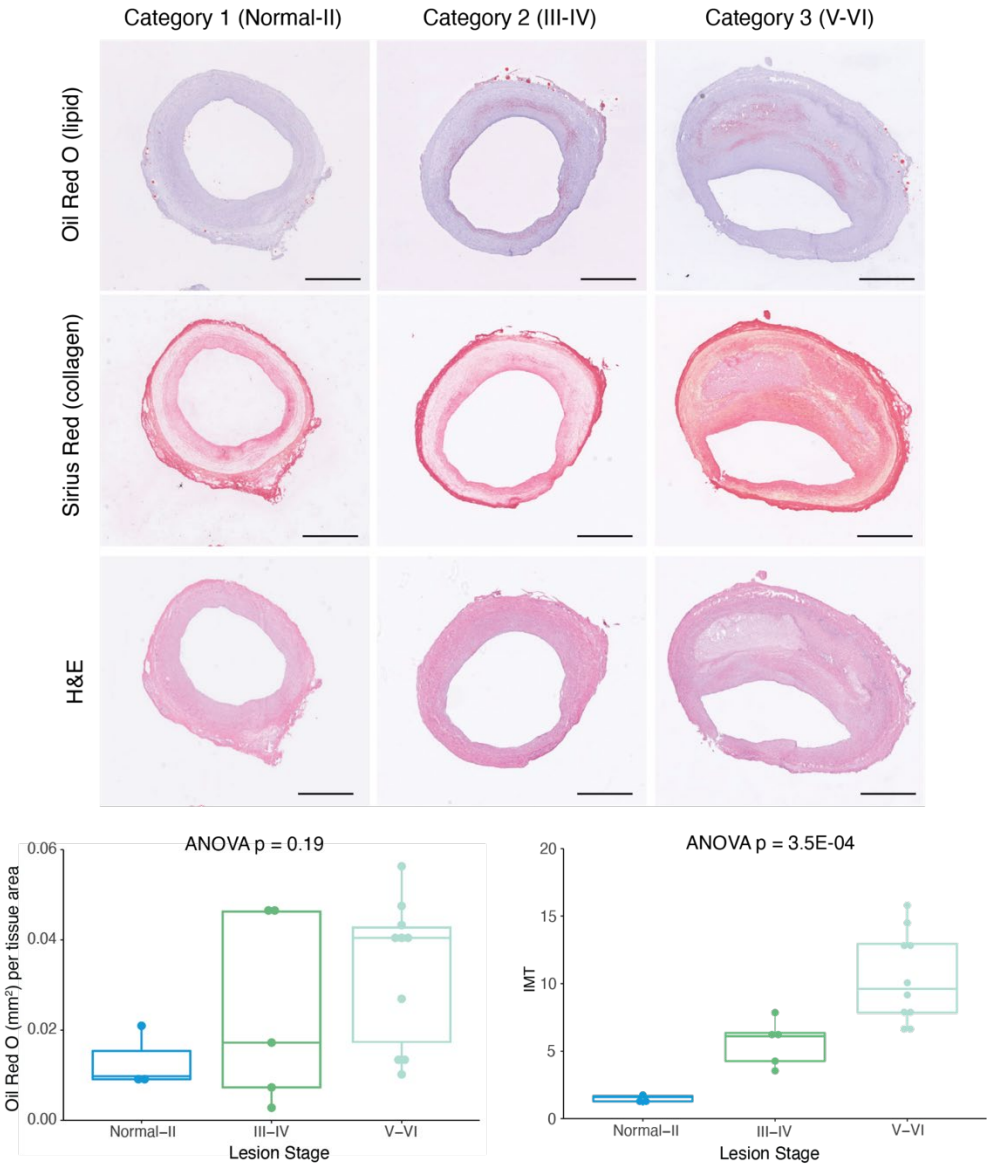
Rebuttal Figure 3. Differentially accessible marker genes between different disease categories. Highlighted genes are immune genes that harbor more accessible peaks in advanced diseased coronary segments (category 3) relative to healthy segments (category 1). Also contractile SMC marker genes are more accessible in healthy artery segments relative to diseased segments. Differential marker genes were detected at FDR <= 0.1 and log2FC > log2(1.5).



Rebuttal Figure 4. Comparison of effect size directions between smooth muscle cell caQTLs (5% FDR) and bulk coronary artery caQTLs (5% FDR), as visualized in scatter plot (a) and donut plot (b). For this analysis, 503 caQTL peaks are shared between both datasets (peaks with a corresponding significant caQTL variant). The rsID reported in the SMC caQTL results (n=40 individuals) was compared with the rsID reported in the bulk caQTL results (n=35 individuals). Two variants were considered to be in linkage disequilibrium (LD) if the r2 value between them was between 0.2 and 1 (in EUR population). If variants had an r2 value < 0.2 (in EUR population), the variants were considered to be in low LD (blue). For the caQTL effect size direction, we considered the RASQUAL Pi statistic. The RASQUAL Pi statistic can range from 0-1, where Pi < 0.5 reflects lower peak accessibility for the alternative allele and Pi > 0.5 reflects higher accessibility for the alternative allele. The effect sizes for linked variants go in the same direction (green) if the Pi values in SMCs and bulk coronary artery are both < 0.5 or both > 0.5. Linear regression line and Pearson correlation coefficient shown in (a).

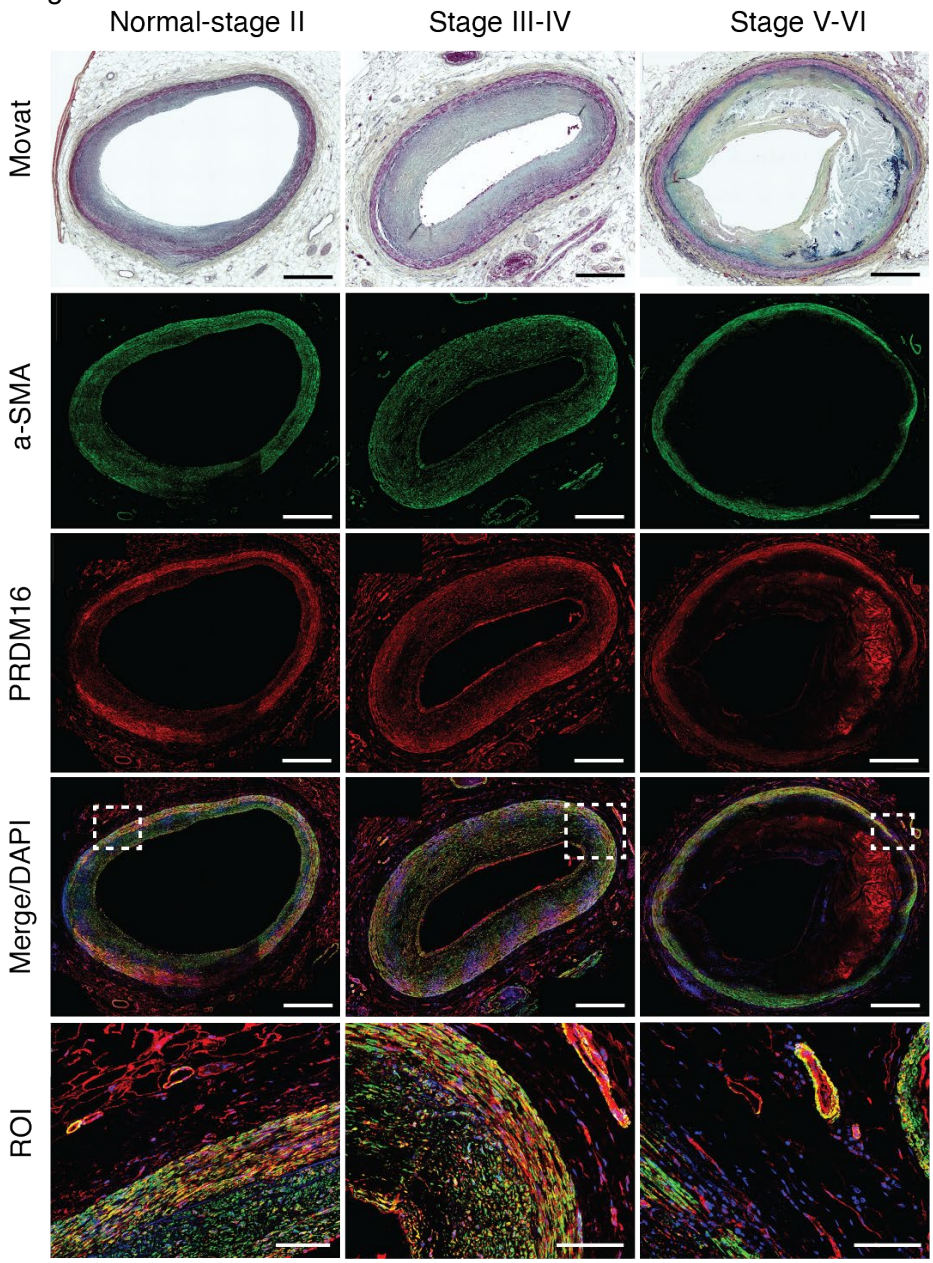


Rebuttal Figure 5. UMAP and box plots of imputed geneScores for known adipocyte marker genes (*UCP1*, *CITED1*, and *ZIC1*) demonstrating the absence of adipocytes in our coronary artery snATAC dataset. This is consistent with the fact that perivascular adipose was trimmed prior to archiving these coronary artery specimens.

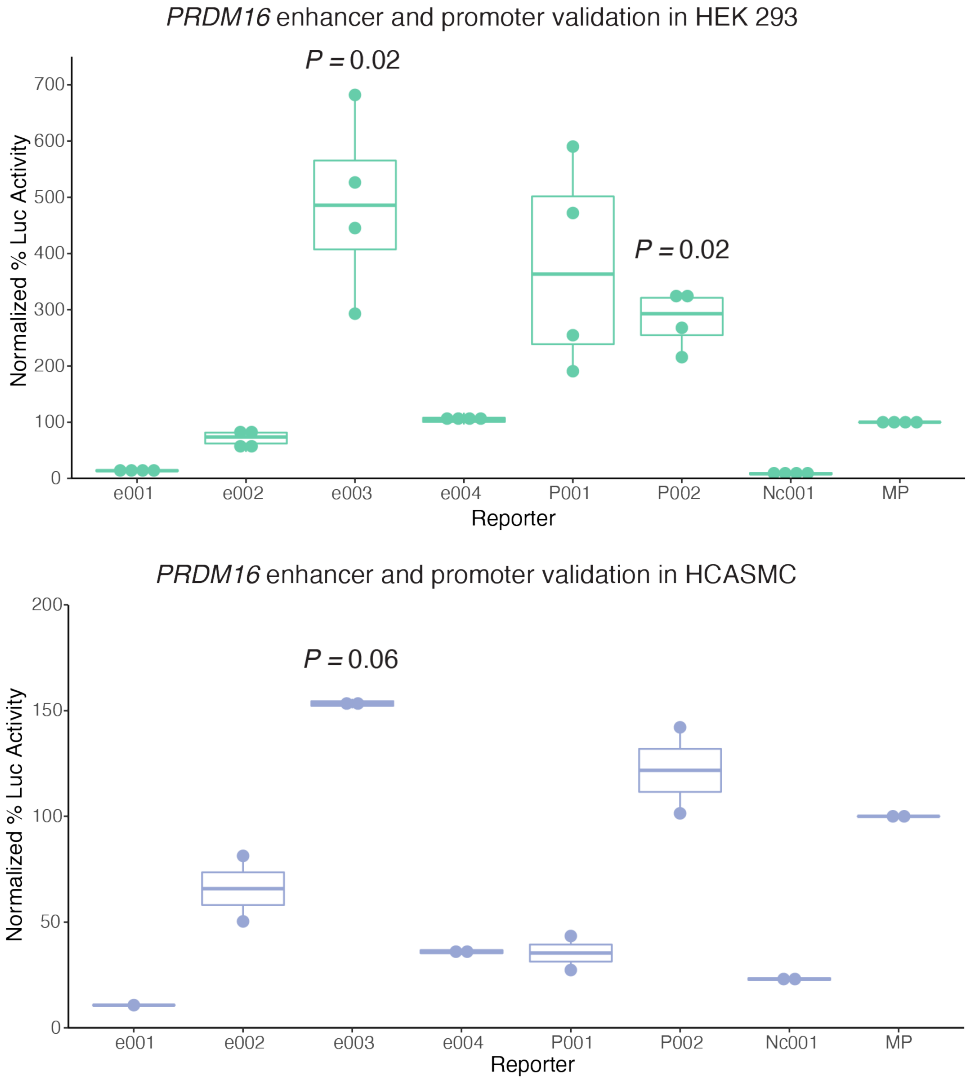


Rebuttal Figure 6. Representative histology staining of adjacent frozen human coronary artery sections at different disease categories used for snATAC profiling (n=4 donors per category). Category 1 reflects normal to Stary atherosclerosis stage I/II lesions with adaptive intimal thickening and early lipid (Oil Red O (ORO)) and collagen (Sirius Red) accumulation in the subintimal layer. Category 2 reflects Stary stage III/IV early/intermediate atheroma lesions with increased lipid and collagen accumulation and proliferation (Hematoxylin & Eosin (H&E)). Category 3 reflects Stary stage V/VI advanced fibroatheroma or complex lesions with more severe lipid and collagen deposition as well as lipid core and thin media layer. (Below) Whole

slide quantitative results of ORO area (mm^2) normalized to overall tissue area and Sirius Red based quantitation of intima-media thickness (IMT) with maximum intima and average media width captured from >6 automatically defined measurements (Methods). n=3, n=5, and n=10 donors per lesion stage, respectively. ANOVA p-values shown for comparisons across lesion stages. Scale bar = 1mm.



Rebuttal Figure 7. Movat pentachrome staining and PRDM16 (red) and alpha-smooth muscle actin (a-SMA) (green) immunofluorescence staining of atherosclerotic human coronary artery segments - left anterior descending (LAD) from normal-Stage I, Stage III-IV, and Stage V-VI lesions based on Stary classification stages. Whole slide images captured from 20x confocal microscopy stitched tiles. PRDM16/a-SMA co-staining (see arrows) depicted in yellow from merged images. DAPI (blue) marks nuclei. n = 4 per group. Scale bar = 1 mm, except for region of interest (ROI): scale bar = 100 μm.



Rebuttal Figure 8. (Top) Luciferase reporter assays evaluating the transcriptional activity of 4 candidate enhancer regulatory regions overlapping CAD variants in lentiviral transduced HEK 293 cells. Candidate regions were cloned into the pLS-mP-Luc plasmid (Addgene plasmid # 106253). Dots represent mean of triplicates from n=4 biological replicate experiments. Boxes and whiskers represent the median with upper and lower quartiles. (Bottom) Luciferase reporter assays evaluating the transcriptional activity of 4 candidate enhancer regulatory regions overlapping CAD variants in lentiviral transduced human coronary artery smooth muscle cells. Dots represent mean of triplicates from n=2 biological replicate experiments. Box and whiskers represent the median with upper and lower quartiles. e001-e004 are CAD enhancer regions, P001 and P002 are endogenous PRDM16/PRDM16-DT promoter sequences, Nc001 is a

negative control enhancer sequence, and MP is the minimal promoter and empty vector from pMCS-Luc. Values are presented as % normalized luciferase of MP reporter. Statistical significance determined from % change of test reporters versus MP reporter using Wilcoxon Rank Test with continuity correction.

Rebuttal Table 1. Candidate regulatory regions at the PRDM16 locus used for luciferase assays in cultured cells.

Enhancer candidate name	**Region (hg38)**	**PRDM16 CAD-associated SNPs**
e001	chr1:2998925-2999935	rs72629460, rs12239064, rs12240128, rs67927838, rs59653178, rs67142023, rs72629462
e002	chr1:3072552-3073451	rs2981890
e003	chr1:3074233-3074650	rs35397508
e004	chr1:3075236-3075530	No SNPs but highly conserved
p001 (*promoter)	chr1:3067941-3069685	PRDM16 promoter
p002 (*promoter)	chr1:3067041-3068642	PRDM16-DT promoter rs2297829
nc001	chr1:3230824-3231620	Negative control sequence

Summary of changes

Figures

Figure	Change/Addition
Figure 3	Added a legend to the trajectory heatmap (Figure 3d)
Figure 4	Changed the start value of the scale from 0 to 1 in Figure 4d
Figure 5	Updated caQTL panels after removing individuals with low numbers of cells
Figure 6	Added additional panels for PRDM16

Supplementary Figures

Supplementary Figure	Change/Addition
4	Added representative histology images for various stages of CAD
6	Added plots showing correlation of snATAC promoter accessibility with integrated scRNA-seq expression
7	Added plots for super enhancer analyses

11	Added donut plot showing the correspondence of effect size directions between smooth muscle caQTLs and bulk coronary artery caQTLs
14	Added UMAP feature plots of PRDM16 and TBX2 in human and mouse atherosclerosis scRNA-seq datasets
15	Added additional representative whole slide histology and confocal immunofluorescence images for PRDM16, a-SMA, and LMOD1

Supplementary Tables/Supplementary Data

We have now moved many of the prior Supplementary Tables to separate Supplementary Data files below. We feel this repackaging makes the supplemental information much more organized than before.

Supplementary Data File	**Details**
1	Top snATAC marker genes in each coronary artery cell type
2	Consensus set of 323,767 coronary artery peaks across all cell types
3	List of enriched transcription factor motifs within coronary artery cell type peaks
4	Differential peak and promoter analysis results between differentiated smooth muscle cells and fibromyocytes
5	Overlap of CAD GWAS SNPs with coronary artery cell type peaks and Peak2Gene link coordinates. For the Peak2Gene links, the peak coordinates now match the peak set used in the rest of the manuscript
6	Chromatin accessibility QTLs within individual cell types calculated using RASQUAL (shown are SNPs passing 5% FDR threshold). For the updated caQTL analysis we excluded individuals from the analysis if they contained <20 nuclei belonging to the respective cell type
7	Chromatin accessibility QTLs from bulk coronary artery samples calculated using RASQUAL (shown are SNPs

	passing 5% FDR threshold)
8	Significant results for functional CAD variant prediction using gkm-SVM, gkmExplain, and DeltaSVM
9	Sample size estimations for top fibromyocyte genes (comparing smooth muscle cells and fibromyocytes)
10	eQTLs for PRDM16 and TBX2 in STARNET and GTEx artery tissues

In the new version of the Supplementary Table file we have changed or added the following:

Table	**Change/Addition**
ST2	Added the number of nuclei in each cell type captured and detected using 10x Genomics Cell Ranger ATAC QC
ST3	Added the number of nuclei analyzed in each sample per cell type (after stringent ArchR QC filtering)
ST5	Summary of snATAC peaks overlapping coronary artery smooth muscle cell super enhancers (SE)

Bibliography

Chiou, J., Geusz, R.J., Okino, M.-L., Han, J.Y., Miller, M., Melton, R., Beebe, E., Benaglio, P., Huang, S., Korgaonkar, K., et al. (2021). Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature *594*, 398–402.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Craps, S., Van Wauwe, J., De Moudt, S., De Munck, D., Leloup, A.J., Boeckx, B., Vervliet, T., Dheedene, W., Criem, N., Geeroms, C., et al. (2021). Prdm16 supports arterial flow recovery by maintaining endothelial function. Circ. Res.

Davis, A., Gao, R., and Navin, N.E. (2019). SCOPIT: sample size calculations for single-cell

sequencing experiments. BMC Bioinformatics *20*, 566.

Depuydt, M.A.C., Prange, K.H.M., Slenders, L., Örd, T., Elbersen, D., Boltjes, A., de Jager, S.C.A., Asselbergs, F.W., de Borst, G.J., Aavik, E., et al. (2020). Microanatomy of the Human Atherosclerotic Plaque by Single-Cell Transcriptomics. Circ. Res. *127*, 1437–1455.

Fernandez, D.M., Rahman, A.H., Fernandez, N.F., Chudnovskiy, A., Amir, E.-A.D., Amadori, L., Khan, N.S., Wong, C.K., Shamailova, R., Hill, C.A., et al. (2019). Single-cell immune landscape of human atherosclerotic plaques. Nat. Med. *25*, 1576–1588.

Franzén, O., Ermel, R., Cohain, A., Akers, N.K., Di Narzo, A., Talukdar, H.A., Foroughi-Asl, H., Giambartolomei, C., Fullard, J.F., Sukhavasi, K., et al. (2016). Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science *353*, 827–830.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

van Kuijk, K., Kuppe, C., Betsholtz, C., Vanlandewijck, M., Kramann, R., and Sluimer, J.C. (2019). Heterogeneity and plasticity in healthy and atherosclerotic vasculature explored by single-cell sequencing. Cardiovasc. Res. *115*, 1705–1715.

Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat. Genet. *48*, 206–213.

Lin, J.-D., Nishi, H., Poles, J., Niu, X., Mccauley, C., Rahman, K., Brown, E.J., Yeung, S.T., Vozhilla, N., Weinstock, A., et al. (2019). Single-cell analysis of fate-mapped macrophages reveals heterogeneity, including stem-like properties, during atherosclerosis progression and regression. JCI Insight *4*.

Lin, M.-E., Chen, T., Leaf, E.M., Speer, M.Y., and Giachelli, C.M. (2015). Runx2 expression in smooth muscle cells is required for arterial medial calcification in mice. Am. J. Pathol. *185*, 1958–1969.

Lin, M.-E., Chen, T.M., Wallingford, M.C., Nguyen, N.B., Yamada, S., Sawangmake, C., Zhang, J., Speer, M.Y., and Giachelli, C.M. (2016). Runx2 deletion in smooth muscle cells inhibits vascular osteochondrogenesis and calcification but not atherosclerotic lesion formation. Cardiovasc. Res. *112*, 606–616.

Liu, B., Pjanic, M., Wang, T., Nguyen, T., Gloudemans, M., Rao, A., Castano, V.G., Nurnberg,

S., Rader, D.J., Elwyn, S., et al. (2018). Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci. Am. J. Hum. Genet. *103*, 377–388.

Long, J.Z., Svensson, K.J., Tsai, L., Zeng, X., Roh, H.C., Kong, X., Rao, R.R., Lou, J., Lokurkar, I., Baur, W., et al. (2014). A smooth muscle-like origin for beige adipocytes. Cell Metab. *19*, 810–820.

Mandric, I., Schwarz, T., Majumdar, A., Hou, K., Briscoe, L., Perez, R., Subramaniam, M., Hafemeister, C., Satija, R., Ye, C.J., et al. (2020). Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. Nat. Commun. *11*, 5504.

Miller, C.L., Pjanic, M., Wang, T., Nguyen, T., Cohain, A., Lee, J.D., Perisic, L., Hedin, U., Kundu, R.K., Majmudar, D., et al. (2016). Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. Nat. Commun. *7*, 12092.

Munz, M., Wohlers, I., Simon, E., Reinberger, T., Busch, H., Schaefer, A.S., and Erdmann, J. (2020). Qtlizer: comprehensive QTL annotation of GWAS results. Sci. Rep. *10*, 20417.

Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238–243.

Numaguchi, R., Furuhashi, M., Matsumoto, M., Sato, H., Yanase, Y., Kuroda, Y., Harada, R., Ito, T., Higashiura, Y., Koyama, M., et al. (2019). Differential phenotypes in perivascular adipose tissue surrounding the internal thoracic artery and diseased coronary artery. J. Am. Heart Assoc. *8*, e011147.

O'Connor, L.J., Schoech, A.P., Hormozdiari, F., Gazal, S., Patterson, N., and Price, A.L. (2019). Extreme polygenicity of complex traits is explained by negative selection. Am. J. Hum. Genet. *105*, 456–476.

Örd, T., Õunap, K., Stolze, L.K., Aherrahrou, R., Nurminen, V., Toropainen, A., Selvarajan, I., Lönnberg, T., Aavik, E., Ylä-Herttuala, S., et al. (2021). Single-Cell Epigenomics and Functional Fine-Mapping of Atherosclerosis GWAS Loci. Circ. Res. *129*, 240–258.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358.

Stary, H.C., Chandler, A.B., Dinsmore, R.E., Fuster, V., Glagov, S., Insull, W., Rosenfeld, M.E., Schwartz, C.J., Wagner, W.D., and Wissler, R.W. (1995). A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation *92*, 1355–1374.

Wang, X., and Goldstein, D.B. (2020). Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. Am. J. Hum. Genet. *106*, 215–233.

Wang, X., Cairns, M.J., and Yan, J. (2019). Super-enhancers in transcriptional regulation and genome organization. Nucleic Acids Res. *47*, 11481–11496.

Wirka, R.C., Wagh, D., Paik, D.T., Pjanic, M., Nguyen, T., Miller, C.L., Kundu, R., Nagao, M., Coller, J., Koyano, T.K., et al. (2019). Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. Nat. Med. *25*, 1280–1289.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952–1958.

Zimmerman, K.D., and Langefeld, C.D. (2021). Hierarchicell: an R-package for estimating power for tests of differential expression with single-cell data. BMC Genomics *22*, 319.

Decision Letter, first revision:

Our ref: NG-A57752R1

15th Dec 2021

Dear Clint,

Thank you for submitting your revised manuscript "Cell-specific chromatin landscape of human coronary artery resolves regulatory mechanisms of disease risk" (NG-A57752R1). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Genetics, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTex)-- we can not proceed with PDFs at this stage.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements soon. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics Please do not hesitate to contact me if you have any questions.

Sincerely,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

Reviewer #1 (Remarks to the Author):

The authors have made exhaustive changes to the manuscript and have made great progress in addressing all the issue. I therefore recommend this work for publication.

Reviewer #2 (Remarks to the Author):

The authors have provided a detailed response to previous reviews with acknowledgment of the caveats inherent in this type of analysis and addition of new and relevant information.

Reviewer #3 (Remarks to the Author):

The authors have responded thoughtfully and completely to the previous reviews and the manuscript is now acceptable. There are just a few minor edits that I would suggest.

1) The manuscript still refers to 'single-cell ATAC-seq' rather than 'single-nucleus ATAC-seq' in several places, including figures.
2) The numbers indicated on the UMAP of Suppl. Fig. 1F do not seem to be necessary and are distracting.
3) Figure 2E needs a more informative axis title. Is this fold enrichment?
4) 'MAM' should be defined in Suppl. Fig. 13.

Author Rebuttal, second revision:

Point-by-point response to Nature Genetics, 2021, Turner and Hu et al.

We thank all of the reviewers for their helpful final comments on our manuscript. Below we provide point-by-point responses (in blue) to each comment/remark from the three reviewers. At the end we also detail specific additions/changes we made to the manuscript. Specific changes to the enclosed manuscript text are also highlighted in blue for clarity.

Reviewer Comments:

Reviewer #1:
Remarks to the Author:
The authors have made exhaustive changes to the manuscript and have made great progress in addressing all the issue. I therefore recommend this work for publication.

Thank you for the positive feedback on our revised manuscript and recommendation for publication.

Reviewer #2:
Remarks to the Author:
The authors have provided a detailed response to previous reviews with acknowledgment of the caveats inherent in this type of analysis and addition of new and relevant information.

Thank you for the positive feedback on our revised manuscript and recommendation for publication.

Reviewer #3:
Remarks to the Author:
The authors have responded thoughtfully and completely to the previous reviews and the manuscript is now acceptable. There are just a few minor edits that I would suggest.

Thank you for the positive feedback on our revised manuscript and additional minor edits.

1) The manuscript still refers to 'single-cell ATAC-seq' rather than 'single-nucleus ATAC-seq' in several places, including figures.

Thank you for this comment. We have now changed all instances of 'single-cell ATAC-seq' or 'scATAC-seq' to 'single-nucleus ATAC-seq' or 'snATAC-seq' throughout the manuscript, including figures and legends.

2) The numbers indicated on the UMAP of Suppl. Fig. 1F do not seem to be necessary and are distracting.

Thank you for this comment. We have now omitted the numbers overlapping the clusters in the UMAP of Supplementary Figure 1F.

3) Figure 2E needs a more informative axis title. Is this fold enrichment?

Thank you for this comment. We have now added a more informative x-axis title for Figure 2E – 'Normalized deviation score' and moved the motif name above the plot.

4) 'MAM' should be defined in Suppl. Fig. 13.

Thank you for this comment. We have now defined ‘MAM’ as ‘mammary artery’ in the figure and legend for Suppl. Fig. 13.

Final Decision Letter:

In reply please quote: NG-A57752R2 Miller

31st Mar 2022

Dear Clint,

I am delighted to say that your manuscript "Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk" has been accepted for publication in an upcoming issue of Nature Genetics.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Genetics style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system.

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press Office (press@nature.com) after sending your e-proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NG-A57752R2) and the name of the journal, which they will need when they contact our Press Office.

Before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that Nature Genetics is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Please note that Nature Research offers an immediate open access option only for papers that were first submitted after 1 January, 2021.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this

method.

If you have not already done so, we invite you to upload the step-by-step protocols used in this manuscript to the Protocols Exchange, part of our on-line web resource, natureprotocols.com. If you complete the upload by the time you receive your manuscript proofs, we can insert links in your article that lead directly to the protocol details. Your protocol will be made freely available upon publication of your paper. By participating in natureprotocols.com, you are enabling researchers to more readily reproduce or adapt the methodology you use. Natureprotocols.com is fully searchable, providing your protocols and paper with increased utility and visibility. Please submit your protocol to https://protocolexchange.researchsquare.com/. After entering your nature.com username and password you will need to enter your manuscript number (NG-A57752R2). Further information can be found at https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#protocols

Sincerely,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087