Single-cell RNA-seq analysis of human coronary arteries using an enhanced workflow reveals SMC transitions and candidate drug targets

Wei Feng Ma[1,2], Chani J. Hodonsky[2], Adam W. Turner[2], Doris Wong[2,3], Yipei Song[2,4], Nelson B. Barrientos[2,5], Clint L. Miller[2,3,5]*

Affiliations

[1]Medical Scientist Training Program, University of Virginia, Charlottesville, VA 22908, USA.

[2]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA.

[3]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA.

[4]Department of Computer Engineering, University of Virginia, Charlottesville, VA 22908, USA.

[5]Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA.

*corresponding author

Correspondence:

Clint L. Miller, PhD. (clintm@virginia.edu)

University of Virginia

Center for Public Health Genomics

PO Box 800717

Charlottesville, VA 22908 USA

Abstract

Recent advances in single-cell RNA sequencing (scRNA-seq) methods have enabled high-resolution profiling and quantification of cellular expression and transcriptional states. Here we incorporate automated cell labeling, pseudotemporal ordering, ligand-receptor evaluation, and drug-gene interaction analysis into an enhanced and reproducible scRNA-seq analysis workflow. We applied this analysis method to a recently published human coronary artery scRNA dataset and revealed distinct derivations of chondrocyte-like and fibroblast-like cells from smooth muscle cells (SMCs). We highlighted several key ligand-receptor interactions within the atherosclerotic environment through functional expression profiling and revealed several attractive avenues for future pharmacological repurposing. This publicly available workflow will also allow for more systematic and user-friendly analysis of scRNA datasets in other disease systems.

Keywords

Atherosclerosis, single-cell RNA-seq, analysis, pseudotemporal, drug-target, workflow

Abbreviations

C7- complement component C7

CAD- coronary artery disease

CH- chondrocytes

CMP- common myeloid progenitor cells

DCN- decorin

DGIdb- drug-gene interaction database

EC- endothelial cells

EGFR- epidermal growth factor receptor

FB- fibroblasts

FBLN1- fibulin 1

GMP- granulocyte-monocyte progenitor cells.

Mø- macrophages

MYH11- myosin heavy chain 11

SC- stem cells

sc/snATAC-seq - single cell/single nucleus assay for transposase-accessible chromatin sequencing

sc/snRNA-seq- single cell/single nucleus RNA sequencing

SMC- smooth muscle cells

UMAP- uniform manifold approximation and projection

Background

Atherosclerosis is a complex process involving chronic inflammation and hardening of the vessel wall and represents one of the major causes of coronary artery disease (CAD), peripheral artery disease, and stroke (Basatemur et al., 2019). Rupture of an unstable atherosclerotic lesion can lead to the formation of a thrombus, causing complete or partial occlusion of a coronary artery (Argraves et al., 2009). The contribution of smooth muscle cells (SMCs) to both lesion stability and progression has recently been established by numerous groups, but the exact mechanisms in which SMCs modulate the atherosclerotic microenvironment and whether pharmacological agents can be used to selectively counter SMC-related deleterious mechanisms are still under investigation (Bennett et al., 2016; Pan et al., 2020; Wirka et al., 2019).

The recent advances in single-cell RNA-sequencing (scRNA-seq) have allowed for ultra-fine gene expression profiling of many diseases at the cellular level, including coronary artery disease (Wirka et al., 2019). As sequencing costs continue to decline, there has also been a consistent growth in scRNA datasets, analysis tools and applications. Currently, a major challenge with scRNA-seq analysis is the inherent bias introduced during manual cell labeling, in which cells are grouped by cluster and their identities called collectively based on their overall differential gene expression profiles (Aran et al., 2019). Another draw-back inherent to commonly used scRNA-seq protocols is their destructive nature to the cells, making time-series analyses of the same samples impossible. Instead, these studies must rely on time-points from separate libraries to monitor processes such as clonal expansion and cell differentiation (DiRenzo et al., 2017; Wang et al., 2020).

Recently, new approaches have been developed to compensate for both of these shortcomings, namely automatic cell labeling and pseudotemporal analysis. Tools such as 'SingleR' and 'Garnett' have been used to assign unbiased identities to individual cells using reference-based and machine learning algorithms, respectively (Aran et al., 2019; Pliner et al., 2019). On the other hand, tools such as 'Monocle3' and 'scVelo' align and project cells onto a pseudotemporal space where each cell becomes a snapshot within the single-cell time continuum (Bergen et al., 2019; Cao et al., 2019). In essence, the single scRNA-seq dataset is transformed into a time series (Bergen et al., 2019; Cao et al., 2019; Manno et al., 2018). Although the pseudotemporal scale does not reflect the actual time scale, it is a reliable approximation to characterize cell fate and differentiation events, e.g., during organogenesis, disease states, or in response to SARS-CoV-2 infection (Cao et al., 2019; Chua et al., 2020).

In this study, we present the application of an enhanced, scalable, and user-friendly scRNA-seq analysis workflow on an existing human coronary artery scRNA-seq dataset. We performed unbiased automatic cell identification at the single-cell level, pseudotemporal analysis, ligand-receptor expression profiling, and drug repurposing analysis. Our results demonstrate potential new mechanisms by which SMCs contribute to the atherosclerotic phenotype and signaling within the lesion microenvironment. More importantly, we reveal attractive candidates for future pre-clinical drug interventional studies. This reproducible analysis workflow can also be easily modified and extended to incorporate different tissue data sources and single-cell modalities such as snATAC-seq (Smith and Sheffield, 2020).

Results and Discussion

Unbiased automatic cell labeling reveals abundant cells with chondrocyte and fibroblast characteristics.

Recently, automatic cell identifications tools have been introduced to compensate the shortcomings of manual, cluster-based cell labeling (Aran et al., 2019). For example, 'SingleR' and 'Garnett' uses reference-based and machine learning algorithms, respectively, to call individual cell identities (Aran et al., 2019; Pliner et al., 2019). Using 'SingleR,' which uses known purified cell expression data as references, we found that endothelial cells (ECs) make up the highest proportion of cells in this dataset (16.21%, Fig. 1), followed by smooth muscle cells (SMCs, 13.8%) and stem cells (SCs, 14.06%), where the latter could be so-called "atherosclerotic stem cells" or normal stromal stem cells but cannot be distinguished until specific expression profile references are developed in the future (Wang et al., 2020). Consistent with recent scRNA-seq studies in atherosclerotic models, we identified abundant fibroblast (FB) and chondrocyte-like (CH) cells, as well as cells with an osteoblast-like (OS) expression profile (Fig. 1) (Pan et al., 2020). In the UMAP clusters reflecting single-cell identities, there was a substantial presence of SMC and FB cells in the OS and SC cluster. Such heterogeneity in cell clusters would have been overlooked in manual cluster-based cell labeling.

Pseudotemporal ordering identifies distinct chondrocyte and fibroblast-like cell differentiation states from smooth muscle cells.

To evaluate putative cell fate decisions or differentiation events (e.g., SMC phenotypic transition states), we performed pseudotemporal analysis and ultra-fine clustering using 'Monocle3', a method previously applied to normal and diseased states, e.g. embryo organogenesis and response to COVID19 infection, respectively (Cao et al., 2019; Chua et al., 2020). We found evidence of SMCs giving rise to CH and FB-like cells (Fig. 2A). This corroborates earlier findings showing that SMCs may transition or de-differentiate into 'fibromyocytes'—SMCs that have undergone a phenotypic modulation within the atherosclerotic lesions (Wang et al., 2020;

Wirka et al., 2019). Genes associated with healthy SMC phenotypes, such as *MYH11* (a canonical marker of SMC), *IGFBP2* (associated with decreased visceral fat), and *PPP1R14A* (which enhances smooth muscle contractions), are decreased by approximately 50-75% along the SMC trajectory as these cells become more FB-like (Table 1, Fig. 2, $p < 0.1E-297$) (Bennett et al., 2016; Carter et al., 2018). Similar results were found by another group using mouse lineage-traced models where MYH11 expression was decreased in SMC-derived modulated "intermediate cell states" (Pan et al., 2020). More importantly, specific inflammatory markers and proteins associated with thrombotic events during CAD, including *C7* and *FBLN1*, are increased along the same trajectory (Argraves et al., 2009; Carter, 2012)(Fig. 2B). Together, and in corroboration of recent studies, our pseudotemporal analysis demonstrates that SMCs could be a source of FB and CH-like cells, with the former associated with an intermediate atherosclerotic cell phenotype, and the latter expresses genes associated with a more advanced atherosclerotic phenotype (Pan et al., 2020; Wirka et al., 2019). This is further supported by the recent study where blocking of SMC-derived intermediate cells coincides with less severe atherosclerotic lesions (Pan et al., 2020). Precisely how these cells might influence the overall stability of the atherosclerotic lesion and patient survival requires additional longitudinal studies using genetic models (Bennett et al., 2016; Pan et al., 2020; Wang et al., 2020).

Comprehensive ligand-receptor analysis shows complex intercellular communications in the human coronary micro-environment and reveals potential drug targets.

To examine the potential cross-talk between different cell types using scRNA-seq data, we compared the ligand and receptor expression profiles of each cell type with experimentally-validated interactions using 'scTalk' (Farbehi et al., 2019). We found that there is an intricate network of signaling pathways connecting different cell types; some cell types, such as OS,

have stronger and more frequent outgoing signals, whereas other cell types such as Macrophages (Mø) have fewer and weaker incoming and outgoing signals (Fig. 3A). SMCs, OSs, and neurons also exhibit a high degree of autocrine signaling profiles (Fig. 3A). Specifically, SMCs are shown to have the highest number of outgoing signals and are among those with the least number of incoming signal weights (Fig. 3B). This suggests that SMCs play an important role in regulating the coronary microenvironment by transducing signals to neighboring cells in the lesion.

Specifically, we identified three significant ligand-receptor interactions between SMCs and FBs, FBLN1-ITGB1, APOD-LEPR, and DCN-EGFR (Fig. 4). We searched for potentially druggable targets to interrupt SMC-FB communication by performing an integrative analysis of the identified ligand-receptor interactions with known drug-gene interactions using the DGIdb 3.0 database (Cotto et al., 2017). Interestingly, anti-EGFR (epidermal growth factor receptor)-based cancer treatments such as erlotinib, cetuximab, and gefitinib were shown as potential key mediators of signaling pathways between SMC and FB via DCN (decorin) and EGFR (Fig. 4). It has been shown that decorin overexpression increases SMC aggregation and SMC-induced calcification at the atherosclerotic plaque (Fischer et al., 2004). Although the overlap between CAD and cancer has been previously noted, the efficacy and mechanisms of chemotherapy, such as erlotinib, in the pathogenesis and stability of CAD or their adjuvant use in cancer patients to treat CAD continues to hold promise for future translational studies (Camaré et al., 2017; DiRenzo et al., 2017; Tapia-Vieyra et al., 2017).

Conclusions

Our findings show that an enhanced, reproducible pipeline for scRNA-seq analysis improve on current standard scRNA-seq bioinformatics protocols. We provide new insights into intricate

cellular differentiation and communication pathways while providing actionable and testable targets for future studies (Fig. 5). In our combined analysis, we found that SMCs give rise to substantial proportions of CH and FB, with the latter associated with worse prognostic markers (Argraves et al., 2009; Carter, 2012; Carter et al., 2018; Pan et al., 2020). Several FDA-approved drugs (e.g., erlotinib, cetuximab, and gefitinib) were shown as potential effectors of SMCs' signaling to FB, and may be used to treat CAD in cancer patients to simplify or augment drug regiments (Camaré et al., 2017). This is consistent with recent reports showing beneficial effects of the acute promyelocytic drug all-trans-retinoic acid (ATRA) in atherosclerosis mouse models (Pan et al., 2020).

Although the utilization of this workflow can compensate for many of the shortcomings of current scRNA-seq analyses, we are still unable to perform cell-lineage tracing that reflects actual timescale without additional molecular experiments. However, leveraging mitochondrial DNA variants in snATAC-seq data has enabled lineage tracing analysis in human cells (Lareau et al., 2020; Xu et al., 2019). Likewise, these analyses can ultimately be extended to integrate spatial data and other multi-modal data (Stuart et al., 2019). In the future as spatial transcriptomics and snATAC-seq data become more widely available, this workflow can be modified to discover signaling pathways or differentiation events at specific tissue locations and time, allowing for more disease-relevant drug-gene interaction analyses (Fig. 5). Nonetheless, this pipeline can be applied immediately to datasets from other tissues/diseases to generate informative directions for follow-up studies, and is more user-friendly and reproducible compared to standard scRNA analyses.

Methods

Data retrieval and pre-processing

Human coronary artery scRNA data read count matrix was retrieved from the Gene Expression Omnibus (GEO) using #GSE131780 and loaded into R 4.0, and was preprocessed using standard parameters of the R packages 'Seurat' v.3, and 'Monocle3' as required (Satija et al., 2015; Team, 2020; Trapnell et al., 2014; Wirka et al., 2019). Uniform manifold approximation projections (UMAP) clusters from 'Seurat' were imported into 'Monocle3' before pseudotemporal analysis.

Automatic cell Identification and pseudotemporal ordering

scRNA read matrices were read into SingleR as previously described for cell labeling (Aran et al., 2019). Briefly, SingleR compares each cell's gene expression profile with known human primary cell atlas data and gives the most likely cell identity independently. SingleR first corrects for batch effects, then calculates the expression correlation scores for each test cell to each cell type in the reference, and the cell identity is called based on reference cell type exhibiting the highest correlation. Then, pseudotemporal analyses were performed as previously described in the analysis of embryo organogenesis (Cao et al., 2019; Trapnell et al., 2014). Briefly, the UMAP clusters were passed into Monocle3 and then the 'learn_graph()' and 'order_cells()' functions. The SMCs and related clusters were then subsetted for detailed subclustering and analysis. For each cluster, Moran's I statistics were calculated, which identify genes that are differentially expressed along their trajectories. Detailed codes to reproduce the figures in this publication can be found at the Miller Lab Github (see Availability of data and materials).

Ligand-receptor cell communication analysis

We analyzed candidate ligand-receptor interactions to infer cell communication using the R package 'scTalk', as previously described in the analysis of glial cells (Farbehi et al., 2019). Briefly, this method is based on permutation testing of random networks, where ligand-receptor

interactions are derived from experimentally derived interactions from the STRING database. We exported statistically significant differentially expressed genes from 'Seurat' using the 'FindMarkers()' function and imported the preprocessed data into 'scTalk'. Then, overall edges of the cellular communication network were calculated using the 'GenerateNetworkPaths()' function, which reflects the overall ligand-receptor interaction strength between each cell type. Then, the cell types of interest were specified and treeplots were generated using the 'NetworkTreePlot()' function.

Gene-drug interaction analysis

The above identified ligand and receptor interaction pairs were fed into the Drug-Gene Interaction database (DGIdb 3.0) to reveal candidate drug-gene interactions (Cotto et al., 2017). Briefly, ligands and receptors that were deemed significant from 'scTalk' were evaluated using the 'queryDGIdb()' function of the 'rDGIdb' R package (Cotto et al., 2017). We included all top FDA-approved drugs produced with verified inhibitory or antagonistic activities. Figure 4 and 5 were modified using BioRender for clarity.

Declarations

No conflicts of interest to disclose.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data are publicly available on GEO, accession number GSE131778, as previously described (Wirka et al., 2019). All codes and analysis pipelines can be viewed at github.com/MillerLab-CPHG/Ma_2020 and repurposed for any scRNA datasets.

Competing interests

The authors declare no competing interests.

Funding

Funding support was provided by grants from the National Institutes of Health (NIH): R00HL125912 (CLM), R01HL148239 (CLM) and T32HL007284 (CJH and DW); American Heart Association (AHA): POST35120545 (AWT) and Leducq Foundation Transatlantic Network of Excellence (PlaqOmics) (CLM and AWT).

Authors' contributions

WFM designed and performed the statistical analysis. CJH, AWT, DW and YS refined the methodology and edited the manuscript. CLM conceived and refined the project and edited the manuscript.

Acknowledgements

Not applicable.

Works cited

Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. 2019. Reference-based analysis of lung single-cell sequencing

reveals a transitional profibrotic macrophage. *Nat Immunol* **20**:163–172. doi:10.1038/s41590-018-0276-y

Argraves WS, Tanaka A, Smith EP, Twal WO, Argraves KM, Fan D, Haudenschild CC. 2009. Fibulin-1 and fibrinogen in human atherosclerotic lesions. *Histochem Cell Biol* **132**:559–565. doi:10.1007/s00418-009-0628-7

Basatemur GL, Jørgensen HF, Clarke MCH, Bennett MR, Mallat Z. 2019. Vascular smooth muscle cells in atherosclerosis. *Nat Rev Cardiol* **16**:727–744. doi:10.1038/s41569-019-0227-9

Bennett MR, Sinha S, Owens GK. 2016. Vascular Smooth Muscle Cells in Atherosclerosis. *Circ Res* **118**:692-- 702. doi:10.1161/circresaha.115.306361

Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. 2019. Generalizing RNA velocity to transient cell states through dynamical modeling. *Biorxiv* 820936. doi:10.1101/820936

Camaré C, Pucelle M, Nègre-Salvayre A, Salvayre R. 2017. Angiogenesis in the atherosclerotic plaque. *Redox Biol* **12**:18-- 34. doi:10.1016/j.redox.2017.01.007

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, Trapnell C, Shendure J. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**:496–502. doi:10.1038/s41586-019-0969-x

Carter AM. 2012. Complement Activation: An Emerging Player in the Pathogenesis of Cardiovascular Disease. *Sci* **2012**:1–14. doi:10.6064/2012/402783

Carter S, Lemieux I, Li Z, Alméras N, Tremblay A, Bergeron J, Poirier P, Després J-P, Picard F. 2018. Changes in IGFBP-2 levels following a one-year lifestyle modification program are independently related to improvements in plasma apo B and LDL apo B levels. *Atherosclerosis* **281**:89–97. doi:10.1016/j.atherosclerosis.2018.12.016

Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, Debnath O, Thürmann L, Kurth F, Völker MT, Kazmierski J, Timmermann B, Twardziok S, Schneider S, Machleidt F, Müller-Redetzky H, Maier M, Krannich A, Schmidt S, Balzer F, Liebig J, Loske J, Suttorp N, Eils J, Ishaque N, Liebert UG, Kalle C von, Hocke A, Witzenrath M, Goffinet C, Drosten C, Laudi S, Lehmann I, Conrad C, Sander L-E, Eils R. 2020. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat Biotechnol* 1–10. doi:10.1038/s41587-020-0602-4

Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL, Griffith M. 2017. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res* **46**:D1068–D1073. doi:10.1093/nar/gkx1143

DiRenzo D, Owens GK, Leeper NJ. 2017. “Attack of the Clones.” *Circ Res* **120**:624--626. doi:10.1161/circresaha.116.310091

Farbehi N, Patrick R, Dorison A, Xaymardan M, Janbandhu V, Wystub-Lis K, Ho JW, Nordon RE, Harvey RP. 2019. Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *Elife* **8**:e43882. doi:10.7554/elife.43882

Fischer JW, Steitz SA, Johnson PY, Burke A, Kolodgie F, Virmani R, Giachelli C, Wight TN. 2004. Decorin Promotes Aortic Smooth Muscle Cell Calcification and Colocalizes to Calcified Regions in Human Atherosclerotic Lesions. *Arteriosclerosis Thrombosis Vasc Biology* **24**:2391–2396. doi:10.1161/01.atv.0000147029.63303.28

Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, Chiang Z, Pelka K, Verboon JM, Luo W, Christian E, Rosebrock D, Getz G, Boland GM, Chen F, Buenrostro JD, Hacohen N, Wu CJ, Aryee MJ, Regev A, Sankaran VG. 2020. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat Biotechnol* 1–11. doi:10.1038/s41587-020-0645-6

Manno GL, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, Bruggen D van, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV. 2018. RNA velocity of single cells. *Nature* **560**:494--498. doi:10.1038/s41586-018-0414-6

Pan H, Xue C, Auerbach BJ, Fan J, Bashore AC, Cui J, Yang DY, Trignano SB, Liu W, Shi J, Ihuegbu CO, Bush EC, Worley J, Vlahos L, Laise P, Solomon RA, Connolly ES, Califano A, Sims PA, Zhang H, Li M, Reilly MP. 2020. Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. *Circulation*. doi:10.1161/circulationaha.120.048378

Pliner HA, Shendure J, Trapnell C. 2019. Supervised classification enables rapid annotation of cell atlases. *Nat Methods* **16**:983–986. doi:10.1038/s41592-019-0535-3

Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**:495-- 502. doi:10.1038/nbt.3192

Smith JP, Sheffield NC. 2020. Analytical Approaches for ATAC-seq Data Analysis. *Curr Protoc Hum Genetics* **106**. doi:10.1002/cphg.101

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-1902.e21. doi:10.1016/j.cell.2019.05.031

Tapia-Vieyra JV, Delgado-Coello B, Mas-Oliva J. 2017. Atherosclerosis and Cancer; A Resemblance with Far-reaching Implications. *Arch Med Res* **48**:12–26. doi:10.1016/j.arcmed.2017.03.005

Team RC. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are

356 revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381--386.
357 doi:10.1038/nbt.2859

358 Wang Y, Nanda V, Direnzo D, Ye J, Xiao S, Kojima Y, Howe KL, Jarr K-U, Flores AM, Tsantilas
359 P, Tsao N, Rao A, Newman AAC, Eberhard AV, Priest JR, Ruusalepp A, Pasterkamp G,
360 Maegdefessel L, Miller CL, Lind L, Koplev S, Björkegren JLM, Owens GK, Ingelsson E,
361 Weissman IL, Leeper NJ. 2020. Clonally expanding smooth muscle cells promote
362 atherosclerosis by escaping efferocytosis and activating the complement cascade. *P Natl*
363 *Acad Sci Usa* **117**:15818–15826. doi:10.1073/pnas.2006348117

364 Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M, Coller J,
365 Koyano TK, Fong R, Woo YJ, Liu B, Montgomery SB, Wu JC, Zhu K, Chang R, Alamprese
366 M, Tallquist MD, Kim JB, Quertermous T. 2019. Atheroprotective roles of smooth muscle cell
367 phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat*
368 *Med* **25**:1280--1289. doi:10.1038/s41591-019-0512-5

369 Xu J, Nuno K, Litzenburger UM, Qi Y, Corces MR, Majeti R, Chang HY. 2019. Single-cell
370 lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial
371 DNA. *Elife* **8**:e45105. doi:10.7554/elife.45105

372

373

Figures and Figure Legends

Figure 1

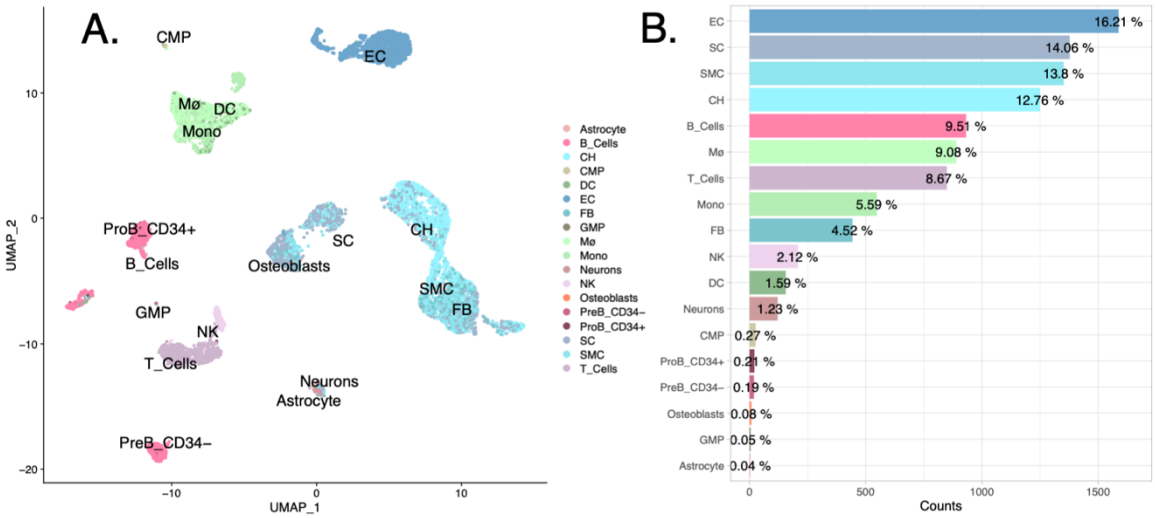
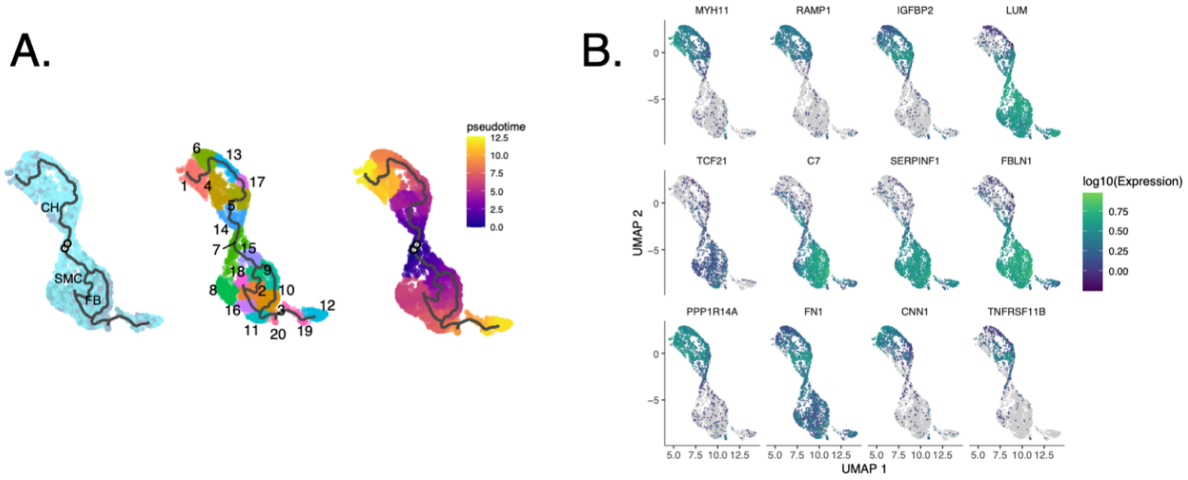


Fig. 1. Unbiased automatic cell labeling of human coronary scRNA data using 'SingleR' reveals abundant cells with chondrocyte (CH) and fibroblast (FB) gene expression patterns. (A) UMAP clustering of 9798 cells derived from human coronary artery explants. **(B)** Population breakdown by percentage. SMC: smooth muscle cells, EC: endothelial cells, CH: chondrocytes, FB: fibroblasts, Mø: macrophages, SC: stem cells, CMP: common myeloid progenitor cells, GMP: granulocyte-monocyte progenitor cells.

384 **Figure 2**

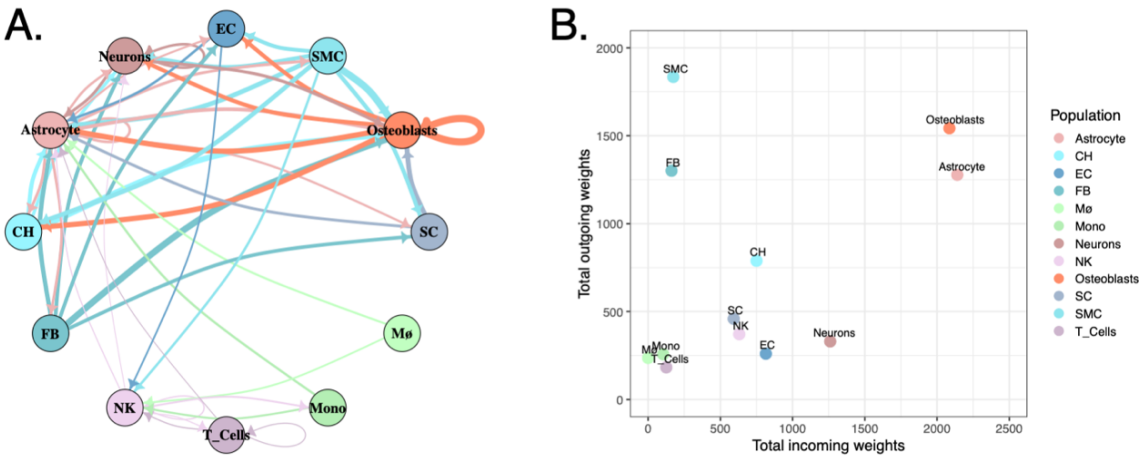


385

386 **Fig. 2. Pseudotemporal and ultra-fine clustering reveals derivation of CH and FB-like**
387 **cells from SMCs. (A) Left,** RNA trajectory (line) shows path two direct paths from the SMC
388 starting nodes (grey circles). **Middle,** ultra-fine clustering shows the logical transition stages
389 from SMC to CH and FB. **Right,** pseudotemporal analysis confirms that the cell and clusters
390 existing along a logical single-cell continuum. **(B)** Selected genes that were shown to vary over
391 pseudotime by Moran's I test were visualized.

392

393 **Figure 3**



394

395 **Fig. 3. Comprehensive ligand-receptor analysis reveals intricate intracellular**

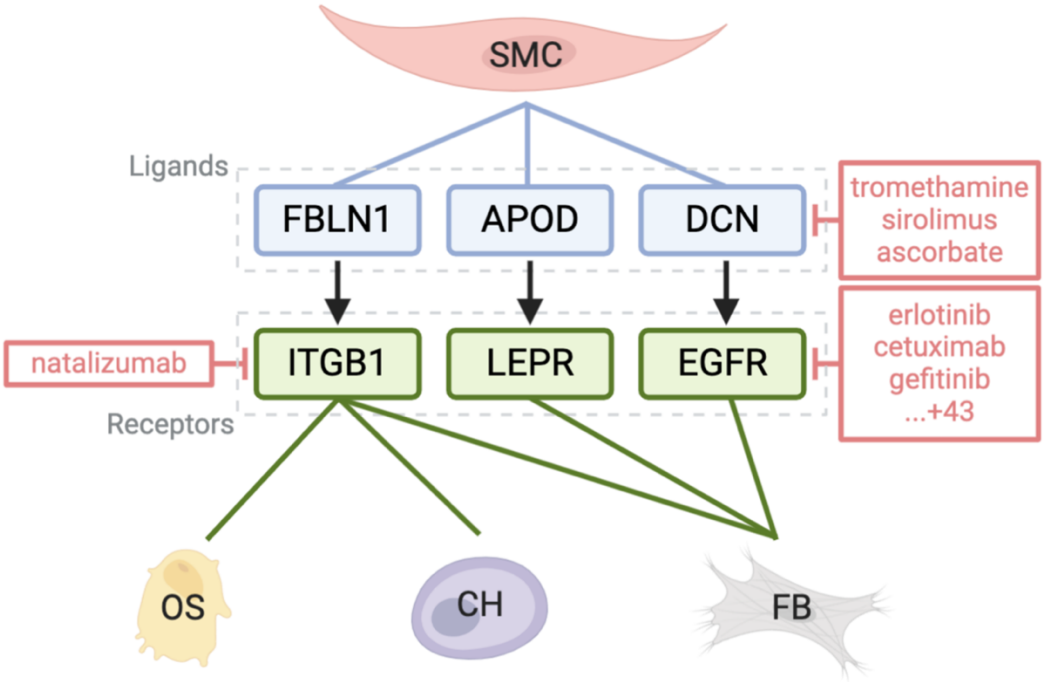
396 **communications in the coronary environment. (A)** circle plot representation of the inferred

397 intercellular communications within the coronary artery environment. **(B)** Incoming and outgoing

398 signal summed weights by cell type.

399

400 **Figure 4**



401

402 **Fig. 4. Fine ligand-receptor analysis coupled with automatic DGIdb analysis reveals**

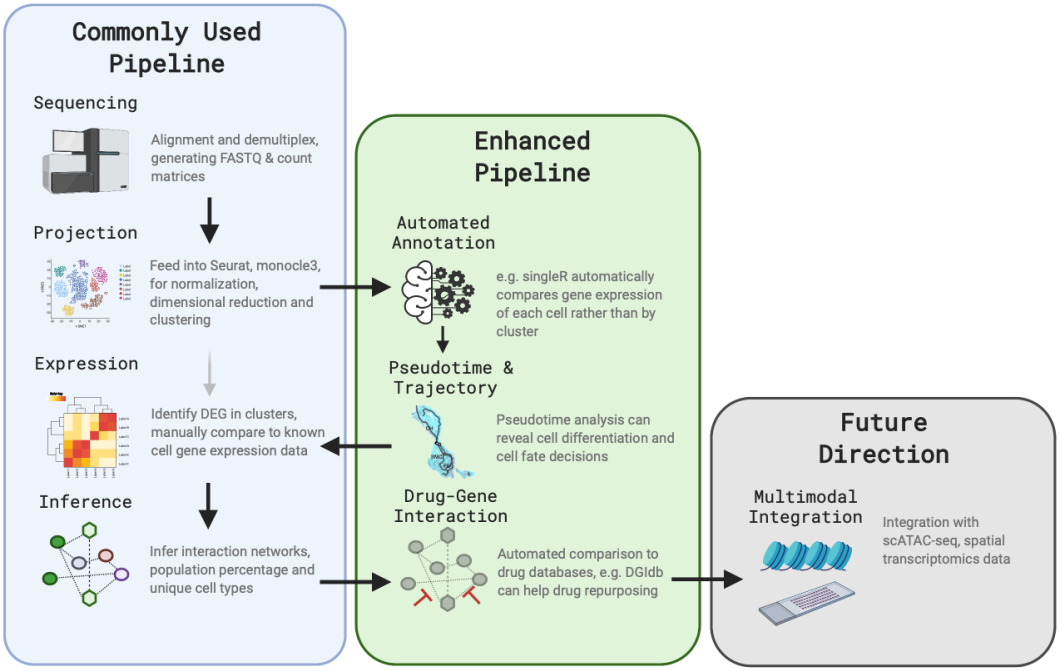
403 **potential drug therapies could interrupt SMC to CH and FB communication.** 'scTalk' shows

404 that SMCs interacts with FB through three pairs of experimentally verified interactions. Of these

405 three, *DCN-EGFR* signaling is the most druggable as revealed by DGIdb 3.0.

406

407 **Figure 5**



408

409 **Fig. 5. Roadmap of enhanced scRNA-seq analysis process.** Instead of cluster-based

410 grouping, our pipeline uses automatic cell labeling, coupled with pseudotime trajectory, cellular

411 network interaction and drug targeting, and provides a reproducible process for scRNA

412 datasets. From this roadmap, it is easy to add additional analysis tools as each cell has its own

413 identity and can be re-projected into different spaces such as t-SNE or 3D spaces.

414

Tables

Table 1.

	p_value	morans_test_statistic	morans_I
MYH11	<1E-297	169.0285919	0.66505236
IGFBP2	<1E-297	129.2299117	0.50844307
PPP1R14A	<1E-297	169.4172149	0.66663148
FBLN1	<1E-297	154.3399713	0.60730135
C7	<1E-297	156.2248703	0.61472417
LUM	<1E-297	170.4635813	0.67075391
RAMP1	<1E-297	161.486949	0.63531256
TCF21	<1E-297	75.75779399	0.29798686
SERPINF1	<1E-297	157.9946249	0.62168635
FN1	<1E-297	125.4354275	0.49354242
CNN1	<1E-297	170.9288702	0.67251583
TNFRSF11B	<1E-297	97.66515083	0.38400709

Table 1. Selected Moran's I statistics for genes listed in Figure 2. Moran's I statistics was used to identify spatial correlations within the single-cell trajectory. +1 indicates that nearby cells are perfectly similar, 0 indicates no similarity or pattern, and -1 indicates total dissimilarity (Cao et al., 2019)