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

39 **Abbreviations**

40 C7- complement component C7

41 CAD- coronary artery disease

42 CH- chondrocytes

43 CMP- common myeloid progenitor cells

44 DCN- decorin

45 DGIdb- drug-gene interaction database

46 EC- endothelial cells

47 EGFR- epidermal growth factor receptor

48 FB- fibroblasts

49 FBLN1- fibulin 1

50 GMP- granulocyte-monocyte progenitor cells.

51 Mø- macrophages

52 MYH11- myosin heavy chain 11

53 SC- stem cells

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

55 sequencing

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

57 SMC- smooth muscle cells

58 UMAP- uniform manifold approximation and projection

59

60 **Background**

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

71

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

83

84 Recently, new approaches have been developed to compensate for both of these shortcomings,
85 namely automatic cell labeling and pseudotemporal analysis. Tools such as 'SingleR' and
86 'Garnett' have been used to assign unbiased identities to individual cells using reference-based
87 and machine learning algorithms, respectively (Aran et al., 2019; Pliner et al., 2019). On the
88 other hand, tools such as 'Monocle3' and 'scVelo' align and project cells onto a pseudotemporal
89 space where each cell becomes a snapshot within the single-cell time continuum (Bergen et al.,
90 2019; Cao et al., 2019). In essence, the single scRNA-seq dataset is transformed into a time
91 series (Bergen et al., 2019; Cao et al., 2019; Manno et al., 2018). Although the pseudotemporal
92 scale does not reflect the actual time scale, it is a reliable approximation to characterize cell fate
93 and differentiation events, e.g., during organogenesis, disease states, or in response to SARS-
94 CoV-2 infection (Cao et al., 2019; Chua et al., 2020).
95
96 In this study, we present the application of an enhanced, scalable, and user-friendly scRNA-seq
97 analysis workflow on an existing human coronary artery scRNA-seq dataset. We performed
98 unbiased automatic cell identification at the single-cell level, pseudotemporal analysis, ligand-
99 receptor expression profiling, and drug repurposing analysis. Our results demonstrate potential
100 new mechanisms by which SMCs contribute to the atherosclerotic phenotype and signaling
101 within the lesion microenvironment. More importantly, we reveal attractive candidates for future
102 pre-clinical drug interventional studies. This reproducible analysis workflow can also be easily
103 modified and extended to incorporate different tissue data sources and single-cell modalities
104 such as snATAC-seq (Smith and Sheffield, 2020).
105
106 **Results and Discussion**
107 **Unbiased automatic cell labeling reveals abundant cells with chondrocyte and fibroblast**
108 **characteristics.**

109 Recently, automatic cell identifications tools have been introduced to compensate the
110 shortcomings of manual, cluster-based cell labeling (Aran et al., 2019). For example, 'SingleR'
111 and 'Garnett' uses reference-based and machine learning algorithms, respectively, to call
112 individual cell identities (Aran et al., 2019; Pliner et al., 2019). Using 'SingleR,' which uses
113 known purified cell expression data as references, we found that endothelial cells (ECs) make
114 up the highest proportion of cells in this dataset (16.21%, Fig. 1), followed by smooth muscle
115 cells (SMCs, 13.8%) and stem cells (SCs, 14.06%), where the latter could be so-called
116 "atherosclerotic stem cells" or normal stromal stem cells but cannot be distinguished until
117 specific expression profile references are developed in the future (Wang et al., 2020).
118 Consistent with recent scRNA-seq studies in atherosclerotic models, we identified abundant
119 fibroblast (FB) and chondrocyte-like (CH) cells, as well as cells with an osteoblast-like (OS)
120 expression profile (Fig. 1) (Pan et al., 2020). In the UMAP clusters reflecting single-cell
121 identities, there was a substantial presence of SMC and FB cells in the OS and SC cluster.
122 Such heterogeneity in cell clusters would have been overlooked in manual cluster-based cell
123 labeling.
124
125 **Pseudotemporal ordering identifies distinct chondrocyte and fibroblast-like cell**
126 **differentiation states from smooth muscle cells.**
127 To evaluate putative cell fate decisions or differentiation events (e.g., SMC phenotypic transition
128 states), we performed pseudotemporal analysis and ultra-fine clustering using 'Monocle3', a
129 method previously applied to normal and diseased states, e.g. embryo organogenesis and
130 response to COVID19 infection, respectively (Cao et al., 2019; Chua et al., 2020). We found
131 evidence of SMCs giving rise to CH and FB-like cells (Fig. 2A). This corroborates earlier
132 findings showing that SMCs may transition or de-differentiate into 'fibromyocytes'—SMCs that
133 have undergone a phenotypic modulation within the atherosclerotic lesions (Wang et al., 2020;

134 Wirka et al., 2019). Genes associated with healthy SMC phenotypes, such as *MYH11* (a
135 canonical marker of SMC), *IGFBP2* (associated with decreased visceral fat), and *PPP1R14A*
136 (which enhances smooth muscle contractions), are decreased by approximately 50-75% along
137 the SMC trajectory as these cells become more FB-like (Table 1, Fig. 2, $p < 0.1E-297$) (Bennett
138 et al., 2016; Carter et al., 2018). Similar results were found by another group using mouse
139 lineage-traced models where MYH11 expression was decreased in SMC-derived modulated
140 "intermediate cell states" (Pan et al., 2020). More importantly, specific inflammatory markers
141 and proteins associated with thrombotic events during CAD, including *C7* and *FBLN1*, are
142 increased along the same trajectory (Argraves et al., 2009; Carter, 2012)(Fig. 2B). Together,
143 and in corroboration of recent studies, our pseudotemporal analysis demonstrates that SMCs
144 could be a source of FB and CH-like cells, with the former associated with an intermediate
145 atherosclerotic cell phenotype, and the latter expresses genes associated with a more
146 advanced atherosclerotic phenotype (Pan et al., 2020; Wirka et al., 2019). This is further
147 supported by the recent study where blocking of SMC-derived intermediate cells coincides with
148 less severe atherosclerotic lesions (Pan et al., 2020). Precisely how these cells might influence
149 the overall stability of the atherosclerotic lesion and patient survival requires additional
150 longitudinal studies using genetic models (Bennett et al., 2016; Pan et al., 2020; Wang et al.,
151 2020).
152
153 **Comprehensive ligand-receptor analysis shows complex intercellular communications in**
154 **the human coronary micro-environment and reveals potential drug targets.**
155 To examine the potential cross-talk between different cell types using scRNA-seq data, we
156 compared the ligand and receptor expression profiles of each cell type with experimentally-
157 validated interactions using 'scTalk' (Farbehi et al., 2019). We found that there is an intricate
158 network of signaling pathways connecting different cell types; some cell types, such as OS,

159 have stronger and more frequent outgoing signals, whereas other cell types such as
160 Macrophages (Mø) have fewer and weaker incoming and outgoing signals (Fig. 3A). SMCs,
161 OSs, and neurons also exhibit a high degree of autocrine signaling profiles (Fig. 3A).
162 Specifically, SMCs are shown to have the highest number of outgoing signals and are among
163 those with the least number of incoming signal weights (Fig. 3B). This suggests that SMCs play
164 an important role in regulating the coronary microenvironment by transducing signals to
165 neighboring cells in the lesion.
166
167 Specifically, we identified three significant ligand-receptor interactions between SMCs and FBs,
168 FBLN1-ITGB1, APOD-LEPR, and DCN-EGFR (Fig. 4). We searched for potentially druggable
169 targets to interrupt SMC-FB communication by performing an integrative analysis of the
170 identified ligand-receptor interactions with known drug-gene interactions using the DGIdb 3.0
171 database (Cotto et al., 2017). Interestingly, anti-EGFR (epidermal growth factor receptor)-based
172 cancer treatments such as erlotinib, cetuximab, and gefitinib were shown as potential key
173 mediators of signaling pathways between SMC and FB via DCN (decorin) and EGFR (Fig. 4). It
174 has been shown that decorin overexpression increases SMC aggregation and SMC-induced
175 calcification at the atherosclerotic plaque (Fischer et al., 2004). Although the overlap between
176 CAD and cancer has been previously noted, the efficacy and mechanisms of chemotherapy,
177 such as erlotinib, in the pathogenesis and stability of CAD or their adjuvant use in cancer
178 patients to treat CAD continues to hold promise for future translational studies (Camaré et al.,
179 2017; DiRenzo et al., 2017; Tapia-Vieyra et al., 2017).
180
181 **Conclusions**
182 Our findings show that an enhanced, reproducible pipeline for scRNA-seq analysis improve on
183 current standard scRNA-seq bioinformatics protocols. We provide new insights into intricate

184 cellular differentiation and communication pathways while providing actionable and testable
185 targets for future studies (Fig. 5). In our combined analysis, we found that SMCs give rise to
186 substantial proportions of CH and FB, with the latter associated with worse prognostic markers
187 (Argraves et al., 2009; Carter, 2012; Carter et al., 2018; Pan et al., 2020). Several FDA-
188 approved drugs (e.g., erlotinib, cetuximab, and gefitinib) were shown as potential effectors of
189 SMCs' signaling to FB, and may be used to treat CAD in cancer patients to simplify or augment
190 drug regiments (Camaré et al., 2017). This is consistent with recent reports showing beneficial
191 effects of the acute promyelocytic drug all-trans-retinoic acid (ATRA) in atherosclerosis mouse
192 models (Pan et al., 2020).
193
194 Although the utilization of this workflow can compensate for many of the shortcomings of current
195 scRNA-seq analyses, we are still unable to perform cell-lineage tracing that reflects actual
196 timescale without additional molecular experiments. However, leveraging mitochondrial DNA
197 variants in snATAC-seq data has enabled lineage tracing analysis in human cells (Lareau et al.,
198 2020; Xu et al., 2019). Likewise, these analyses can ultimately be extended to integrate spatial
199 data and other multi-modal data (Stuart et al., 2019). In the future as spatial transcriptomics and
200 snATAC-seq data become more widely available, this workflow can be modified to discover
201 signaling pathways or differentiation events at specific tissue locations and time, allowing for
202 more disease-relevant drug-gene interaction analyses (Fig. 5). Nonetheless, this pipeline can be
203 applied immediately to datasets from other tissues/diseases to generate informative directions
204 for follow-up studies, and is more user-friendly and reproducible compared to standard scRNA
205 analyses.
206
207 **Methods**
208 **Data retrieval and pre-processing**

209 Human coronary artery scRNA data read count matrix was retrieved from the Gene Expression
210 Omnibus (GEO) using #GSE131780 and loaded into R 4.0, and was preprocessed using
211 standard parameters of the R packages 'Seurat' v.3, and 'Monocle3' as required (Satija et al.,
212 2015; Team, 2020; Trapnell et al., 2014; Wirka et al., 2019). Uniform manifold approximation
213 projections (UMAP) clusters from 'Seurat' were imported into 'Monocle3' before pseudotemporal
214 analysis.
215

216 **Automatic cell Identification and pseudotemporal ordering**

217 scRNA read matrices were read into SingleR as previously described for cell labeling (Aran et
218 al., 2019). Briefly, SingleR compares each cell's gene expression profile with known human
219 primary cell atlas data and gives the most likely cell identity independently. SingleR first corrects
220 for batch effects, then calculates the expression correlation scores for each test cell to each cell
221 type in the reference, and the cell identity is called based on reference cell type exhibiting the
222 highest correlation. Then, pseudotemporal analyses were performed as previously described in
223 the analysis of embryo organogenesis (Cao et al., 2019; Trapnell et al., 2014). Briefly, the
224 UMAP clusters were passed into Monocle3 and then the 'learn_graph()' and 'order_cells()'
225 functions. The SMCs and related clusters were then subsetted for detailed subclustering and
226 analysis. For each cluster, Moran's I statistics were calculated, which identify genes that are
227 differentially expressed along their trajectories. Detailed codes to reproduce the figures in this
228 publication can be found at the Miller Lab Github (see Availability of data and materials).
229

230 **Ligand-receptor cell communication analysis**

231 We analyzed candidate ligand-receptor interactions to infer cell communication using the R
232 package 'scTalk', as previously described in the analysis of glial cells (Farbehi et al., 2019).
233 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

284 reveals a transitional profibrotic macrophage. *Nat Immunol* **20**:163–172. doi:10.1038/s41590-
285 018-0276-y

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

289 Basatemur GL, Jørgensen HF, Clarke MCH, Bennett MR, Mallat Z. 2019. Vascular smooth
290 muscle cells in atherosclerosis. *Nat Rev Cardiol* **16**:727–744. doi:10.1038/s41569-019-0227-
291 9

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

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

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

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

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

303 Carter S, Lemieux I, Li Z, Alméras N, Tremblay A, Bergeron J, Poirier P, Després J-P, Picard F.
304 2018. Changes in IGFBP-2 levels following a one-year lifestyle modification program are
305 independently related to improvements in plasma apo B and LDL apo B levels.
306 *Atherosclerosis* **281**:89–97. doi:10.1016/j.atherosclerosis.2018.12.016

307 Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, Debnath O, Thürmann L,
308 Kurth F, Völker MT, Kazmierski J, Timmermann B, Twardziok S, Schneider S, Machleidt F,
309 Müller-Redetzky H, Maier M, Krannich A, Schmidt S, Balzer F, Liebig J, Loske J, Suttorp N,
310 Eils J, Ishaque N, Liebert UG, Kalle C von, Hocke A, Witzenrath M, Goffinet C, Drosten C,
311 Laudi S, Lehmann I, Conrad C, Sander L-E, Eils R. 2020. COVID-19 severity correlates with
312 airway epithelium–immune cell interactions identified by single-cell analysis. *Nat Biotechnol*
313 1–10. doi:10.1038/s41587-020-0602-4

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

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

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

323 Fischer JW, Steitz SA, Johnson PY, Burke A, Kolodgie F, Virmani R, Giachelli C, Wight TN.
324 2004. Decorin Promotes Aortic Smooth Muscle Cell Calcification and Colocalizes to Calcified
325 Regions in Human Atherosclerotic Lesions. *Arteriosclerosis Thrombosis Vasc Biology*
326 **24**:2391–2396. doi:10.1161/01.atv.0000147029.63303.28

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

331 Manno GL, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti
332 ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, Bruggen D van, Guo J, He X, Barker R,
333 Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV.
334 2018. RNA velocity of single cells. *Nature* **560**:494--498. doi:10.1038/s41586-018-0414-6

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

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

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

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

346 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M,
347 Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-
348 1902.e21. doi:10.1016/j.cell.2019.05.031

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

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

354 Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ,
355 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

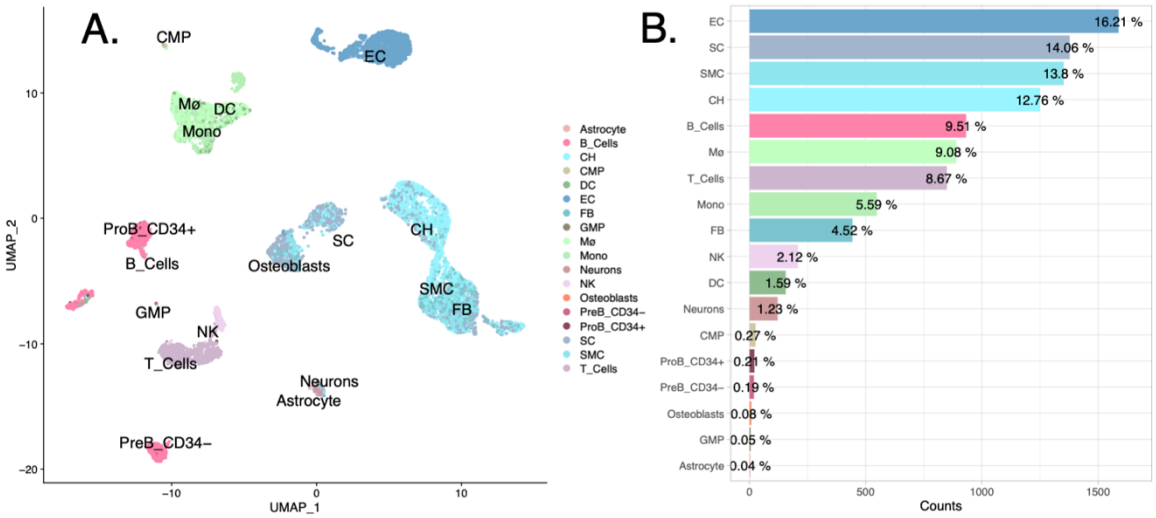
372

373

374

375 **Figures and Figure Legends**

376 **Figure 1**



377

378 **Fig. 1. Unbiased automatic cell labeling of human coronary scRNA data using 'SingleR'**

379 **reveals abundant cells with chondrocyte (CH) and fibroblast (FB) gene expression**

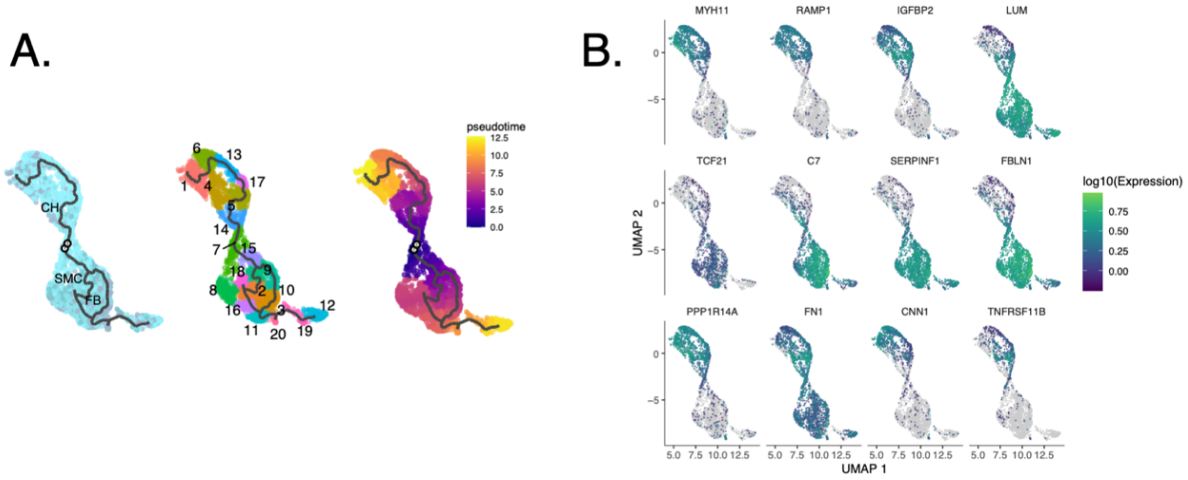
380 **patterns. (A)** UMAP clustering of 9798 cells derived from human coronary artery explants. **(B)**

381 Population breakdown by percentage. SMC: smooth muscle cells, EC: endothelial cells, CH:

382 chondrocytes, FB: fibroblasts, Mø: macrophages, SC: stem cells, CMP: common myeloid

383 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

Figure 3

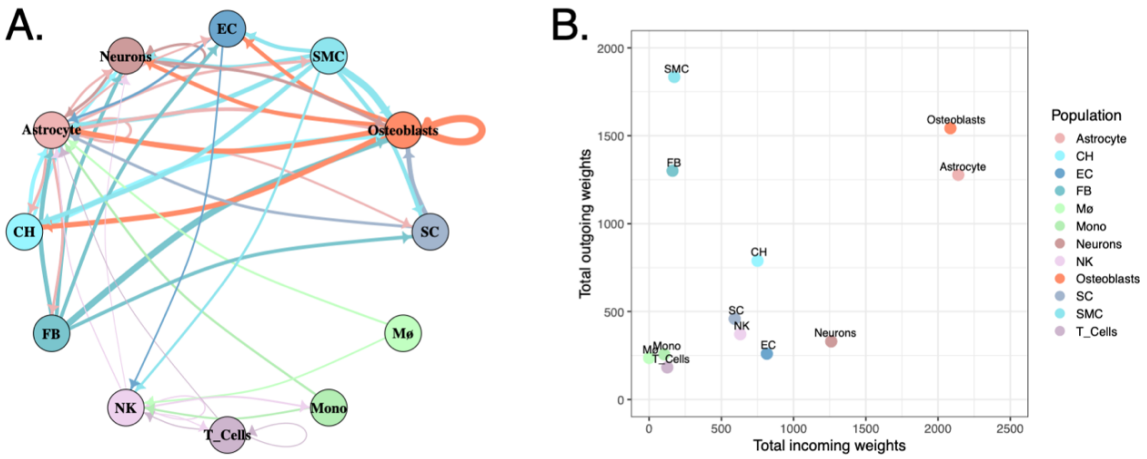
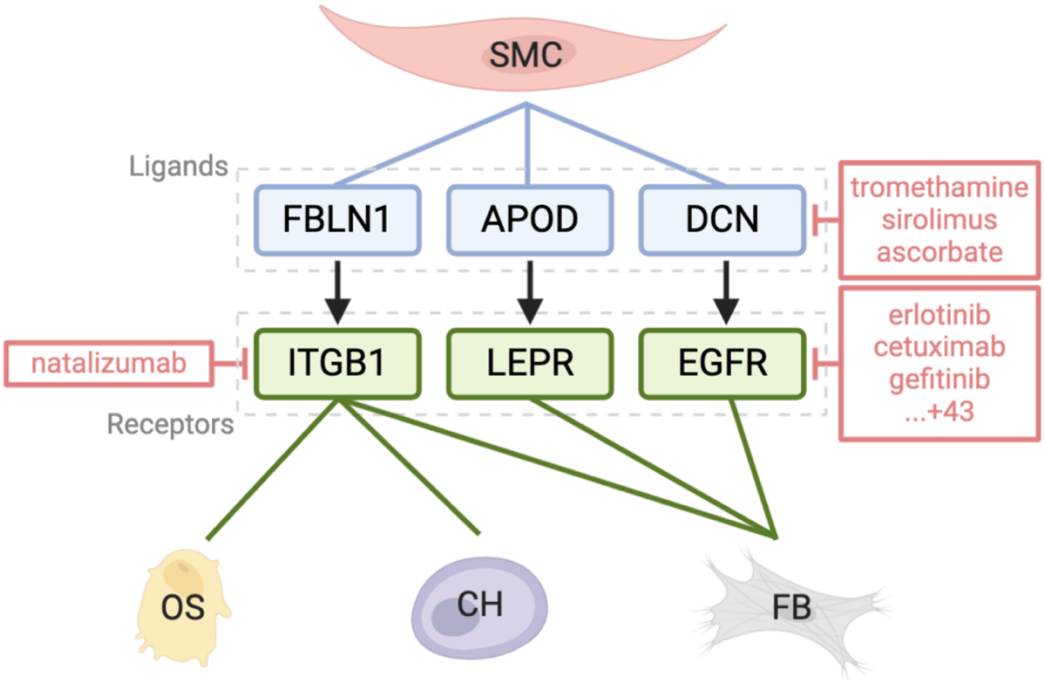


Fig. 3. Comprehensive ligand-receptor analysis reveals intricate intracellular communications in the coronary environment. (A) circle plot representation of the inferred intercellular communications within the coronary artery environment. **(B)** Incoming and outgoing signal summed weights by cell type.

400 **Figure 4**



401

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

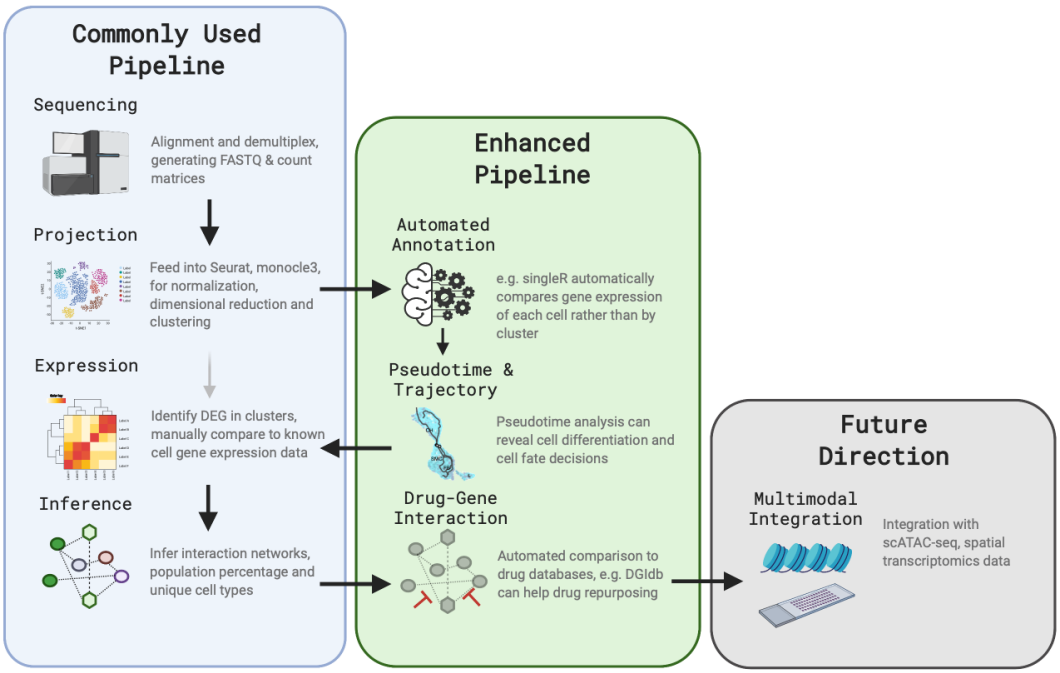
403 **potential drug therapies could interrupt SMC to CH and FB communication.** 'scTalk' shows

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

415 **Tables**

416 **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

417

418 **Table 1. Selected Moran's I statistics for genes listed in Figure 2.** Moran's I statistics was

419 used to identify spatial correlations within the single-cell trajectory. +1 indicates that nearby cells

420 are perfectly similar, 0 indicates no similarity or pattern, and -1 indicates total dissimilarity (Cao

421 et al., 2019)