SUPPLEMENTAL MATERIALS

Aortic Cellular Diversity and Quantitative GWAS Trait Prioritization through Single Nuclear RNA Sequencing (snRNA-Seq) of the Aneurysmal Human Aorta

Elizabeth L. Chou, MD[1,3,4†], Mark Chaffin, MS[4,5†], Bridget Simonson, PhD[4,5†], James P. Pirruccello, MD[2-6], Amer-Denis Akkad, PhD[7], Mahan Nekoui, MD[4,6], Christian Lacks Lino Cardenas, PharmD, PhD[2-3], Kenneth C. Bedi Jr., BS[8], Craig Nash, PhD[4,5], Dejan Juric, MD[9], James R. Stone, MD, PhD[10], Eric M. Isselbacher, MD[2-3,11], Kenneth B. Margulies, MD[8], Carla Klattenhoff, PhD[7], Patrick T. Ellinor, MD, PhD[2-6††], and Mark E. Lindsay, MD, PhD[2-4,6,11††]

1. Division of Vascular and Endovascular Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA
2. Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts, USA
3. Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA
4. Cardiovascular Disease Initiative, Broad Institute, Cambridge, Massachusetts, USA
5. Precision Cardiology Laboratory, The Broad Institute, Cambridge, MA, USA 02142
6. Demoulas Center for Cardiac Arrhythmias, Massachusetts General Hospital, Boston, Massachusetts, USA
7. Precision Cardiology Laboratory, Bayer US LLC, Cambridge, MA, USA 02142
8. Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA 19104
9. Cancer Center, Massachusetts General Hospital, Boston, Massachusetts, USA
10. Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA
11. Thoracic Aortic Center, Massachusetts General Hospital, Boston, Massachusetts, USA

Major Resources Table

In order to allow validation and replication of experiments, all essential research materials listed in the Methods should be included in the Major Resources Table below. Authors are encouraged to use public repositories for protocols, data, code, and other materials and provide persistent identifiers and/or links to repositories when available. Authors may add or delete rows as needed.

Animals (in vivo studies)

Species	Vendor or Source	Background Strain	Sex	Persistent ID / URL
N/A				

Genetically Modified Animals

	Species	Vendor or Source	Background Strain	Other Information	Persistent ID / URL
Parent - Male	N/A				
Parent - Female					

Antibodies

Target antigen	Vendor or Source	Catalog #	Working concentration	Lot # (preferred but not required)	Persistent ID / URL
N/A					

DNA/cDNA Clones

Clone Name	Sequence	Source / Repository	Persistent ID / URL
N/A			

Cultured Cells

Name	Vendor or Source	Sex (F, M, or unknown)	Persistent ID / URL
N/A			

Data & Code Availability

Description	Source / Repository	Persistent ID / URL
Single nucleus RNA sequencing data	Broad Institute	(https://singlecell.broadinstitute.org/single_cell).
Raw dataset	NCBI GEO	(accession #GSE207784)

Other

Description	Source / Repository	Persistent ID / URL
RNAscope 2.5 HD Duplex kit: cat no 322430	ACD bio	https://acdbio.com/store/catalog/product/view/id/6/
Hs-DCN-C2: cat no 589521-C2	ACD bio	https://acdbio.com/search/site/%252A589521-C2%252A/cms/probes
Hs-ITGA8-C2: cat no 417951-C2	ACD bio	https://acdbio.com/search/site/%252A417951-C2%252A/cms/probes
Hs-PLIN2: cat no 508741	ACD bio	https://acdbio.com/search/site/%252A508741%252A/cms/probes/platform/manual-assay-rnascope
Hs-ERBB4: cat no 407831	ACD bio	https://acdbio.com/search/site/%252A407831%252A/cms/probes/channel/1
Hs-ADAMTS4: cat no 537341	ACD bio	https://acdbio.com/search/site/%252A537341%252A/cms/probes/platform/manual-assay-rnascope/channel/1
Hs-CNTN4: cat no 438491	ACD bio	https://acdbio.com/search/site/%252A438491%252A/cms/probes
Hs-CFH: cat no 428731	ACD bio	https://acdbio.com/search/site/%252A428731%252A/cms/probes/channel/1
Hs-CD163: cat no 417061	ACD bio	https://acdbio.com/search/site/%252A417061%252A/cms/probes/platform/manual-assay-rnascope/channel/1
Hs-JUN: cat no 470541	ACD bio	https://acdbio.com/search/site/%252A470541%252A/cms/probes/platform/manual-assay-rnascope/channel/1
Hs-LTBP4: cat no 893621	ACD bio	https://acdbio.com/search/site/%252A893621%252A/cms/probes
Hs-LRP1-C2: cat no 836621-C2	ACD bio	https://acdbio.com/search/site/%252A836621-C2%252A/cms/probes

Supplemental Table 1. Patient Characteristics for Aortic Samples (n=13)

	Ascending Thoracic Aortic Aneurysm (ATAA)						Control Ascending Aorta						
	1	2	3	4	5	6	C1	C2	C3	C4	C5	C6	C7
Study ID	608	718	774	C2	C3	C6	Ao6	Ao12	Ao13	1817	1820	1825	1827
Sex	F	M	F	M	M	F	F	M	F	M	M	F	F
Ethnicity	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	African American	Non-Hispanic	Non-Hispanic
Race	White	White	White	White	White	White	White	White	White	White	Black	White	White
Age, y	56	55	63	55	53	69	63	53	71	52	64	40	58
Diagnosis	ATAA, DTAA	ATAA, root aneurysm	ATAA, root aneurysm	ATAA, root aneurysm	ATAA	ATAA	Metastatic renal cell carcinoma	Lung adenocarcinoma	Metastatic thyroid carcinoma	Drug Intoxication	CVA/Stroke	Drug Intoxication	CVA/Stroke
Asc aortic diam (cm)	5.2	5.5	5.2	5	4.6	6	3	3.2	3.9	“WNL” (<3.8)	“WNL” (<3.8)	2.7	2.6
Smoking status	Past	Never	Past	Never	Never	Past	Never	Never	Never	40 pack years	No	10 pack years	39 pack years
Diabetes	No	No	No	No	No	No	No	No	No	No	Yes	No	No
HTN	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	No	Yes
COPD	No	No	No	No	No	Yes	No	No	No	No	No	No	Yes
AVR	Yes	No	Yes	Yes	No	Yes	No	No	No	No	No	No	No
Bicuspid aortic valve	No	No	No	No	No	No	No	No	No	No	No	No	No
Reoperation	No	No	No	No	No	No	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Supplemental Table 2. Sample Cell Quality Control (excel)

Supplemental Table 3. Global Marker Genes (excel)

Supplemental Table 4. Differentially Expressed Genes Between Aneurysm and Control Aortic Tissue (excel)

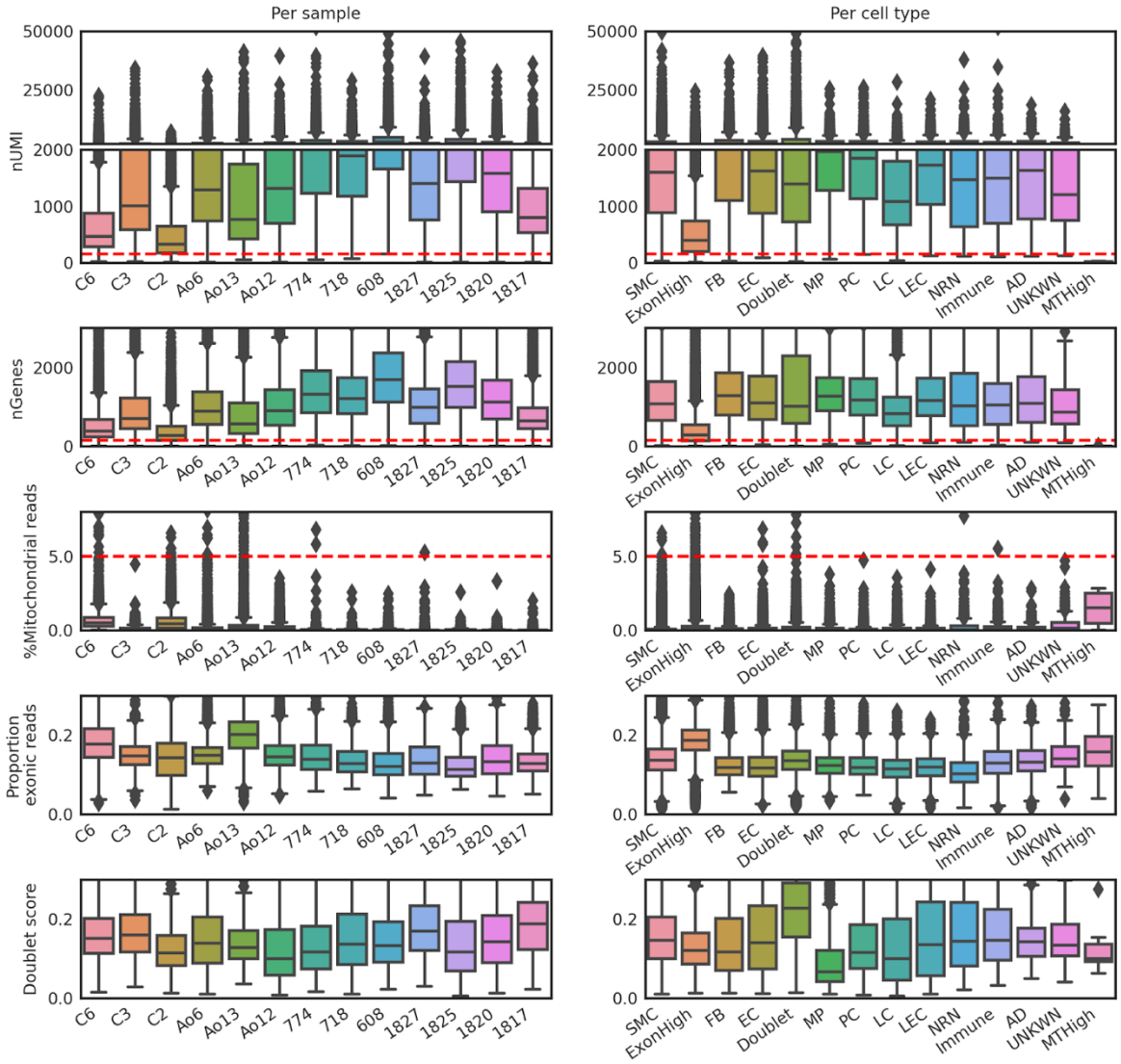
Supplemental Table 5. Gene Ontology Analysis of Differentially Expressed Genes (excel)

Supplemental Table 6. Marker Genes of VSMC Sub-clusters (excel)

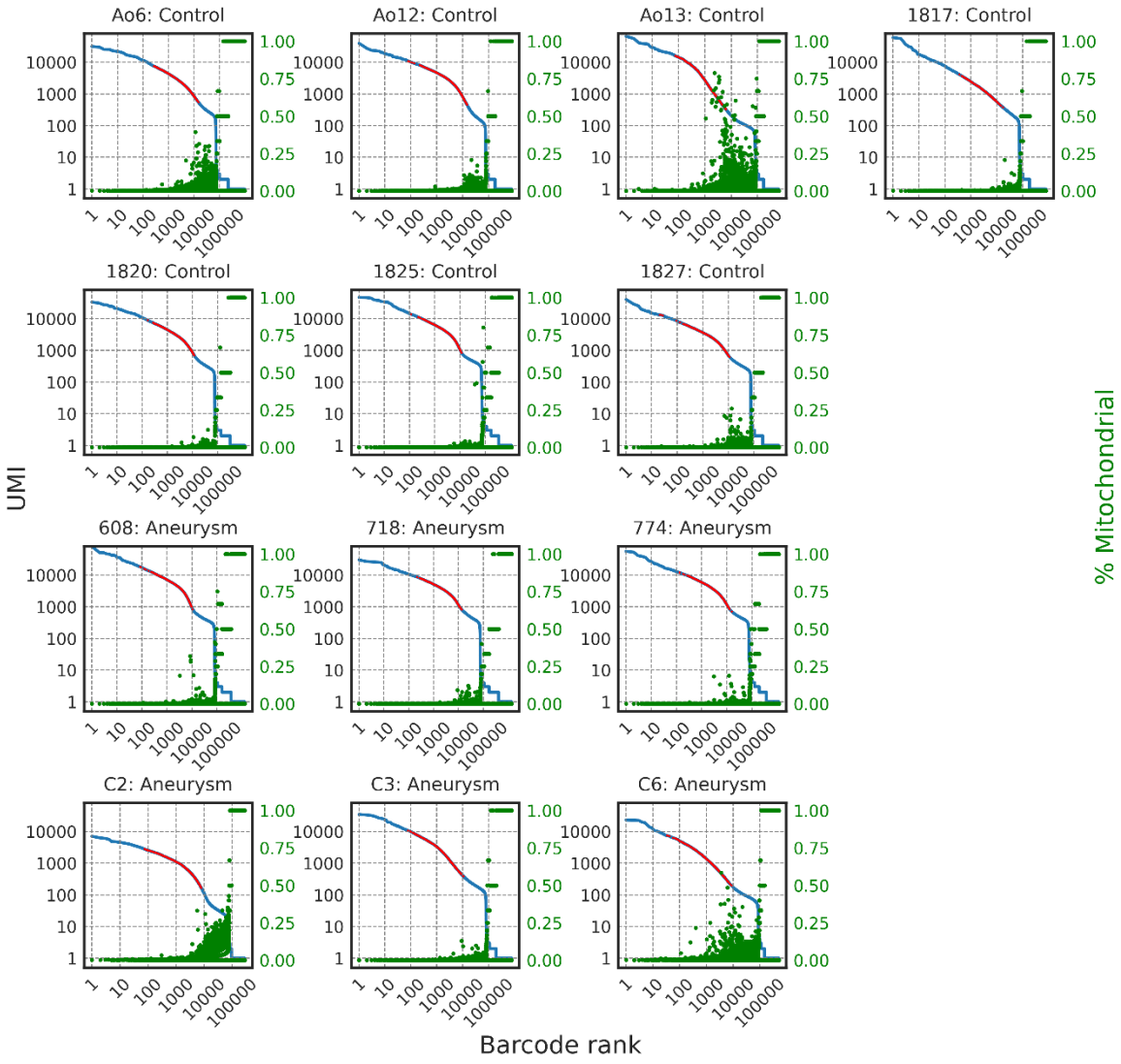
Supplemental Table 7. VSMC Sub-cluster Gene Ontology (excel)

Supplemental Table 8. Marker Genes of Fibroblast Sub-clusters (excel)

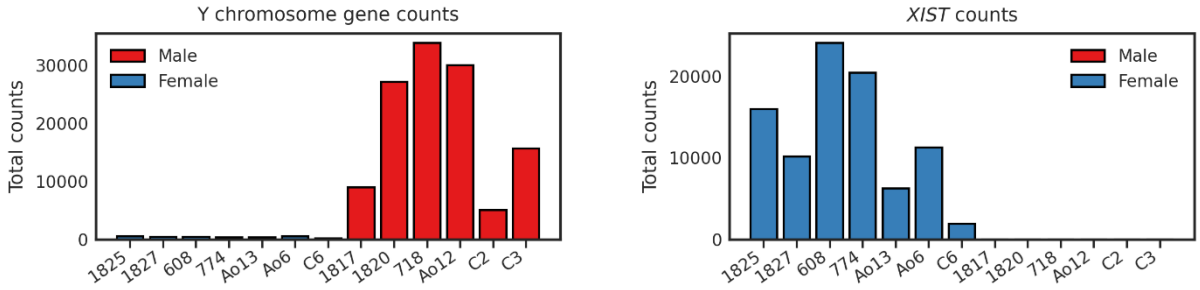
Supplemental Figure 1. snRNAseq quality control metrics. The distribution of quality control metrics across 123,578 non-empty droplets prior to quality control are shown. Quality control metrics include the total number of unique molecular identifiers per nucleus (“nUMI”), the total number of unique genes detected per nucleus (“nGenes”), the percent of reads mapping to mitochondrial genes (“%Mitochondrial reads”), the proportion of reads mapping exclusively to exons (“Proportion of exonic reads”), and the Scrublet estimate doublet score (“Doublet score”). Distributions are split by sample (left) and by crude cell type identities (right). For nUMI, y-axis is split at nUMI=2000 for clarity. For other metrics, the y-axis is truncated to focus on the majority of the distribution. Red lines indicate the hard cutoff for the metric, when a hard cutoff is employed. SMC, Smooth muscle cell; ExonHigh, Low quality clusters with elevated proportions of exonic mapping reads; FB, Fibroblast; EC, Endothelial cell; MP, Macrophage; PC, Pericyte; LC, Lymphocyte; LEC, Lymphatic endothelial cell; NRN, Neuronal; AD, Adipocyte; UNKWN, Unknown; MTHigh, Low quality clusters with elevated levels of mitochondrial mapping reads.



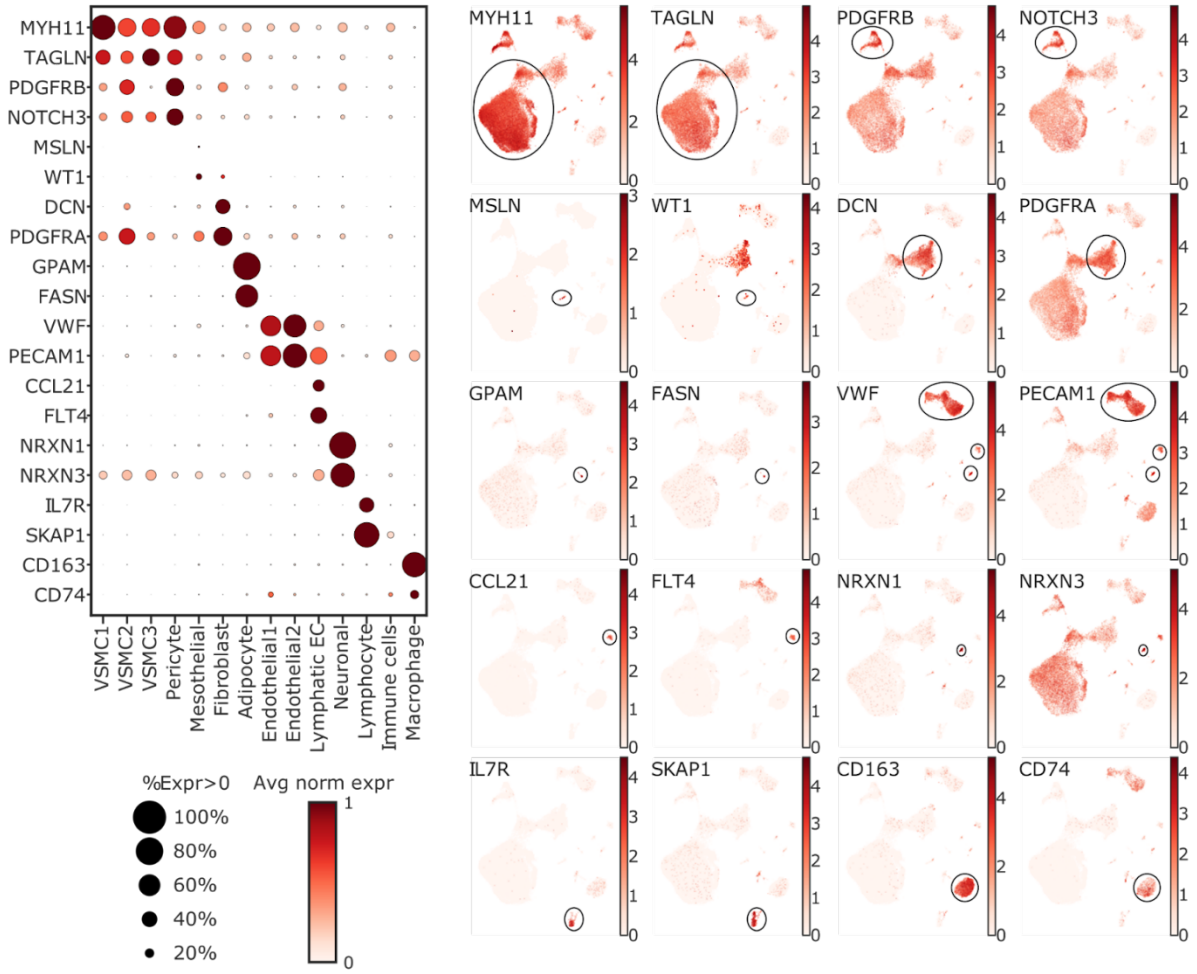
Supplemental Figure 2 Cell barcode unique molecular identifier (UMI) curves for each patient included in the analysis. Each droplet barcode is ranked by the total UMI observed in the given droplet. Red dots represent droplets retained in the analysis. Green dots represent the proportion of reads mapping to mitochondrial genes for each droplet barcode.



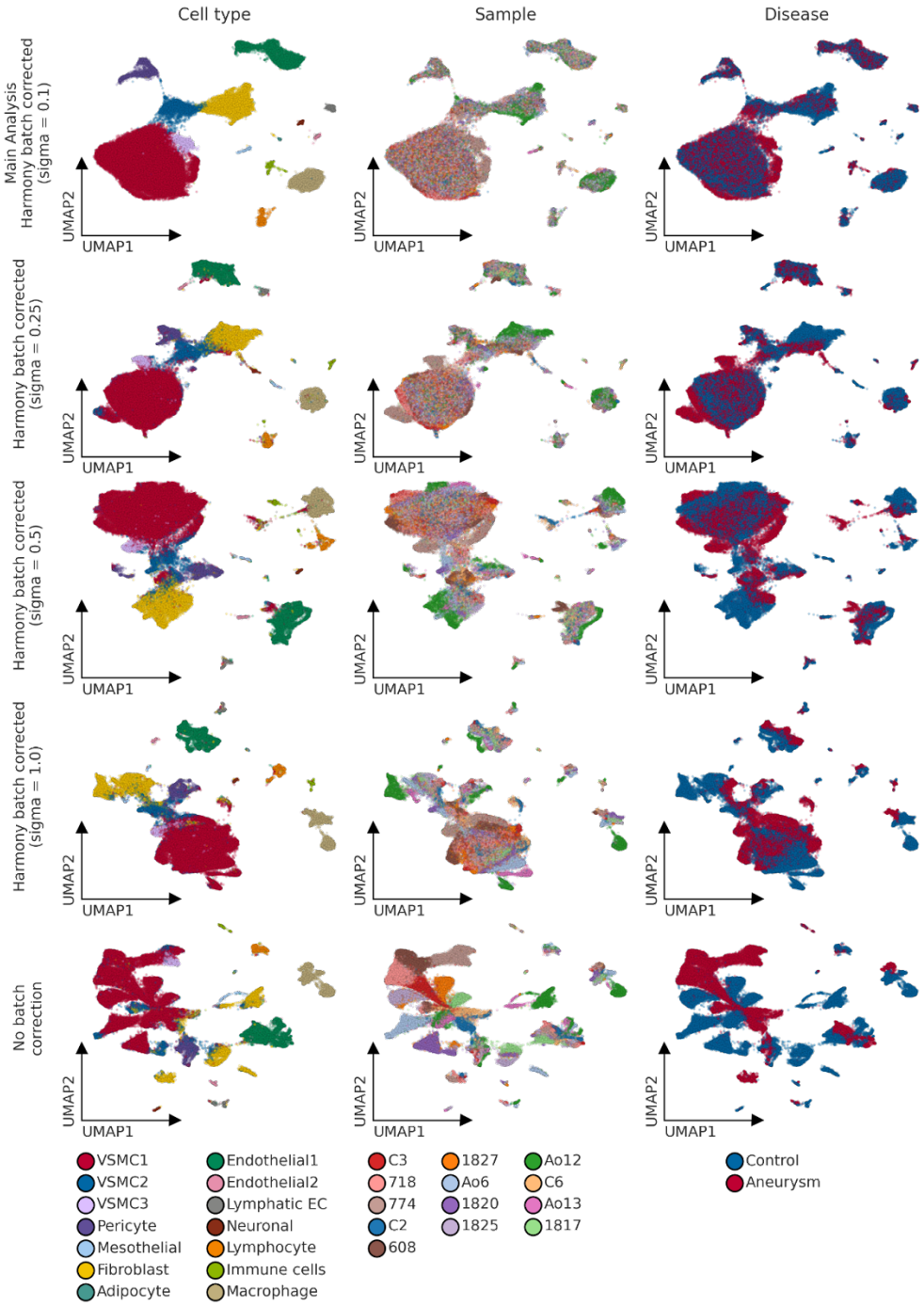
Supplemental Figure 3. Total counts for Y chromosome genes (left) and *XIST* (right) in each patient included in the analysis. Each patient is colored by the phenotypic sex.



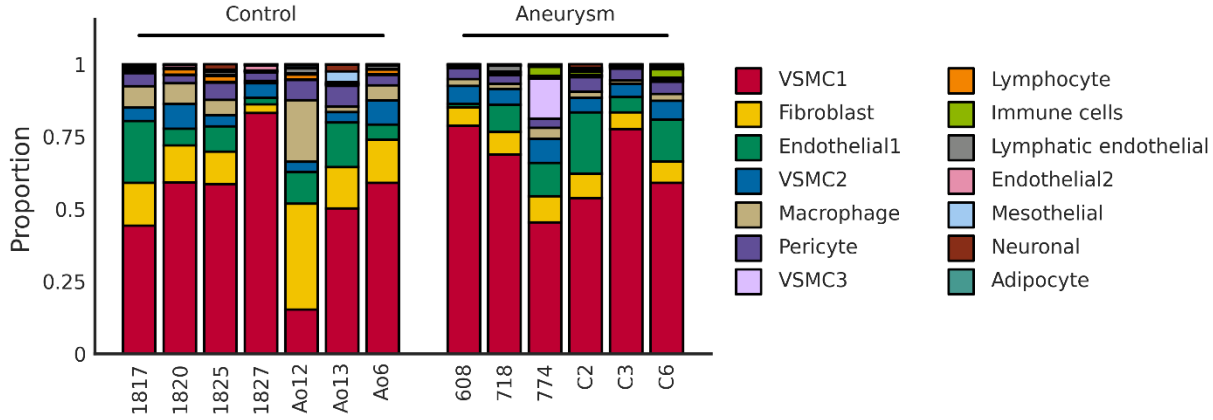
Supplemental Figure 4. Expression of canonical cell type marker genes. Dot plot (left) showing the expression of canonical markers for smooth muscle cells (*MYH11*, *TAGLN*), pericytes (*PDGFRB*, *NOTCH3*), mesothelial cells (*MSLN*, *WT1*), fibroblasts (*DCN*, *PDGFRA*), adipocytes (*GPAM*, *FASN*), endothelial cells (*VWF*, *PECAM1*), lymphatic endothelial cells (*CCL21*, *FLT4*), neuronal cells (*NRXN1*, *NRXN3*), lymphocytes (*IL7R*, *SKAP1*), and macrophages (*CD163*, *CD74*). The size of the dot represents the proportion of nuclei expressing the gene in the given cell type, and the shade represents the average expression of the gene in the given cell type, normalized across cell types. Feature plots (right) showing the overlap of the log normalized expression of each gene with the UMAP visualization. VSMC, Vascular smooth muscle cell; EC, Endothelial cell; %Expr>0, Percent of nuclei expressing gene; Avg norm expr, Average log normalized expression scaled across cell types.



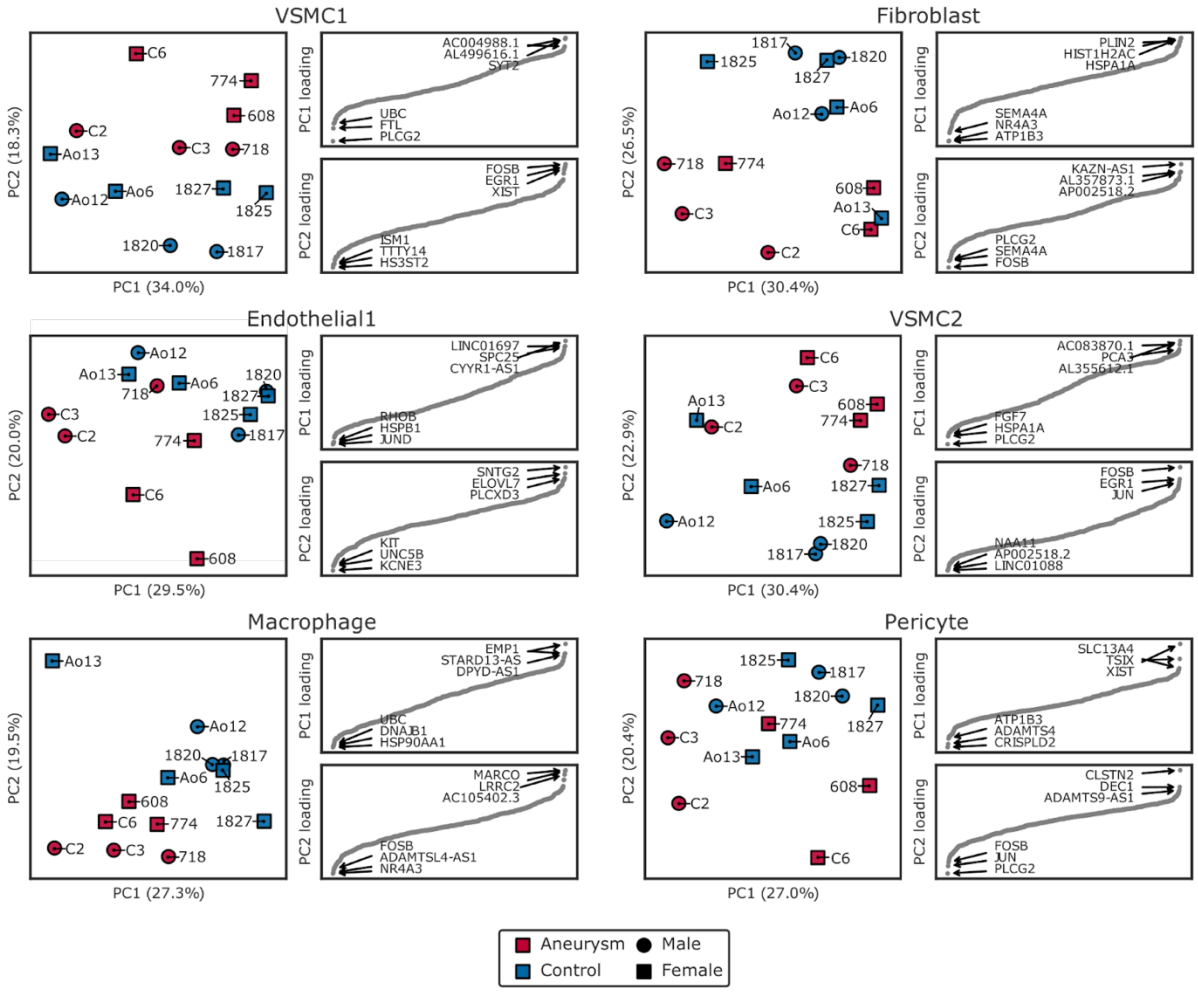
Supplemental Figure 5. Sample alignment in snRNAseq map construction. UMAP visualizations showing the overlay of cell type as derived from the primary map (left), sample (middle), and disease status (right). Each row represents a different degree of sample alignment with *Harmony*. The top row is the primary map presented in the main analysis, the next three rows represent various degrees of batch correction using *Harmony* by adjusting the hyperparameter σ (0.25, 0.5, 1.0), and the bottom row is the map with no sample alignment. VSMC, Vascular smooth muscle cell; EC, Endothelial cell.



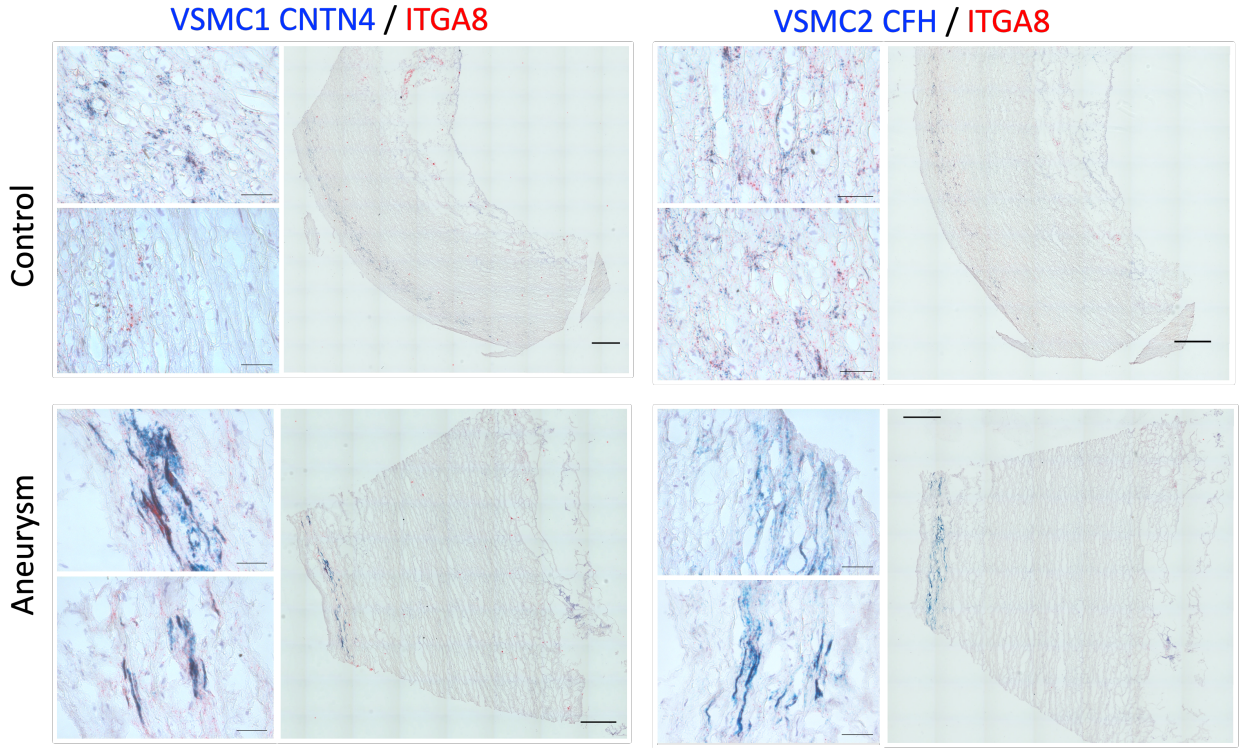
Supplemental Figure 6. Stacked bar plots demonstrating the proportion of each cell type found in each sample. VSMC, Vascular smooth muscle cell.



Supplemental Figure 7. Sample level principal component analysis by cell type. For each cell type, principal component (PC) analysis was performed on the aggregation of all nuclei per sample. Plots of PC1 and PC2 are shown with colors representing disease status and shape representing sex. Gene loadings for the 500 highly variable genes used in the analysis of each cell type are shown for each PC, labeling the top 3 most up- and down-weighted genes. VSMC, Vascular smooth muscle cell.

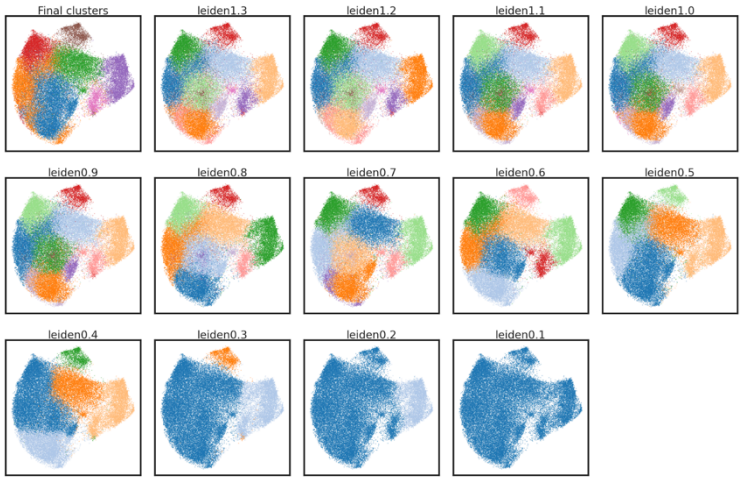


Supplemental Figure 8. RNA labeling of aortic tissue using RNAscope in situ hybridization. Control and aneurysm aortic tissue labeled with ITGA8 (red, global VSMC marker), CNTN4 (blue, VSMC1 marker), CFH (blue, VSMC2 marker). Identified VSMC1 and VSMC2 populations appear blue. Images show increased VSMC1 in control tissue and increased VSMC2 in aneurysm tissue, both in medial layer. 40x single images = 200 um; 20x tiled images = 1000 um

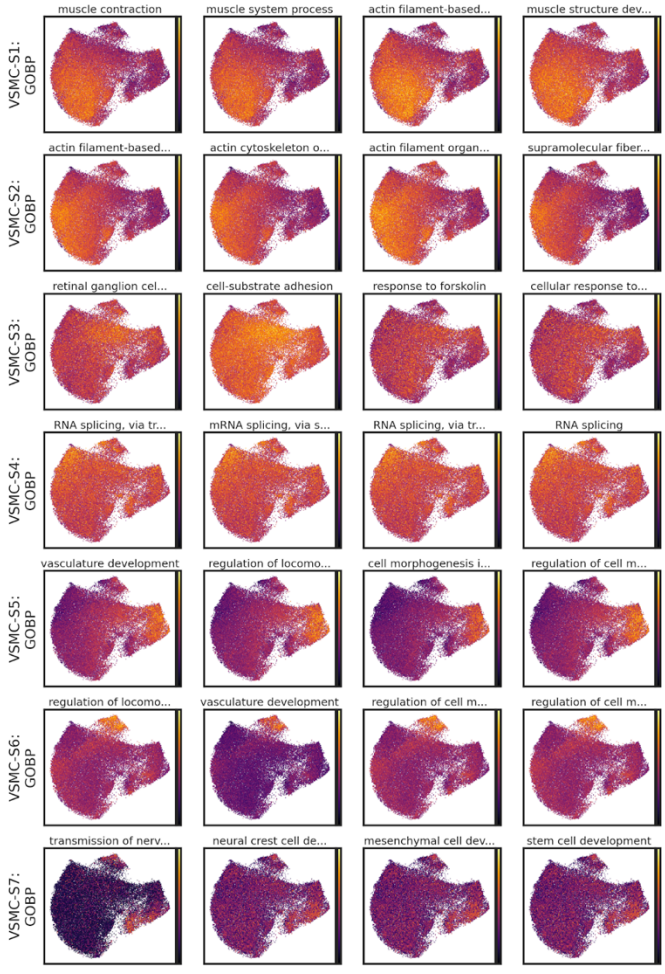


Supplemental Figure 9. a. VSMC sub-clustering at various resolutions. UMAP visualizations of VSMC sub-clustering using the Leiden algorithm at various resolutions. Top left represents the final sub-clustering analysis used in the manuscript. **b. Gene ontology biological process scoring for VSMCs.** Sub-cluster marker genes identified as driving the top four gene ontology enrichments (see **Figure 5f**) for each sub-cluster of VSMCs were scored collectively across all VSMC nuclei using the algorithm implemented as *sc.tl.score_genes()* in *scanpy*. Relative scoring of each set of genes is overlaid on UMAP visualizations of all VSMC nuclei to show localization of gene set enrichment. Some ontologies may be shown for multiple sub-clusters, but only the significant marker genes driving the enrichment for that sub-cluster (labeled left) are used in the scoring procedure. VSMC, Vascular smooth muscle cell; GOBP, Gene ontology biological process.

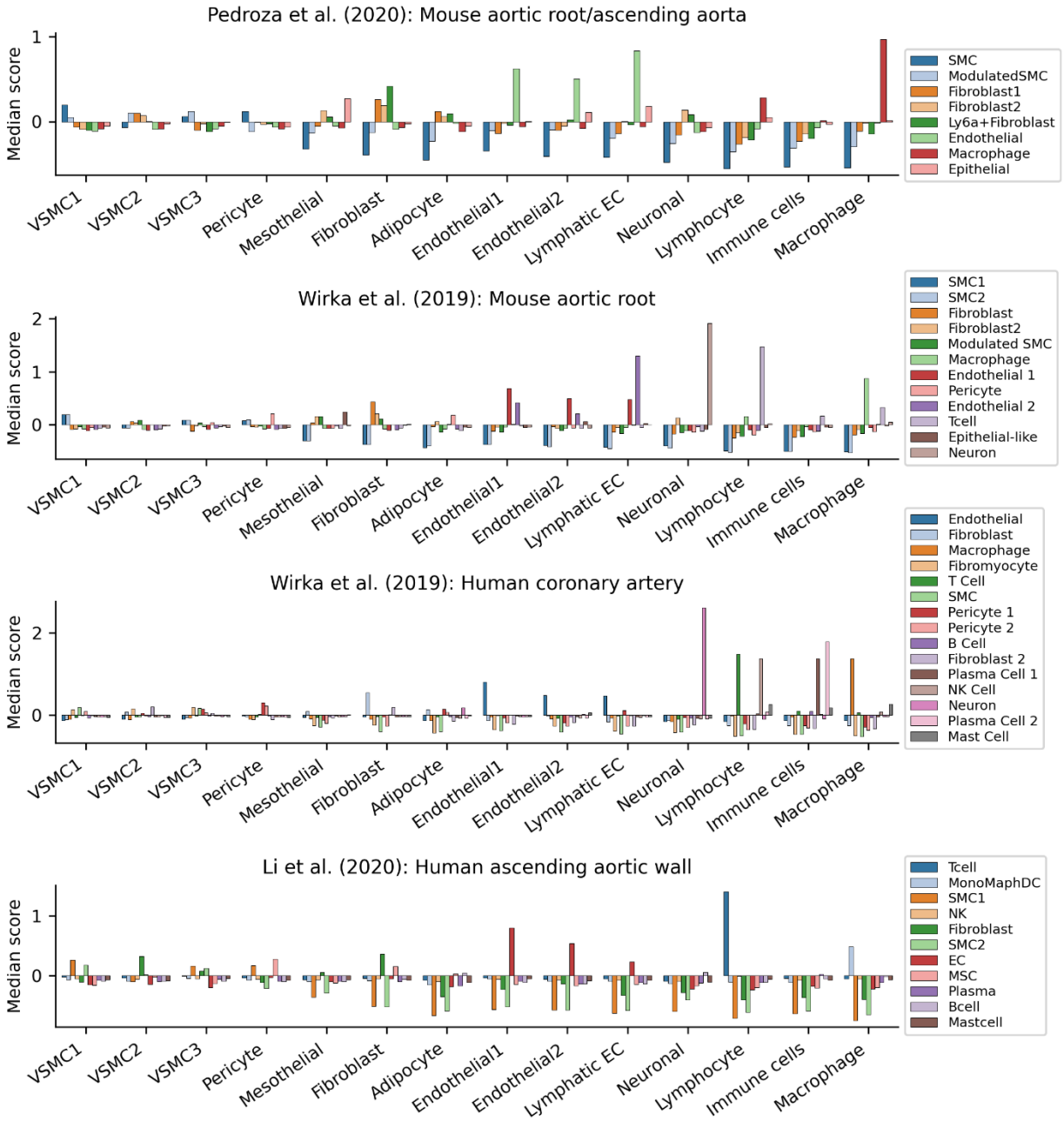
a.



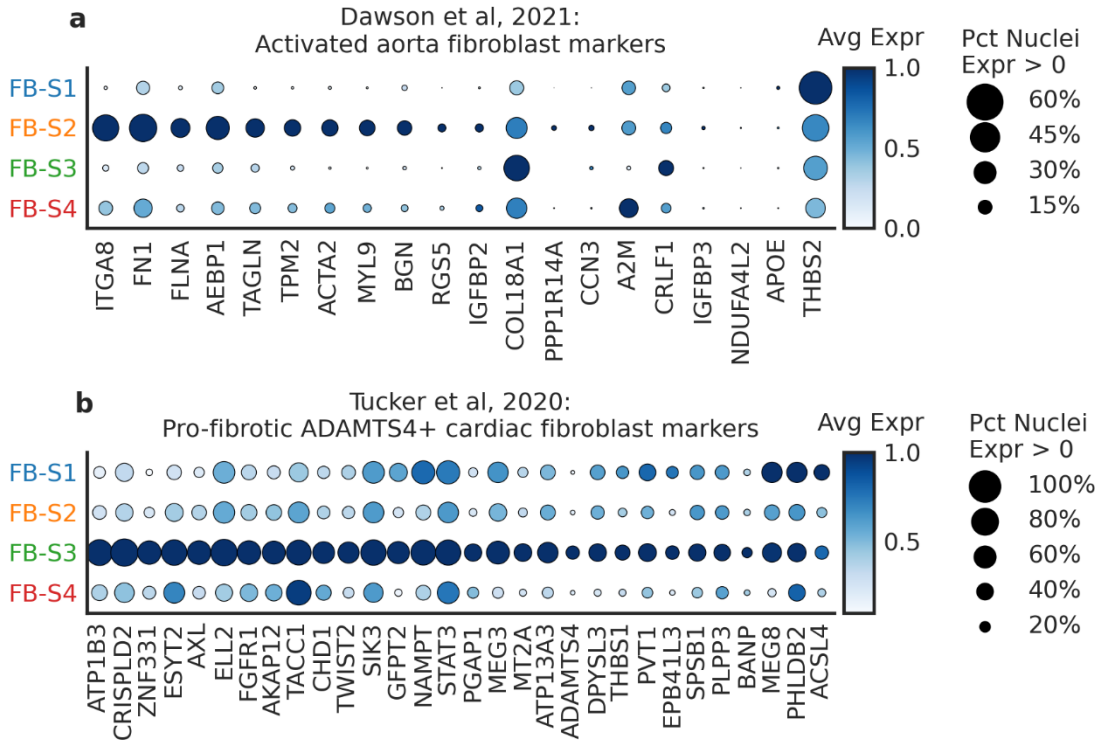
b.



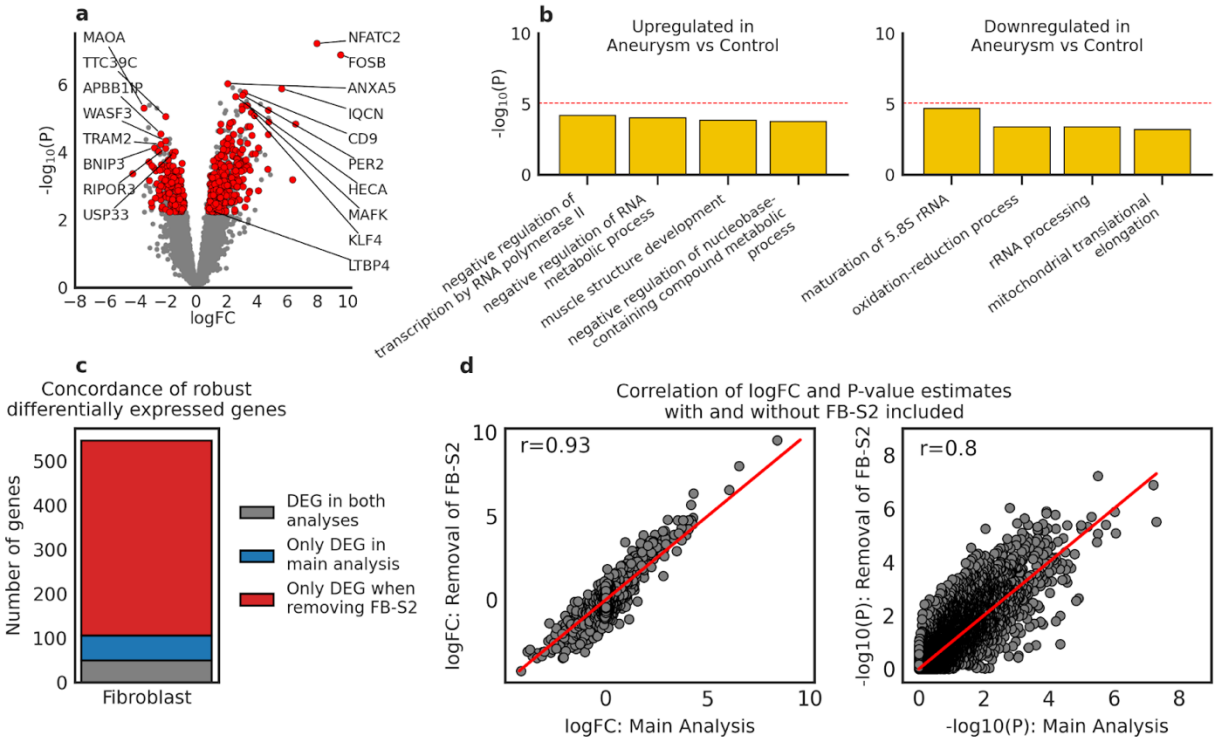
Supplemental Figure 10. Scoring of public aorta single cell RNA-sequencing cell type expression profiles identified clusters using the algorithm *sc.tl.score_genes*() as implemented in scanpy. The median score for each public cell type is shown across all nuclei from each cell type in our dataset. SMC, smooth muscle cell; MonoMaphDC, combination of monocytes, macrophages, and dendritic cells; NK, natural killer; EC, endothelial cell.



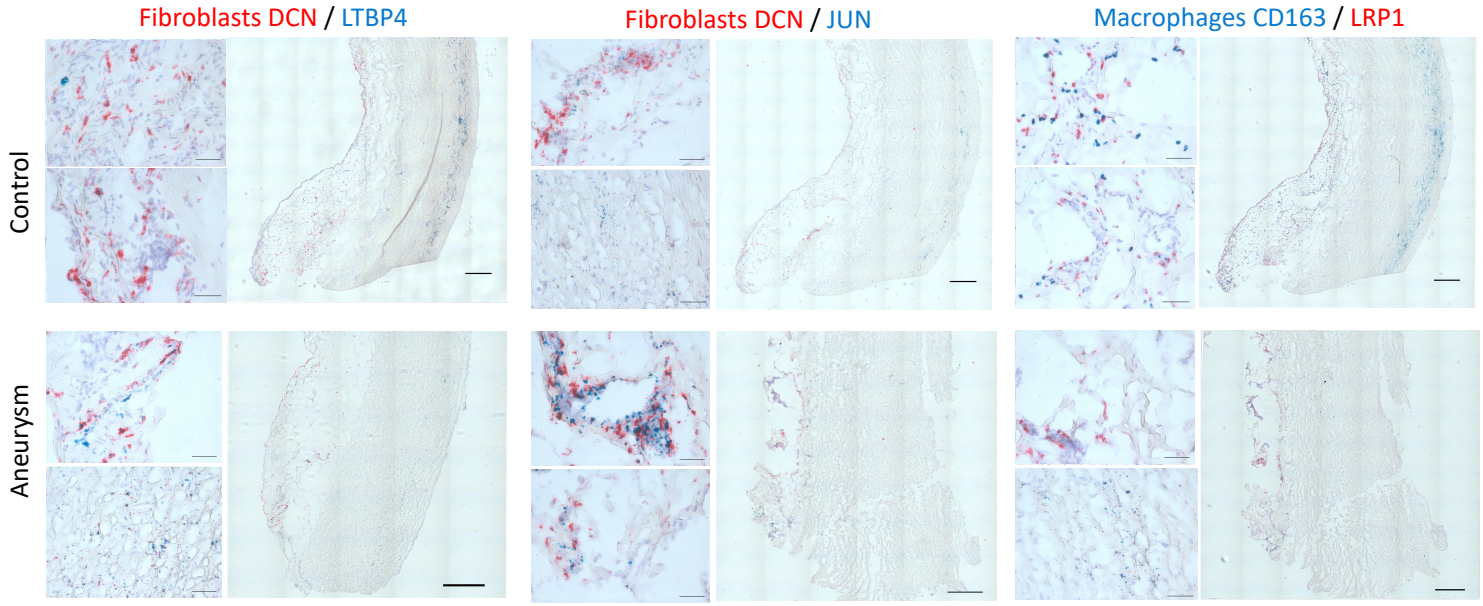
Supplemental Figure 11. a. Expression of genes upregulated in activated, human aortic fibroblasts previously reported in Dawson *et al*, 2021[65] across fibroblast sub-clusters. **b.** Expression of genes from a profibrotic, human cardiac fibroblast population previously reported in Tucker *et al*, 2020[46] across fibroblast sub-clusters. Avg Expr, Average log-normalized expression scaled to the maximum expression in any sub-cluster; Pct Nuclei Expr > 0, Percent of nuclei in a given sub-cluster that express the gene at non-zero levels.



Supplemental Figure 12. Differential expression in fibroblasts between aneurysm cases and controls after removal of FB-S2. a. Volcano plot showing the differentially expressed genes in fibroblasts between aneurysm cases and controls after removal of FB-S2. Top differentially expressed genes are labeled along with other interesting genes. The x-axis represents a log fold-change (logFC) estimate and the y-axis is the $-\log_{10}(P)$. Significant genes after multiple testing correction with the Benjamini-Hochberg correction are highlighted in red. **b.** The top up-regulated (left) and down-regulated (right) gene ontologies in fibroblasts after the removal of FB-S2. Red lines indicate a false discovery rate of 0.05 based on the Benjamini-Hochberg procedure. **c.** Concordance of differentially expressed genes from the main fibroblast analysis presented and this sensitivity analysis removing FB-S2. The grey bar represents the number of significantly differentially expressed genes in both the main analysis and the sensitivity analysis, the blue bar represents genes only found in the main analysis and not identified in the sensitivity analysis removing FB-S2, and the red bar represents genes only found in the sensitivity analysis removing FB-S2 and not the main analysis. **d.** Scatterplot showing the correlation from the main analysis with the analysis removing FB-S2 based on the log fold-change (logFC) estimate (left) and $-\log_{10}(P)$ (right). Pearson correlation coefficients (r) are shown in the upper left of each panel.



Supplemental Figure 13. RNA labeling of aortic tissue using RNAscope in situ hybridization of prioritized genes identified from GWAS (Figure 7). Control and aneurysm aortic tissue labeled with DCN (red, global fibroblast marker), LTBP4 (green-blue), JUN (green-blue); CD163 (green-blue, global macrophage marker) and LRP1 (red). Images show the relative increased expression of LTBP4 and JUN in fibroblasts of aneurysm tissue compared to controls and relative decreased expression of LRP1 in macrophages of aneurysm tissue compared to controls. 40x single images = 200 um; 20x tiled images = 1000 um



Supplemental Figure 14. Differential expression analysis after removal of patient 774. Concordance of differentially expressed genes in main analysis compared to sensitivity analysis removing patient 774 (top left). Gray bars are genes that are found to be significantly differentially expressed in both the main analysis and the sensitivity analysis, blue bars are genes only found in the main analysis and not identified in the sensitivity analysis removing patient 774, and red bars are genes only found in the sensitivity analysis removing patient 774 and not the main analysis. Remaining panels show scatterplots for each cell type comparing the log fold-change estimates and *P*-values from the main analysis compared to the sensitivity analysis removing patient 774 based on the model including a *duplicateCorrelation()* effect for study of origin. Pearson correlations (*r*) are shown in the top left of panels. VSMC, Vascular smooth muscle cell; DEG, Differentially expressed gene; logFC, Log fold-change.