Title: Multiomics of human aortic endothelial cells reveals cell subtypes with heterogeneous responses to canonical endothelial-to-mesenchymal perturbations.

Authors: Maria L. Adelus[1,2], Jiacheng Ding[1], Binh T. Tran[1], Austin C. Conklin[1], Anna K. Golebiewski[1], Lindsey K. Stolze[1], Michael B. Whalen[1], Darren A. Cusanovich[1,3], Casey E. Romanoski[1, 2, 3, 4]

Affiliations:

[1] The Department of Cellular and Molecular Medicine, The University of Arizona, Tucson, AZ 85721, USA.
[2] The Clinical Translational Sciences Graduate Program, The University of Arizona, Tucson, AZ, 85721, USA
[3] Asthma and Airway Disease Research Center, The University of Arizona, Tucson, AZ, 85721, USA
[4] Corresponding Author: cromanoski@arizona.edu

ABSTRACT

Objective: Endothelial cells (ECs) regulate atherogenesis with Endothelial-to-Mesenchymal (EndMT) transition correlating with disease. Single cell (sc) effects of EndMT perturbations *in vitro,* with quantitative comparison to signatures of human lesions *in vivo* is lacking.

Approach and Results: Multiomic profiling that concurrently measures sc RNA and ATAC was performed on 8 distinct primary cultures of human aortic ECs exposed to 7-day of EndMT-promoting perturbations: IL1B, TGFB2, and ERG knockdown (siERG). Meta-analysis of sc transcriptomes across 17 human arterial specimens was performed and two quantitative measures assessed the similarities of *ex vivo* versus *in vitro* molecular profiles. Primary HAEC cultures were reproducibly populated by 4 major clusters, termed EC1-4: EC1-angiogenic; EC2-proliferative; EC3-activated/mesenchymal-like; and EC4-mesenchymal. Independent exposure to siERG, IL1B and TGFB2 elicited mostly distinct transcriptional and chromatin accessible responses. EC1 and EC2, the most canonically 'healthy' EC populations were affected predominantly by siERG; the activated cluster EC3 was most responsive to IL1B; and the mesenchymal population EC4 was most affected by TGFB2. Quantitative comparisons between *in vitro* and *ex vivo* transcriptomes confirmed EC1 and EC2 as most canonically EC-like, and EC4 as most mesenchymal with minimal effects elicited by siERG, IL1B and TGFB2. Lastly, accessible chromatin regions unique to EC2 and EC4 were most enriched for CAD-associated SNPs from GWAS suggesting these cell phenotypes harbor CAD-modulating mechanisms.

Conclusion: Primary EC cultures contain markedly heterogenous cell subtypes defined by their molecular profiles. Surprisingly, pro-EndMT exposures for 7 days were inadequate to shift cells from one subpopulation to another suggesting relatively stable molecular phenotypes in culture. Interpretations could be that EndMT acts on a modest number of transcripts or that the *in vitro* systems used herein fail to recapitulate the complex EndMT-promoting microenvironment of human atherosclerotic lesions. Recognizing and leveraging heterogeneity *in vitro* should improve fidelity of these systems for modeling *in vivo* biology.

INTRODUCTION

Atherosclerosis is the underlying pathobiology of ischemic heart disease, the leading cause of morbidity and mortality worldwide (1-3). Atherosclerosis is characterized by plaque formation in the artery wall (4), with both inflammation (5) and lipid deposition (6) being considered important contributors to its progression. Endothelial-to-mesenchymal transition (EndMT), a process in which endothelial cells (ECs) are transformed into mesenchymal-like cells with enhanced ability to migrate and form extracellular matrix (ECM) components, has also been proposed to explain (at least in part) the progression of atherosclerosis, although some have proposed that EndMT is beneficial for stabilizing plaques (7, 8).

The heterogeneity of ECs and the prevalence of EndMT are being increasingly appreciated within the context of atherosclerosis. In a lineage-traced mouse model using the *Cdh5* (VE-cadherin) promoter, luminal aortic ECs from Apoe-/- animals on a high fat diet underwent EndMT as marked by expression of *Notch3*, while EndMT was not detected in animals on a chow diet (9). In another lineage-traced mouse model using the *CD31* (PECAM-1) promoter, one-third of fibroblasts and “mesenchymal” cells were reported to be of endothelial origin (8). Although lineage tracing is not possible in humans, immunocytochemical techniques have suggested that EndMT frequently occurs in atherosclerotic vessels. These studies have described an unexpectedly high number of cells co-expressing pairs of endothelial and mesenchymal proteins, including FAP/VWF, FSP-1/VWF, FAP/CD31, FSP-1/CD31 (8), p-SMAD2/FGFR1 (9), and αSMA/PECAM-1. Notably, a strong correlation (r = 0.84) was observed between the severity of coronary artery disease (CAD) and the extent of EndMT, as measured by reduction in FGFR1 and induction of p-SMAD2 in humans (9).

Efforts to recapitulate EndMT with *in vitro* cell culture models are important for experimentally generating and testing hypotheses about the driving molecular mechanisms. Transforming growth factor-β (TGFB) signaling is one of the hallmarks of EndMT and a regulator of EC heterogeneity (8, 10). TGFB1 administration to human umbilical vein endothelial cells (HUVECs) has been shown to induce a fibroblast-like phenotype with upregulation of markers for collagen and matrix metalloproteinases (MMP) versus control, the latter providing a potential mechanism by which EndMT could destabilize plaques (9). Moreover, in a murine model, plaque area decreased 37.5% and the amount of fibrosis decreased 49.6% after Apoe$^{-/-}$ mice were injected with recombinant soluble TGFB receptor II, which inhibits TGFB signaling, underscoring a causal relationship between TGFB, EndMT, and plaque features *in vivo* (11). These findings highlight the relevance for use of TGFB to model EndMT *in vitro* (12-18).

Treatment with interleukin-1 beta (IL1B) for at least 24 hours has also been used to model EndMT which results in decreased endothelial markers *CDH5* and *PECAM1*, and an upregulation of mesenchymal genes *CHD2* (N-cadherin) and *FN1* (fibronectin) (19-23). Notably, the Canakinumab Antiinflammatory Thrombosis Outcome Study (CANTOS) clinical trial showed the incidence of cardiovascular events was reduced with administration of anti-IL1B therapy (24). The promotion of EndMT by IL1B supports the hypothesis that EndMT drives atherosclerotic progression and plaque rupture.

A third molecular pathway shown to promote EndMT is loss of function for ETS related gene (*ERG*), which encodes a transcription factor of critical importance for EC fate specification and homeostasis (25-29).

Decreased *ERG* expression is correlated with EndMT in tissue from humans with end-stage liver fibrosis (30) and HUVECs *in vitro* (31). Moreover, loss of *ERG* enhances paracrine fibroblast activation *in vitro* and impairs lung fibrosis resolution in young mice *in vivo* (32). The relationship between *ERG* and EndMT is also being studied within the cardiovascular system. Inducible endothelial *Erg* deletion ($ERG^{iEC\text{-}KO}$) in mice is associated with spontaneous thrombosis, hemorrhages and systemic coagulopathy (33). Finally, in humans, *ERG* was observed to be absent in ECs located in the most vulnerable regions of atherosclerotic plaques (25, 31).

With the advent of single cell technologies, there has been an increase in the number of papers exploring EC diversity *in vitro* and *in vivo* (34-39). Several EC subtypes have been discovered using single-cell methods, including, but not limited to, proliferative (40, 41), mesenchymal-like (42, 43), and angiogenic (36, 41, 44) EC subtypes. While new insights have realized the heterogeneity of ECs at single cell resolution, the majority of these studies have been performed in rodents (26, 34-37), and have used sole modality of single cell RNA sequencing (scRNA-seq) technologies (34-39), which unlike multiome (or multimodal) applications, lack additional epigenetic information for the same cells. Precisely how EC heterogeneity is influenced by *in vitro* models of EndMT, the complexity of EC subtype response, and how this heterogeneity relates to human atherosclerotic plaques remains largely unexplored. Here, we characterize how multiomic profiles of human ECs from eight genetically diverse cultures are influenced at single cell resolution by EndMT-promoting (pro-EndMT) perturbations including TGFB2, IL1B, and *ERG* siRNA (siERG). Additionally, we perform a meta-analysis of four single cell transcriptomic datasets, containing 17 arterial samples, from mild-to-moderate calcified atherosclerotic plaques to evaluate the ability of the *in vitro* EC models to recapitulate molecular signatures observed in atherosclerosis. Lastly, we identify open chromatin regions unique to EC subsets that are enriched for CAD-associated single nucleotide polymorphisms (SNPs).

RESULTS

EC Transcriptomic Profiles Reveal a Heterogenous Population

To systematically uncover the molecular landscape of ECs at single cell resolution, we cultured primary human aortic ECs isolated from luminal digests of ascending aortas of eight de-identified heart transplant donors at low passage (p3-6) (45) (**Figure 1A**). Using the 10X Genomics multiome kit (46), single nucleus mRNA expression (snRNA-seq) and chromatin accessibility (snATAC-seq) data are collected simultaneously for a total of 15,220 nuclei after stringent quality control (**Methods**). RNA and ATAC data are integrated separately by treatment condition and then with each other according to an established pipeline (**Methods**) (47).

snRNA-seq libraries were sequenced as seven runs to a median depth of 29,732-84,476 reads and a median transcript range of 2,481-3,938 per nucleus (**Table S1** and **Table S2** in the **Data Supplement**). Five distinct EC subtypes (EC1-5) are detected from the fully integrated dataset, which includes all donors, treatments, and data types (**Figure 1B**). Subtypes EC1 and EC3 comprise cells from all donors, whereas EC2 and EC4 comprise most donors, and EC5 is nearly exclusively populated by cells from a single donor (**Figure 1C**). Because we do not observe EC5 across multiple individuals, we chose not to focus our analysis

on this subtype beyond multimodal profiling. Pathway enrichment of marker genes revealed EC1 to exhibit an angiogenic phenotype (WP4331, p-value 4.0×10^{-9}; GO:0038084, p-value 1.5×10^{-9}) with enriched transcripts including *KDR*, *NOTCH4*, *VEGFA*, and *VEGFC* (**Figure 1D-G**). EC2 was enriched in proliferation (GO:1903047, p-value 7.4×10^{-35}) with characteristic markers *CENPE*, *CENPF*, *KIF11*, *KIF4A* and *TOP2A* (**Figure 1D-G**). EC3 displayed a more mesenchymal phenotype than EC1-2 with enrichment in smooth muscle cell proliferation (GO:0048660; p-value 1.1×10^{-10}) (**Figure 1F**). From the top 200 differentially expressed genes (DEGs) for EC3 we observed additional mesenchymal pathways enriched, including NABA CORE MATRISOME (M5884; p-value 1×10^{-34}) and locomotion (GO:0040011; p-value 1.2×10^{-15}). EC3 highly expressed activated mesenchyme markers including *BMP2*, *ATP8A1*, and *SERPINE1*, suggesting an activated mesenchymal-like phenotype (**Figure S1A-B** in the **Data Supplement**). A fourth subset, EC4, demonstrates enrichment in ECM organization (GO:0097435; p-value 3.2×10^{-19}), a process characteristic of mesenchymal cells, with distinctive expression of collagen genes, including *COL1A1*, *COL1A2*, *COL3A1*, and *COL5A1* (**Figure 1D-G**) (48, 49). EC5 displays an apparent procoagulatory phenotype (GO:0090330; p-value 1.3×10^{-6}), marked by *CD9*, *IL6ST*, *EMILIN1*, *MMRN1*, *LAMA2* and *C1QTNF1* (**Figure S1C** in the **Data Supplement**). Top marker genes and pathways for each EC subtype can be found in **Table S3-4** in the **Data Supplement**. These observations are in line with previous reports of angiogenic, proliferative, mesenchymal, and procoagulatory EC subtypes (36, 37, 41, 43, 50).

EC Subtypes Exhibit Distinct Open Chromatin Profiles and Enriched Motifs

To investigate how different transcriptional signatures across ECs correspond to distinct chromatin states, we utilize the snATAC-seq portion of the multiome dataset. The snATAC-seq data is sequenced as seven runs to a median depth range of 22,939-126,122 reads and median overlapping peak range of 3,480-19,259 per nucleus (**Table S1** and **S5** in the **Data Supplement**). Of 204,904 peaks from the entire dataset, we identifed 13,731 differential peaks, with 79 to 8,091 peaks differentially accessible per EC subtype (**Table S7** in the **Data Supplement**). Most differential peaks were identified from EC4 and EC2. Over 80% of total peaks are in intergenic or non-coding intragenic regions (**Figure 2B**).

Transcription factor (TF) motif enrichment analysis was performed using Signac (51) using differentially accessible regions (DARs) for each EC subtype (**Figure 2C**). It is important to note that TFs within a TF family may share DNA-binding motifs and cannot be distinguished by motifs alone. As a result, TF names from the Jaspar database (52) indicate the family. Finding the precise TF binding the motifs requires additional information, including expression patterns or chromatin immunoprecipitation followed by sequencing (ChIP-seq) of the factor in the proper cell types.

We find the basic helix-loop-helix (*bHLH*) motif defined by the core sequence CANNTG enriched in EC1 peaks, including enrichments for ASCL2 (adjusted p-value 3.9×10^{-50}), TCF12 (adjusted p-value 1.7×10^{-21}), and BHLHE22(var.2) (adjusted p-value 5.7×10^{-48}) (**Figure 2C-D**). ETS motifs, including ETV1 (adjusted p-value 3.2×10^{-42} and 5.3×10^{-249}, for EC1-2, respectively), SPIB (adjusted p-value 7.9×10^{-22} and 2.5×10^{-236}, respectively), and GABPA (adjusted p-value 2.7×10^{-41} and 4.3×10^{-244}, respectively), were also enriched in

EC1 as well as in EC2 peaks. These data are consistent with known roles for ETS TFs including ERG and FLI1 in governing angiogenic and homeostatic endothelial phenotypes (31). Given the relationship in expression of *ERG* (**Figure 1E**) and the ETS motif distribution in open chromatin (**Figure 2D**) across the dataset, ERG is likely driving the EC1-2 sub-phenotypes. The near-exact match in motifs between the ETV1 motif position weight matrix in Jaspar and the *de novo* enriched motif from ERG ChIP-seq in human aortic ECs (29) further supports this conclusion (**Figure 2F**). In addition to ETS motifs, EC2 is enriched in ZFX (adjusted p-value $4.2x10^{-86}$) and ZNF148 (adjusted p-value $1.1x10^{-126}$) which are C2H2 zinc finger motifs. EC5 is also enriched in ZFX (adjusted p-value $9.6x10^{-38}$) and ZNF148 (adjusted p-value $4.5x10^{-43}$), but again, these cells were derived from a single donor and so were not considered further. C2H2 zinc finger motifs, as well as KLF4 (adjusted p-value $5.4x10^{-32}$ and $8.4x10^{-135}$, for EC1-2, respectively), also show enrichment in EC1-2. EC3 peaks are enriched for GATA motifs including GATA4 (adjusted p-value $3.1x10^{-8}$), GATA5 (adjusted p-value $8x10^{-11}$), GATA1::TAL1 (adjusted p-value $1.8x10^{-6}$), and bHLH motif BHLHE22(var.2) (adjusted p-value 0.01). EC4 open regions were uniquely enriched for TEA domain (TEAD) factors comprised of motifs named TEAD2 (adjusted p-value $1.2x10^{-238}$), TEAD3 (adjusted p-value $2.1x10^{-306}$), and TEAD4 (adjusted p-value $6.9x10^{-252}$) (**Figure 2C-D, F**). Notably, TEAD factors have been found as enriched in vascular smooth muscle cells (VSMCs) (53, 54), which is consistent with EC4 having the most mesenchymal phenotype of our EC subtypes.

Taken together, these data demonstrate that EC1-2 are the subtypes most canonically like 'healthy' or angiogenic ECs insofar as they exhibit ETS motif enrichments. Additionally, we conclude that EC4 is the most mesenchymal EC insofar as it exhibits TEAD factor enrichments.

Pro-EndMT Perturbations *In Vitro* Elicit EC Subtype-Specific Transcriptional Responses

Given the heterogeneity in molecular profiles driving EC subtype, we sought to evaluate the similarities and differences among pro-EndMT perturbations and evaluate each EC subtype's transcriptomic response to these perturbations. The three pro-EndMT perturbations that were independently administered include: (1) 7-day exposure of human aortic ECs to IL1B (10 ng/ml), (2) 7-day TGFB2 (10 ng/mL), and (3) 7-day siRNA-mediated knock-down of ERG (siERG), with serial transfections on days 0 and 4 (**Figure 1A, Methods**). We compare perturbed transcriptomes to control cultures grown for 7 days in media lacking cytokine (for IL1B and TGFB2) or transfected with scrambled siRNA (siSCR) for siERG. Multiple genetically distinct human aortic EC donors were used for each comparison, with four donors included for IL1B, five for TGFB2, and three for siERG (**Table S6**). Principal component (PC) analysis reveals that pro-EndMT perturbations elicited greater variance in RNA expression (38-55% of variance) than donor (17%-27% variance) (**Figures S2A-C** in the **Data Supplement**). DEG analysis using pseudo-bulked profiles grouped by donor, subcluster, and perturbation combinations (**Table S8** in the **Data Supplement**).

We observe heterogeneity in transcriptional responses across EC subtypes. While EC1-2 transcripts are predominantly perturbed by siERG, the greatest number of transcripts differentially expressed in EC3 are responsive to IL1B, though siERG and TGFB2 also regulate tens to hundreds of transcripts in EC3. In

contrast, transcripts in EC4 are predominantly responsive to TGFB2 (**Figure 3A**, **Table S8** in the **Data Supplement**).

We considered whether pro-EndMT perturbations caused cells to alter their phenotypes so much that they take on transcriptional profiles of another EC subtype. By evaluating proportions of cells in each cluster by donor and perturbation, we find that treatment with IL1B for six hours was the only perturbation that statistically changed subtype membership – a shift of 9% average decrease in EC2 cells (±3% st.dev; p-value $4.4x10^{-3}$ from paired 2-sided t-test) and an average 12% increase in EC3 cells (±4% st.dev.; p-value $1.7x10^{-3}$ from paired 2-sided t-test) with IL1B exposure compared to control (**Figure S2D** in the **Data Supplement**). These data suggest that IL1B treatment for six hours is able to shift EC subtype to the more mesenchymal phenotype in EC3. Still, this shift was not observed after 7 day IL1B exposure, suggesting a transient state. Importantly, we do not observe cells shifting to EC4, which is the most mesenchymal sub-cluster.

When comparing enriched pathways across perturbations, we observe that over 80% of transcripts differentially expressed by a treatment in EC4 were in response to TGFB2 (**Figure 3A**, **Table S8** in the **Data Supplement**). TGFB2-affected transcripts for EC4 are enriched in invadopodia formation (R-HAS-8941237; p-value $2.7x10^{-7}$) and anchoring fibril formation (R-HAS-2214320; p-value $3.6x10^{-7}$) (**Figure 3B**). Notably, TGFB2-affected genes for EC3 share several mesenchymal-related enriched pathways with TGFB2-affected genes for EC4, including actin cytoskeleton organization (GO:0030036; p-value $4.4x10^{-7}$), NABA CORE MATRISOME (M5884; p-value $2.8x10^{-7}$), and ECM organization (R-HSA-1474244; p-value $5.4x10^{-7}$). However, TGFB2-attenuated transcripts unique to EC3 were enriched in platelet activation (GO:0030168; p-value $1.4x10^{-4}$) (**Figure 3B**).

Most transcripts affected in EC3 were responsive to IL1B (**Figure 3A**). Importantly, several EC3 genes differentially expressed with IL1B were also affected with siERG (**Figure 3A**). IL1B-affected transcripts in EC3 are not enriched in mesenchymal-like pathways (**Figure 3C**). However, EC3 IL1B-attenuated genes are enriched in blood vessel development (GO:0032502; p-value $5.1x10^{-11}$), indicating that this perturbation still has anti-endothelial effects (**Figure 3C**).

Most genes significantly affected by perturbations in EC1-2 were responsive to siERG, likely due to their more endothelial-like phenotypes compared to EC3-4 (**Figure 3A**). siERG-affected genes in EC1-2 were enriched in COVID-19 adverse outcome pathway (55) (WP4891; p-values $5x10^{-9}$ and $8.3x10^{-5}$, for EC1-2 respectively) and AGE-RAGE signaling in diabetes (56) (hsa04933; p-values $8.9x10^{-16}$ and $1.9x10^{-20}$, respectively), while EC3 siERG-perturbed genes are enriched with a unique metabolic profile demonstrated by enrichment in monosaccharide metabolic process (GO:0005996; p-value $1x10^{-6}$), carbohydrate metabolic process (GO:0005975; p-value $6.6x10^{-7}$), and aerobic glycolysis (WP4629; p-value $4.1x10^{-5}$) (**Figure 3D**). In contrast, EC4 siERG-induced genes are enriched in positive regulation of angiogenesis (GO:0045766; p-value $4.5x10^{-6}$), a function normally impaired in *ERG*-depleted endothelial cells (**Figure 3D**) (26).

Due to the role that ERG plays in inhibiting NF-KB-dependent inflammation *in vitro* and *in vivo* (25), we set out to characterize mutually exclusive and shared pathways between IL1B and siERG (**Figure 3E**). Importantly, siERG but not IL1B-perturbed genes involve several previously mentioned metabolic processes

including carbohydrate metabolic process (GO:0005975; p-value $6.6x10^{-7}$), aerobic glycolysis (WP4629; p-value $4.1x10^{-5}$), and monosaccharide metabolic process (GO:0005996; p-value $1x10^{-6}$). This suggests differences in the ability of ERG and IL1B to modify metabolism. Interestingly, IL1B but not siERG upregulated Interferron signaling and viral responsive pathways (GO:0051607, p-value $1x10^{-37}$; R-HSA-913531, p-value $1x10^{-41}$). Shared IL1B- and siERG-upregulated genes were enriched in COVID-19 adverse outcome pathway (WP4891; p-value $1.9x10^{-9}$) (55). Shared IL1B- and siERG-attenuated genes are enriched in several processes involving ribosomal proteins, including ribosome, cytoplasmic (CORUM:306; p-value $3.3x10^{-7}$), cytoplasmic ribosomal proteins (WP477; p-value $5.3x10^{-7}$), and peptide chain elongation (R-HSA-156902; p-value $5.9x10^{-7}$) (**Figure 3E**). This finding indicates that the downregulation of ribosomal genes is a hallmark of inflammatory and *ERG*-depleted endothelium. Altogether, these data demonstrate the heterogeneity in EC subtype response to EndMT perturbations.

In Vitro **EC EndMT Models Reorganize Epigenetic Landscapes with Subtype Specificity**

To gain insight into gene regulatory mechanisms responsible for EC subtype transcriptional responses to IL1B, TGFB2, and siERG, we compare the effects of these perturbations on chromatin accessibility. Across all three treatments, we identified 4,034 differentially accessible regions (DARs, **Table S9** in the **Data Supplement**, **Methods**). The majority of DARs for EC1-2 were responsive to siERG, while the majority of DARs for EC3 were responsive to IL1B (**Figure 4A**, **Table S9** in the **Data Supplement**). Interestingly, the epigenetic landscape of EC4 differs from its transcriptional response, insofar as most peaks were responsibe to IL1B (not TGFB2) (**Figure 4A**, **Table S9** in the **Data Supplement**). To inform the TFs likely bound to differentially accessible regulatory elements, motif enrichment analysis was performed (**Figure 4B-D**). Several distinct TF motifs were enriched across EC subtypes. For IL1B,, we observed enrichment in KLF15 (adjusted p-value $5x10^{-10}$) (kruppel like factor 15) in EC2 alone (**Figure 4B**). siERG induced peaks showed subtype-specific motif enrichments, including TWIST1 (adjusted p-value $2.5x10^{-22}$) (twist family bHLH transcription factor 1), HAND2 (adjusted p-value $2.3x10^{-19}$) (heart and neural crest derivatives expressed 2) for EC1, RELA (adjusted p-value $9.5x10^{-20}$) (RELA proto-oncogene, NF-KB subunit) for EC2, and CEBPD (adjusted p-value $1.6x10^{-29}$) for EC3(**Figure 4C**). Minimal motif enrichment was observed with siERG for EC4.

We also found several TF motifs enriched across more than one EC subtype upon perturbation. IL1B-affected peaks gained in EC1 and EC2 shared enrichments for TFDP1 (adjusted p-value $1.3x10^{-4}$ and $9x10^{-4}$ for EC1-2, respectively) (transcription factor Dp1) and ZBTB14 motifs (adjusted p-value $2.2x10^{-4}$ and $2x10^{-8}$, respectively) (zinc finger and BTB domain containing 14). IL1B-induced peaks in EC3 and EC4 shared enrichment for CEBPD (adjusted p-value $4.4x10^{-73}$ and $1.6x10^{-33}$ for EC3-4, respectively) and CEBPG motifs(adjusted p-value $5.4x10^{-45}$ and $7.1x10^{-18}$, respectively) (CCAAT enhancer binding protein delta and gamma) (**Figure 4B**). TGFB2-affected peaks in EC1, EC2, and EC3 shared enrichment for ZBTB14 (adjusted p-values $6.8x10^{-31}$, $5.1x10^{-12}$, and $2x10^{-8}$, for EC1-3, respectively) while TGFB2-induced peaks in EC3 and EC4 shared enrichment for the SMAD5 motif (adjusted p-value $7.4x10^{-6}$ and $4.2x10^{-11}$, for EC3-4, respectively)

(SMAD family member 5) (**Figure 4D**). Taken together, while several enriched motifs are shared across EC subtypes, divergent epigenetic landscapes are also induced with pro-EndMT perturbations. We therefore conclude that different transcriptional responses to these perturbations across EC subtypes are elicited by distinct TFs, including members of families of the KLF, TWIST, HAND, p65, and CEBP families.

Meta-Analysis of *Ex Vivo* Human Atherosclerotic Plaque snRNA-seq Datasets

To understand the diversity of ECs in human atherosclerotic plaques and evaluate their relationships to oour *in vitro* system, we performed a meta-analysis of data from recent publications that utilized scRNA-seq from human atherosclerotic lesions (53, 57-59) (accessions in **Table S10** in the **Data Supplement**). We identified 24 diverse clusters among 58,129 cells after integration of 17 different coronary and carotid samples (**Figure 5A** and **Table S11** in the **Data Supplement**). Clusters were annotated using a combinatorial approach including canonical marker genes, CIPR (60), and the original publications (**Figure 5B**). Clusters were annotated as: T-lymphocytes, natural killer T-cells, ECs, macrophages, VSMCs, fibroblasts, B-lymphocytes, basophils, neurons, and plasmacytoid dendritic cells (PDCs) (**Figure 5A**). We find the greatest proportion of cells belonging to each major cell type derive from carotid arteries, except for neurons which derive exclusively from coronary arteries, and PDCs which derive exclusively from carotid arteries (**Figures S3B-C** in the **Data Supplement**). Expected pathway enrichments are observed for annotated cell types, including NABA CORE MATRISOME (M5884; p-value $4.8x10^{-41}$) for fibroblasts, blood vessel development (GO:0001568; p-value $5.6x10^{-33}$) for ECs, and actin cytoskeleton organization (GO:0030036; p-value $1.3x10^{-15}$) for and VSMCs (**Figure S3D-G** in the **Data Supplement**). These observations support the diverse composition of human atherosclerotic lesions.

We evaluated what pathways distinguished the Endothelial Cells 1 (Endo1) and Endothelial Cells 2 (Endo2) subtypes from our *ex vivo* meta-analysis (**Figure 5C**). We found Endo2 has an EndMT-related phenotype, with enrichment in mesenchymal pathways including NABA MATRISOME ASSOCIATED (M5885; p-value $1.6x10^{-14}$), ECM organization (R-HSA-1474244; p-value $6x10^{-17}$), skeletal system development (GO:0001501; p-value $3.4x10^{-13}$) and network map of SARS-CoV-2 signaling pathway (55) (WP5115; p-value $1.3x10^{-11}$) (**Figure 5C-D**). Additionally, we observe enrichment for inflammatory pathways in Endo2 including prostaglandin synthesis and regulation (WP98; p-value $1.2x10^{-7}$), and complement and coagulation cascades (hsa04610; $1x10^{-10}$) (**Figure 5C, D**) (61, 62). On the contrary, Endo1 was highly enriched in multicellular organismal homeostasis (GO:0048871; p-value $3.3x10^{-8}$) and lowly enriched in mesenchymal pathways (M5885; p-value $1x10^{-3}$; no enrichment for R-HSA-1474244, GO:0001501, or WP5115), indicating an EC phenotype which has not undergone EndMT (**Figure 5C-D**). Interestingly, Endo1, but not Endo2, is highly enriched in ribosome, cytoplasmic pathway (CORUM:306; p-value $9.3x10^{-96}$) and TRBP containing complex (CORUM:5380; DICER, RPL7A, EIF6, MOV10 and subunits of the 60S ribosomal particle; p-value $1.5x10^{-22}$), suggesting a potential protective role for this complex along with ribosomal gene expression (63, 64). The depletion of these pathways may serve as a hallmark of activated endothelium (**Figure 5C-E**). We interpret

these results to suggest that Endo1 is a classical endothelial state, while Endo2 appears to be characterized by EndMT.

Ex Vivo-derived Module Score Analysis Reveals Differences among *In Vitro* EC Subtypes and EndMT Stimuli

To directly evaluate relationships between the *ex vivo* and *in vitro* cell subpopulations, we utilized module scores. These quantitative scores are based on the *ex vivo* marker genes for each cluster and are evaluated for each *in vitro* cell subcluster. Unexpectedly, the *ex vivo* cluster that consistently generates the greatest module scores for *in vitro* ECs is the VSMCs cluster 5 (VSMC5) (**Figure 5A**; **Figure S4A** in the **Data Supplement**). VSMC5 bridges the EC to SMC and fibroblast clusters in the *ex vivo* analysis (**Figure 5A**). Marker genes of VSMC5 are expressed across *ex vivo* and *in vitro* clusters (**Figure S5A** in the **Data Supplement**) and include important regulators of ECM, such as *BGN*, *VCAN*, *FN1*, as well as several collagen genes (*COL1A1*, *COL1A2*, *COL3A1*, *COL6A1*) (**Figure S5B** in the **Data Supplement**). VSMC marker transcripts also include several lncRNAs and mitochondrial transcripts (CARMN, MALAT1, NEAT1; MT-ATP6, MT-ND4, and MT-CYB). *Ex vivo* Endo1 and Endo2 module scores are the second highest scoring across *in vitro* clusters. Cells scoring high for Endo1 are concentrated in the *in vitro* EC1 cluster, while cells scoring high in Endo2 are concentrated to the *in vitro* EC3 locale (**Figure S4B-E** in the **Data Supplement**). This supports that EC3 is a more activated subtype than EC1. Finally, among *in vitro* cells, those with highest VSMC5 module scores are concentrated in EC4, underscoring that EC4 is a more mesenchymal sub-phenotype *in vitro* (**Figure S4B-E** in the **Data Supplement**).

We stratify these analyses by pro-EndMT treatment and find greater VSMC5 module scores with TGFB2 treatment versus control for EC4 (adjusted p-value = 9.9×10^{-15}) (**Figure S6A-C** in the **Data Supplement**). However, there is no difference in VSMC5 module scores for EC1-3 between control and TGFB2 treatment. Moreover, we observe lower Endo1 module scores with TGFB2 for EC4 (adjusted p= 2.53×10^{-8}), and higher scores for EC1 (adjusted p=8.0×10^{-8}), suggesting protective mechanisms against EndMT in more endothelial-like EC subtypes (i.e., EC1-3) that are absent in more mesenchymal-like EC (i.e.,EC4) subtypes (**Figure S6A-C** and **Table S12-13** in the **Data Supplement**). We observe siERG lowers Endo1 scores across all EC subtypes (adjusted p=9.9×10^{-15} for EC1-4), indicating ERG depletion decreases endothelial-likeness across all EC subtypes (**Figure S6A-C** and **Table S12-13** in the **Data Supplement**). Moreover, siERG increases VSMC5 scores for EC2 (adjusted p=2.8×10^{-9}) and EC3, indicating siERG also induces EndMT for proliferative and activated mesenchymal-like EC phenotypes (adjusted p-valulue 0.04) (**Figure S6A-C** and **Table S12-13** in the **Data Supplement**).

EC Subtype is a Major Determinant in Modeling Cell States Observed in Atherosclerosis

In addition to module score analysis, we apply a complementary approach to quantitatively relate *in vitro* EC subtypes and EndMT perturbations to *ex vivo* cell types. We calculate average expression profiles for all major cell populations in both *ex vivo* and *in vitro* datasets and examine the comprehensive pairwise

relationship among populations with hierarchical clustering of Spearman Correlation (**Figure 6A**). All *in vitro* transcripts significantly regulated across all pro-EndMT perturbations at 5% False Discovery Rate (FDR) (65) are used in this analysis, although several additional means to select transcripts did show similar results. This analysis reveals three major observations. First, *in vitro* EC4-5 cells are most like mesenchymal *ex vivo* cell types including VSMCs and fibroblasts (indicated by the yellow block of correlations in the bottom left of the heatmap in **Figure 6A**). Second, *in vitro* EC1-3 are most like *ex vivo* Endo1 and Endo2 populations, especially among the siSCR and 7-day control cells. Moreover, cells in the siSCR condition in EC1 are most like *ex vivo* Endo1, reinforcing that these two populations are the most canonically 'healthy' endothelial populations. Third, while EndMT treatments did elicit variation in how similar cells in EC1-5 resembled *ex vivo* transcriptomic signatures, these effects are secondary to which subtype the cells belonged (**Figure 6A**). Taken together, these findings underscore that EC subtype, versus perturbation, is a greater determinant of similarity to *ex vivo* cell types.

CAD-Associated Genetic Variants Are Enriched Across EC Subtype Epigenomes

Genetic predisposition to CAD is approximately 50% heritable with hundreds to thousands of genetic loci supposed to be involved in shaping an individual's propensity for disease (66, 67). Most CAD-associated variants are not protein coding, suggesting they perturb cellular function through gene regulatory functions. We therefore ask whether the open chromatin regions in this *in vitro* dataset coincide with locations of single nucleotide polymorphisms (SNPs) reported in the latest CAD meta-GWAS analysis from the Millions Veterans Project (MVP), which includes datasets from CARDIoGRAMplusC4D 1000G study, UK Biobank CAD study, and Biobank Japan (68). We find significant enrichment in CAD-associated SNPs for the complete set of accessible regions across all EC subtypes (termed "panEC"; adjusted p-value $1.5ex10^{-93}$; Odds Ratio (OR)= 1.8; **Figure 6B**, **Table S14-15** in the **Data Supplement**) when comparing CAD SNPs exceeding the genome-wide significance threshold of $p<5x10^{-8}$ versus non-significant SNPs (**Methods**). Among accessible regions unique to EC subtypes, EC4 shows the greatest enrichment (adjusted p-value $7.85x10^{-6}$; OR=1.74). Additionally, EC2 is also enriched for CAD SNPs (adjusted p-value $6.3x10^{-8}$; OR=2.15), supporting a role for proliferative ECs in CAD. Of all accessible regions influenced by pro-EndMT perturbations, siERG and TGFB2 sets are most enriched for CAD variants (**Figure 6B, Table S14-15** in the **Data Supplement**).

The measurement of both gene expression and DNA accessibility in the same cell enables testing for direct links between noncoding DNA elements and their potential regulatory targets (i.e., gene promoters). This is achieved by testing for correlation between DNA accessibility and the expression of a nearby gene across single cells (51, 69). Focusing on EC4, we search for EC4-specific sites of correlated chromatin accessibility and linked target gene expression. Upon restricting linked peaks overlapping CAD SNPs, we identify 81 significant SNP-peak-gene trios ($p < 0.05$) representing 46 unique genes with specific activity in EC4 (**Table S16** in the **Data Supplement**). We submit the 46 unique genes to Metascape (70) and observe enrichment in EndMT-related pathways including blood vessel development (GO:0001568; p-value $2.1x10^{-10}$), crosslinking of collagen fibrils (R-HSA-2243919; p-value $1.4x10^{-8}$), and canonical and non-canonical TGF-

B signaling (WP3874; p-value 2.2×10^{-6}) (**Figure S7** in the **Data Supplement**). Literature review of this gene list further confirms several linked EC4-restricted genes associated with cardiovascular disease, including *COL4A1*, *COL4A2*, *PECAM1,* DSP, and*BMP6*, (**Figure 6C-E**) (71-73).

Altogether, these data underscore that common genetic variation influences individual propensities for CAD through ECM-organizing functions evidenced by the EC4 phenotype.

DISCUSSION

The major goals of this study were fourfold: (1) to quantitatively assess molecular heterogeneity of cultured human aortic ECs *in vitro*, (2) to evaluate and compare molecular changes elicited by EndMT perturbations at single cell resolution, (3) to assess similarities between *in vitro* and *ex vivo* EC signatures to inform the extent to which *in vitro* models recapitulate *ex vivo* biology, and (4) investigate how heterogenous EC populations are enriched for genetic associations to CAD. Findings for each of these goals are discussed below along with important implications and questions arising from this work.

The multiomic single cell profiles of 15,220 cells cultured *in vitro* from eight individuals enabled the discovery of 5 EC subpopulations, named EC1-5. Except for EC5, EC subpopulations were comprised of cells from multiple donors and EndMT perturbations, which lends credence to the reproducibility of these biological states. The loosely defined phenotypes, based on pathway enrichment analysis, are healthy/angiogenic for EC1, proliferative for EC2, activated mesenchymal-like for EC3, mesenchymal for EC4, and procoagulatory for EC5. Angiogenic (36, 37, 41), proliferative (43, 74), mesenchymal (43), and procoagulatory (50) ECs have been previously reported in literature. Pro-EndMT perturbations have markedly unique effects on different EC subclusters, highlighting the fact that *in vitro* systems contain populations of discrete cell subtypes, or states, that respond divergently to even reductionistic experimental conditions. Implications of such heterogeneity include both a need to elucidate what factors dictate treatment responsiveness as well as experimental design and data interpretation that considers heterogeneity of response. The exact origin of EC heterogeneity observed in this study is unknown. We consider it likely that EC1-4 populations, which were populated by most donors, date back to the original isolation of ECs from aortic trimmings, which would imply that different states were preserved across passage in the culture conditions. However, we cannot exclude the possibility that some of the subpopulations have expanded since seeding of the cultures. If that were the case, EC1-4 represent reproducible cell states consequent to primary culture of arterial cells. In fact, the limited correlation with *ex vivo* data supports (at least partially) this interpretation. Future studies will be required to delineate the exact source of heterogeneity in these systems.

This study set out to elucidate whether EndMT is an end-stage phenotype for ECs and whether it can be achieved *in vitro* upon stimulation with different perturbations: notably, TGFB2, IL1B, and siERG. The duration and doses employed were established based on literature reports from which each perturbation was found to elicit EndMT phenotypes, albeit usually EndMT was quantified by only a few marker genes (16, 21, 31). We have found that TGFB2, IL1B, and siERG have many distinct effects on EC molecular profiles (**Figures 3-4**). In general, TGFB2 elicits a greater transcriptomic and epigenomic response in the

mesenchymal EC subtype, EC4, while siERG and IL1B regulate the greatest numbers of shared transcripts and chromatin regions in more endothelial clusters EC1-3. One interpretation for this finding is that IL1B treatment and depletion of *ERG* directly affect rewiring transcription in ECs while TGFB2 may affect other cell types in the vascular wall (or culture plate) that in turn affect ECs through paracrine interactions. Part of the similarities between IL1B and siERG responses may be explained by the fact that *ERG* depletion increases IL1B production (29). We did not expect to find that none of the EndMT treatments significantly shift EC cluster membership. Although on average 12% more cells were observed in EC3 after six hours IL1B exposure, this change was not consistent at the 7-day IL1B timepoint, suggesting a transient shift in EC state. Therefore, we conclude that the perturbations in this study were not effective at converting ECs to fully mesenchymal cells. If that were the case, we would have seen an enrichment in cells in EC4 or another exclusively mesenchymal cluster after EndMT perturbation. Nonetheless, several alternative possibilities could explain why perturbations did not result in complete EndMT: other combinations of perturbations, doses, or timepoints may be necessary; EndMT may represent a transient state *in vivo*; and/or interactions with immune cells or hemodynamic shear stress may be necessary.

The current study sought to evaluate similarities and differences between *in vitro* primary cultures of human aortic ECs to *ex vivo* single cell signatures of cells from human lesions. First, we leveraged transcriptomic profiles from clusters in the scRNA meta-analysis of human lesions and evaluated each *in vitro* cluster using a module score (**Figures 5** and **Figure S6** in the **Data Supplement**). The three *ex vivo* clusters with greatest similarity to *in vitro* clusters were Endo1, Endo2, and VSMC5. Pathway enrichment analysis suggested that the *ex vivo* Endo1 cluster is close to the classic “healthy” EC state relative to Endo2, which returned pathway enrichments consistent with activated endothelium (**Figure 5C**-**D**). Interestingly, Endo 2 is depleted in in ribosome transcripts as well as transcripts in the Dicer complex (**Figure 5C**-**E**), which may serve as hallmarks of dysregulated endothelium *in vivo*. VSMC5 is an interesting *ex vivo* cluster insofar as it spans the endothelial, fibroblast, and VSMC clusters (**Figure 5A**) and is enriched for genes in actin cytoskeleton, extracellular matrix organization, and more (**Figure S5** in the **Data Supplement**). I*n vitro* EC clusters EC1-3 score generally greater in Endo1 and Endo2 relative to the more mesenchymal EC4-5 (**Figure S4** in the **Data Supplement**). Consistent with the intent of the pro-EndMT treatments, they generally decrease Endo1 and Endo2 scores and increase VSMC5 scores. However, these effects are unexceptional in comparison to effects of EC subtype. In addition to module scores, we also utilized unsupervised clustering of Spearman correlation coefficients across *ex vivo* and *in vitro* average gene expression profiles, finding again that EC1-3 are more like Endo1 and Endo2 and EC4-5 are more like VSMCs (**Figure 6A**). As expected, the control (siSCR) cells are most correlated to healthy Endo1 transcriptomes; however, the correlation coefficient achieved is modest, at rho = 0.56. While reinforcing that *in vitro* cell cultures best resemble ECs isolated *ex vivo* – regardless of EndMT perturbation –, this finding accentuates how different cultured cells are and paves the way for quantitatively evaluating and improving *in vitro* models.

Finally, GWAS studies have established that hundreds of independent common genetic variants in human populations affect risk for CAD, yet discovering the causal mechanisms remains a major challenge

given that most of the risk is in non-coding regions of the genome. One approach to prioritize causal variants in regulatory elements is through integration of open chromatin regions from the cell type and states of interest followed by expression quantitative trait loci (eQTL) or other linking evidence to target gene (75, 76). In the current study, we find significant enrichment for CAD-risk variants in open chromatin regions across all clusters ("panEC") as well as specifically for EC2 and EC4 subpopulations (**Figure 6B; Table S14-16** in the **Data Supplement**). Taken together, these data emphasize the value in multimodal datasets in human samples for prioritizing disease-associated SNPs and mechanisms.

METHODS

Tissue Procurement and Cell Culture

Primary human aortic ECs were isolated from a total of eight de-identified deceased heart donor aortic trimmings (belonging to three females and five males) at the University of California Los Angeles Hospital as described previously (45). An additional two cell lines, including immortalized human aortic ECs (teloHAECs) and teloHAECs that underwent ERG knockout (KO) were included in the dataset during integration, but not used for the purposes of analysis in this study. ECs were isolated with a device developed to allow initial exposure of only the endothelium to enzymatic digestion. After washing, cells were seeded in wells previously coated with 1% gelatin and treated with 1ug/cm^2 human fibronectin. ECs were identified by their typical cobblestone morphology, presence of Factor VIII-related antigen, and uptake of acetylated LDL labeled with 1,1'-dioctadecyl-1-3,3,3',3'-tetramethyl-indo-carbocyan-ine perchlorate (Di-acyetl-LDL), and grown and propagated in growth medium (M199 containing 20% FBS, supplemented with 2 mM L-glutamine, 1mM sodium pyruvate, 100 U/ml penicillin, and 100 ug/mL streptomycin) supplemented with heparin and EC growth supplement. ECs were treated prior to harvest for 7 days with either 10ng/mL TGFB2, IL1B, or no additional protein; two doses of small interfering RNA for ERG locus (siERG), or randomized siRNA (siSCR); or 6 hours with either 1ng/mL IL1B, or no additional protein.

Nuclear Dissociation and Library Preparation

Nuclei from primary cells were isolated according to 10x Genomics *Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing* Demonstrated Protocol (CG000365, Rev C) (77). Nuclei were pooled isolated with lysis buffer consisting of 10 mM Tris-HCl (pH 7.5, Invitrogen, cat. no. 15567027), 10 mM NaCl (Invitrogen, cat. no. AM9759), 3 mM $MgCl_2$ (Alfa Aesar, cat. no. J61014), 0.1% Tween-20 (Thermo Scientific, cat. no. 9005-64-5), 0.1% IGEPAL CA-630 (Thermo Scientific, cat. no. J61055.AP), 0.01% Digitonin (Thermo Fisher, cat. no. BN2006), 1% BSA (Sigma Aldrich, cat. no. A2153), 1 mM DTT (Thermo Fisher Scientific, cat. no. 707265ML), 1U/µl RNase inhibitor (Sigma Protector RNase inhibitor; cat. no. 3335402001), and nuclease-free water (Invitrogen, cat. no. 10977015). The seven pooled samples were incubated on ice for 6.5 minutes with 100 µl lysis buffer and washed three times with 1 mL wash buffer consisting of 10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 1% BSA, 0.1% Tween-20, 1 mM DTT, 1U/µl RNase inhibitor, and nuclease-free water. Samples were centrifuged at 500 rcf for 5 minutes at 4C, and the pellets

were resuspended in chilled Diluted Nuclei Buffer consisting of 1X Nuclei Buffer (20X) (10X Genomics), 1 mM DTT (Thermo Fisher Scientific, cat. no. 707265ML), 1U/µl RNase inhibitor, and nuclease-free water. The homogenate was filtered through a 40-µm cell strainer (Flowmi, cat. no. BAH136800040) prior to proceeding immediately to 10X Chromium library preparation according to manufacturer protocol (CG000338).

Genotyping and Multiplexing Cell Barcodes for Donor Identification

Genotyping of EC donors was performed as described previously (75). Briefly, IMPUTE2 (78) was used to impute genotypes utilizing all populations from the 1000 Genomes Project reference panel (phase 3) (79). Genotypes were called for imputed SNPs with allelic R2 values greater than 0.9. Mapping between genomic coordinates was performed using liftOver (80). VCF files were subset by genotypes for the donors of interest using VCFtools (81).

To identify donors across the *in vitro* dataset, snATAC- and snRNA-seq output BAM files from Cell Ranger ARC (10X Genomics, v.2.0.0) (46) were concatenated, sorted, and indexed using samtools (82). The concatenated BAM files were input – along with the genotype VCF file – to demuxlet (83) to identify best matched donors for each cell barcode using options "–field GT". Verification of accurate donor identification was confirmed by visualizing female sex specific *XIST* for the known donor sexes (**Figure S8** in the **Data Supplement**).

snRNA-seq Bioinformatics Workflow

A target of 10,000 nuclei were loaded onto each lane. Libraries were sequenced on NovaSeq6000. Reads were aligned to the GRCh38 (hg38) reference genome and quantified using Cell Ranger ARC (10X Genomics, v.2.0.0) (46). Datasets were subsequently preprocessed for RNA individually with Seurat version 4.3.0 (47). Seurat objects were created from each dataset, and cells with < 500 counts were removed. This is a quality control step, as it is thought that cells with low number of counts are poor data quality. Similarly, for each cell, the percentage of counts that come from mitochondrial genes was determined. Cells with > 20% mitochondrial gene percent expression (which are thought to be of low quality, possibly due to membrane rupture) were excluded. Demuxlet (83) was next used to remove doublets. The filtered library was subset and merged by pro-EndMT perturbation. Data were normalized with NormalizeData, and cell cycle regression was performed by generating cell cycle phase scores for each cell using CellCycleScoring, followed by regression of these using ScaleData (84). Batch effects by treatment were corrected using FindIntegrationAnchors using 10,000 anchors, followed by IntegrateData.

snATAC-seq Bioinformatics Workflow

A target of 10,000 nuclei were loaded onto each lane. Libraries were sequenced on an NovaSeq6000. Reads were aligned to the GRCh38 (hg38) reference genome and quantified using Cell Ranger ARC (10X Genomics, v.2.0.0) (46). Datasets were subsequently preprocessed for ATAC individually with Seurat v4.3.0 (47) and Signac v1.6.0 (85) to remove low-quality nuclei (nucleosome signal > 2, transcription start site

enrichment < 1, ATAC count < 500, and % mitochondrial genes > 20) (47). Next, demuxlet (83) was used to remove doublets. A common peak set was quantified across snATAC-seq libraries using FeatureMatrix, prior to merging each lane. A series of two iterative peak calling steps were performed. The first step consisted of calling peaks for every EndMT perturbation, and the second involved calling peaks for every cluster generated from Weighted Nearest Neighbor Analysis (WNN) (**Methods**, **"Integration and Weighted Nearest Neighbor Analyses"**). Latent semantic indexing (LSI) was computed after each iterative peak calling step using Signac standard workflow (51). Batch effects by treatment were finally corrected using FindIntegrationAnchors using 10,000 anchors, followed by IntegrateData.

Integration and Weighted Nearest Neighbor Analyses

Following snRNA-seq and snATAC-seq quality control filtering, barcodes for each modality were matched, and both datasets were combined by adding the snATAC-seq assay and integrated LSI to the snRNA-seq assay. WNN (47) was next calculated on the combined dataset, followed by joint UMAP ($_{WNN}$UMAP) visualization using Signac (51) functions FindMultimodalNeighbors, RunUMAP, and FindClusters, respectively. WNN is an unsupervised framework to learn the relative utility of each data type in each cell, enabling an integrative analysis of multimodal datasets. This process involves learning cell-specific modality "weights" and constructing a $_{WNN}$UMAP that integrates the modalities. The subtypes discovered in the first round of WNN were utilized in an additional peak calling step for snATAC-seq, followed by latent semantic indexing (LSI) computation, re-integration, and a final round of WNN to achieve optimal peak predictions (**Methods, "Single Nucleus ATAC Sequencing Bioinformatics Workflow**") (86).

Differential Expression and Accessibility Region Analyses Across EC Subtypes and EndMT Perturbation-Subtype Combinations

Differential expression between clusters was computed by constructing a logistic regression (LR) model predicting group membership based on the expression of a given gene in the set of cells being compared, with pro-EndMT perturbation included as a latent variable in the model and comparing this with a null model using a likelihood ratio test (LRT). This was performed using Seurat FindMarkers, with "test.use = LR" and "latent.vars" set to perturbation. Differential expression between perturbation and control for each cluster was performed using pseudobulk method with DESeq2 (87). Raw RNA counts were extracted for each EndMT perturbation-subtype combination and counts, and metadata were aggregated to the sample level.

Differential accessibility between EC subtypes was performed using FindMarkers, with "test.use = LR" and latent.vars set to both the number of reads in peaks and perturbation. Finally, differential accessibility between perturbation and control for each cluster was performed using FindMarkers, with "test.use = LR" and latent.vars set to the number of reads in peaks.

Bonferroni-adjusted p-values were used to determine significance at adjusted p-value < 0.05 for differential expression, and p-value < 0.005 for differential accessibility (65).

Pathway Enrichment Analysis

Pathway enrichment analysis (PEA) was performed using Metascape (70). Top DEGs for each EC subtype or subtype-perturbation were sorted based on ascending p-value. Genes listed for each pathway were pulled from the Metacape results file, "_FINAL_GO.csv". For heatmaps produced by metascape, top 20 or 100 pathways were pulled from Metascape .png files, "HeatmapSelectedGO.png", "HeatmapSelectedGOParent.png", or "HeatmapSelectedGOTop100.png".

Motif Enrichment Analysis

A hypergeometric test was used to test for overrepresentation of each DNA motif in the set of differentially accessible peaks compared to a background set of peaks. We tested motifs present in the Jaspar database (2020 release) (52) by first identifying which peaks contained each motif using motifmatchr R package (https://bioconductor.org/packages/motifmatchr). We computed the GC content (percentage of G and C nucleotides) for each differentially accessible peak and sampled a background set of 40,000 peaks matched for GC content (51). Per-cell motif activity scores were computed by running chromVAR (88), and visualized using Seurat (47) function FeaturePlot.

Human Atherosclerosis scRNA-seq Public Data Download, Mapping, and Integration Across Samples

Count matrices of 17 samples taken from four different published scRNA-seq datasets were downloaded from the NCBI Gene Expression Omnibus (accessions listed in **Table S10** in the **Data Supplement**), processed using Cell Ranger (10x Genomics Cell Ranger 6.0.0) (89) with reference GRCh38 (version refdata-gex-GRCh38-2020-A, 10X Genomics), and analyzed using Seurat version 4.3.0 (47). Seurat objects were created from each dataset, and cells with < 500 counts and > 20% mitochondrial gene percent expression were excluded. Additionally, doublets were removed using DoubletFinder (90), which predicts doublets according to each real cell's proximity in gene expression space to artificial doublets created by averaging the transcriptional profile of randomly chosen cell pairs. Next, normalization and variance stabilization, followed by principal component (PC) analysis for 30 PCs were performed in Seurat (47) using default parameters. Batch effects across the 17 samples were corrected using Seurat functions (47) FindIntegrationAnchors using 10,000 anchors, followed by IntegrateData. During the integration step, cell cycle regression was performed by assigning cell cycle scores with Seurat (47) function CellCycleScoring. The *ex vivo* dataset first visualized, and canonical markers were identified for annotating cell types using FindAllMarkers.

Module Scoring

FindAllMarkers was used to identify top DEGs between each *ex vivo* cell subtype. Cells from the *in vitro* dataset were assigned an *ex vivo* cell subtype module score using Seurat (47) function AddModuleScore. The difference in module score between each *in vitro* EC subtype was established using Wilcoxon rank sum test with continuity correction and a two-sided alternative hypothesis.

Comparison of *Ex Vivo* snRNA-seq Data to *In Vitro* snRNA-seq Data

Meta-analyzed *ex vivo* human scRNA-seq data and *in vitro* snRNA-seq data were compared. Gene expression values for each *ex vivo* cell subtype and *in vitro* EC subtype-perturbation were produced using the AverageExpression function in Seurat (47) (which exponentiates log data, therefore output is depth normalized in non-log space). **Figure 6A** was generated using hclust function in R (91). Spearman correlation was used as the distance metric. Sample clustering was performed using all significant genes (adjusted p-value < 0.05) induced and attenuated across all *in vitro* EC subtypes for each pro-EndMT perturbation versus its respective control. **Figure S5A** was made using average expression data for marker genes for each *ex vivo* cell subtype. Hierarchical clustering across *ex vivo* cell subtypes was performed using hclust function in R (91), using average expression as the distance metric for a given gene.

GWAS SNP Enrichment Analysis

The SNPs associated with CAD were extracted from the most recent available meta-analysis (68). We utilized a matched background of SNPs pulled from 1000 Genomes Project reference panel (phase 3) (79) which were filtered using PLINK (92) v1.90b5.3 with the following settings: "--maf 0.01", "--geno 0.05". Mapping between genomic coordinates was performed using liftOver (80). To evaluate for enrichment in CAD-associated SNPs for each EC subtype and perturbation-subtype peak set, traseR package in R (traseR) (93) was used with the following: 'test.method' = "fisher", 'alternative' = "greater".

Peak-To-Gene Linkage

We estimated a linkage score for each peak-gene pair using the LinksPeaks function in Signac (51). For each gene, we computed the Pearson correlation coefficient r between the gene expression and the accessibility of each peak within 500 kb of the gene TSS. For each peak, we then computed a background set of expected correlation coefficients given properties of the peak by randomly sampling 200 peaks located on a different chromosome to the gene, matched for GC content, accessibility, and sequence length (MatchRegionStats function in Signac). We then computed the Pearson correlation between the expression of the gene and the set of background peaks. A z score was computed for each peak as $z = (r - \mu)/\sigma$, where μ was the background mean correlation coefficient and σ was the s.d. of the background correlation coefficients for the peak. We computed a P value for each peak using a one-sided z-test and retained peak-gene links with a p-value < 0.05 and a Pearson correlation coefficient. The results were restricted to peak regions which overlapped with significant CAD-associated SNPs (**Methods**, **"GWAS SNP Enrichment Analysis"**).

Data Visualization

Data visualizations were performed using Seurat functions DimPlot, DotPlot, FeaturePlot, and VlnPlot. Other data visualizations were performed using ggplot2 (for stacked bargaphs) (94), UpSetR (for UpSet plots)

(95), pheatmap (for DEG and DAR analysis heatmaps) and heatmap.2 (for Spearman's rank correlation coefficient heatmap and **Figure S5A**) (96).

DATA AVAILABILITY

Data produced in this study is made public in the GEO accession GSE228428.

ACKNOWLEDGEMENTS

Funding for this study was provided by grants from the National Insitututes of Health through R01HL147187 (CER), R35GM137896 (DAC), F30HL162469 (MLA), T32HL7249-45 (MLA), and from the Geneen Charitable Trust Awards Program for Coronary Heart Disease Research (CER).

REFERENCES

1. Brown JC, Gerhardt TE, Kwon E. Risk factors for coronary artery disease. StatPearls [Internet]. 2020.
2. Hajra L, Evans AI, Chen M, Hyduk SJ, Collins T, Cybulsky MI. The NF-κB signal transduction pathway in aortic endothelial cells is primed for activation in regions predisposed to atherosclerotic lesion formation. Proceedings of the National Academy of Sciences. 2000;97(16):9052-7.
3. Birdsey GM, Shah AV, Dufton N, Reynolds LE, Almagro LO, Yang Y, et al. The endothelial transcription factor ERG promotes vascular stability and growth through Wnt/β-catenin signaling. Developmental cell. 2015;32(1):82-96.
4. Rafieian-Kopaei M, Setorki M, Doudi M, Baradaran A, Nasri H. Atherosclerosis: process, indicators, risk factors and new hopes. International journal of preventive medicine. 2014;5(8):927.
5. Zhu Y, Xian X, Wang Z, Bi Y, Chen Q, Han X, et al. Research progress on the relationship between atherosclerosis and inflammation. Biomolecules. 2018;8(3):80.
6. Ahotupa M. Oxidized lipoprotein lipids and atherosclerosis. Free radical research. 2017;51(4):439-47.
7. Wesseling M, Sakkers T, De Jager S, Pasterkamp G, Goumans M. The morphological and molecular mechanisms of epithelial/endothelial-to-mesenchymal transition and its involvement in atherosclerosis. Vascular pharmacology. 2018;106:1-8.
8. Evrard SM, Lecce L, Michelis KC, Nomura-Kitabayashi A, Pandey G, Purushothaman K-R, et al. Endothelial to mesenchymal transition is common in atherosclerotic lesions and is associated with plaque instability. Nature communications. 2016;7(1):1-16.
9. Chen P-Y, Qin L, Baeyens N, Li G, Afolabi T, Budatha M, et al. Endothelial-to-mesenchymal transition drives atherosclerosis progression. The Journal of clinical investigation. 2015;125(12):4514-28.
10. van Meeteren LA, Ten Dijke P. Regulation of endothelial cell plasticity by TGF-β. Cell and tissue research. 2012;347(1):177-86.
11. Lutgens E, Gijbels M, Smook M, Heeringa P, Gotwals P, Koteliansky VE, et al. Transforming growth factor-β mediates balance between inflammation and fibrosis during plaque progression. Arteriosclerosis, thrombosis, and vascular biology. 2002;22(6):975-82.
12. Arciniegas E, Sutton AB, Allen TD, Schor AM. Transforming growth factor beta 1 promotes the differentiation of endothelial cells into smooth muscle-like cells in vitro. Journal of cell science. 1992;103(2):521-9.
13. Frid MG, Kale VA, Stenmark KR. Mature vascular endothelium can give rise to smooth muscle cells via endothelial-mesenchymal transdifferentiation: in vitro analysis. Circulation research. 2002;90(11):1189-96.
14. Ishisaki A, Hayashi H, Li A-J, Imamura T. Human umbilical vein endothelium-derived cells retain potential to differentiate into smooth muscle-like cells. Journal of Biological Chemistry. 2003;278(2):1303-9.
15. Krenning G, Moonen J-RA, van Luyn MJ, Harmsen MC. Vascular smooth muscle cells for use in vascular tissue engineering obtained by endothelial-to-mesenchymal transdifferentiation (EnMT) on collagen matrices. Biomaterials. 2008;29(27):3703-11.

16. Medici D, Potenta S, Kalluri R. Transforming growth factor-β2 promotes Snail-mediated endothelial–mesenchymal transition through convergence of Smad-dependent and Smad-independent signalling. Biochemical Journal. 2011;437(3):515-20.
17. Moonen J-RA, Krenning G, Brinker MG, Koerts JA, Van Luyn MJ, Harmsen MC. Endothelial progenitor cells give rise to pro-angiogenic smooth muscle-like progeny. Cardiovascular research. 2010;86(3):506-15.
18. Paranya G, Vineberg S, Dvorin E, Kaushal S, Roth SJ, Rabkin E, et al. Aortic valve endothelial cells undergo transforming growth factor-β-mediated and non-transforming growth factor-β-mediated transdifferentiation in vitro. The American journal of pathology. 2001;159(4):1335-43.
19. Bujak M, Dobaczewski M, Chatila K, Mendoza LH, Li N, Reddy A, et al. Interleukin-1 receptor type I signaling critically regulates infarct healing and cardiac remodeling. The American journal of pathology. 2008;173(1):57-67.
20. Bujak M, Frangogiannis NG. The role of IL-1 in the pathogenesis of heart disease. Archivum immunologiae et therapiae experimentalis. 2009;57(3):165-76.
21. Maleszewska M, Moonen J-RA, Huijkman N, van de Sluis B, Krenning G, Harmsen MC. IL-1β and TGFβ2 synergistically induce endothelial to mesenchymal transition in an NFκB-dependent manner. Immunobiology. 2013;218(4):443-54.
22. Chaudhuri V, Zhou L, Karasek M. Inflammatory cytokines induce the transformation of human dermal microvascular endothelial cells into myofibroblasts: a potential role in skin fibrogenesis. Journal of cutaneous pathology. 2007;34(2):146-53.
23. Sánchez-Duffhues G, García de Vinuesa A, van de Pol V, Geerts ME, de Vries MR, Janson SG, et al. Inflammation induces endothelial-to-mesenchymal transition and promotes vascular calcification through downregulation of BMPR2. The Journal of pathology. 2019;247(3):333-46.
24. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. New England journal of medicine. 2017;377(12):1119-31.
25. Sperone A, Dryden NH, Birdsey GM, Madden L, Johns M, Evans PC, et al. The transcription factor Erg inhibits vascular inflammation by repressing NF-κB activation and proinflammatory gene expression in endothelial cells. Arteriosclerosis, thrombosis, and vascular biology. 2011;31(1):142-50.
26. Fish JE, Cantu Gutierrez M, Dang LT, Khyzha N, Chen Z, Veitch S, et al. Dynamic regulation of VEGF-inducible genes by an ERK/ERG/p300 transcriptional network. Development. 2017;144(13):2428-44.
27. Lathen C, Zhang Y, Chow J, Singh M, Lin G, Nigam V, et al. ERG-APLNR axis controls pulmonary venule endothelial proliferation in pulmonary veno-occlusive disease. Circulation. 2014;130(14):1179-91.
28. Vijayaraj P, Le Bras A, Mitchell N, Kondo M, Juliao S, Wasserman M, et al. Erg is a crucial regulator of endocardial-mesenchymal transformation during cardiac valve morphogenesis. Development. 2012;139(21):3973-85.
29. Hogan NT, Whalen MB, Stolze LK, Hadeli NK, Lam MT, Springstead JR, et al. Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. Elife. 2017;6:e22536.
30. Dufton NP, Peghaire CR, Osuna-Almagro L, Raimondi C, Kalna V, Chauhan A, et al. Dynamic regulation of canonical TGFβ signalling by endothelial transcription factor ERG protects from liver fibrogenesis. Nature communications. 2017;8(1):1-14.
31. Nagai N, Ohguchi H, Nakaki R, Matsumura Y, Kanki Y, Sakai J, et al. Downregulation of ERG and FLI1 expression in endothelial cells triggers endothelial-to-mesenchymal transition. PLoS genetics. 2018;14(11):e1007826.
32. Caporarello N, Lee J, Pham TX, Jones DL, Guan J, Link PA, et al. Dysfunctional ERG signaling drives pulmonary vascular aging and persistent fibrosis. Nature communications. 2022;13(1):4170.
33. Peghaire C, Dufton N, Lang M, Salles-Crawley I, Ahnström J, Kalna V, et al. The transcription factor ERG regulates a low shear stress-induced anti-thrombotic pathway in the microvasculature. Nature communications. 2019;10(1):1-17.
34. Kalucka J, de Rooij LP, Goveia J, Rohlenova K, Dumas SJ, Meta E, et al. Single-cell transcriptome atlas of murine endothelial cells. Cell. 2020;180(4):764-79. e20.

35. Rohlenova K, Goveia J, García-Caballero M, Subramanian A, Kalucka J, Treps L, et al. Single-cell RNA sequencing maps endothelial metabolic plasticity in pathological angiogenesis. Cell metabolism. 2020;31(4):862-77. e14.
36. Kalluri AS, Vellarikkal SK, Edelman ER, Nguyen L, Subramanian A, Ellinor PT, et al. Single-cell analysis of the normal mouse aorta reveals functionally distinct endothelial cell populations. Circulation. 2019;140(2):147-63.
37. Zhao G, Lu H, Chang Z, Zhao Y, Zhu T, Chang L, et al. Single-cell RNA sequencing reveals the cellular heterogeneity of aneurysmal infrarenal abdominal aorta. Cardiovascular research. 2021;117(5):1402-16.
38. Xu K, Xie S, Huang Y, Zhou T, Liu M, Zhu P, et al. Cell-type transcriptome atlas of human aortic valves reveal cell heterogeneity and endothelial to mesenchymal transition involved in calcific aortic valve disease. Arteriosclerosis, thrombosis, and vascular biology. 2020;40(12):2910-21.
39. Cheng J, Gu W, Lan T, Deng J, Ni Z, Zhang Z, et al. Single-cell RNA sequencing reveals cell type-and artery type-specific vascular remodelling in male spontaneously hypertensive rats. Cardiovascular Research. 2021;117(4):1202-16.
40. Khan S, Taverna F, Rohlenova K, Treps L, Geldhof V, de Rooij L, et al. EndoDB: a database of endothelial cell transcriptomics data. Nucleic acids research. 2019;47(D1):D736-D44.
41. Li Z, Solomonidis EG, Meloni M, Taylor RS, Duffin R, Dobie R, et al. Single-cell transcriptome analyses reveal novel targets modulating cardiac neovascularization by resident endothelial cells following myocardial infarction. European heart journal. 2019;40(30):2507-20.
42. Andueza A, Kumar S, Kim J, Kang D-W, Mumme HL, Perez JI, et al. Endothelial reprogramming by disturbed flow revealed by single-cell RNA and chromatin accessibility study. Cell reports. 2020;33(11):108491.
43. Tombor L, John D, Glaser S, Luxan G, Forte E, Furtado M, et al. Single cell sequencing reveals endothelial plasticity with transient mesenchymal activation after myocardial infarction. European Heart Journal. 2020;41(Supplement_2):ehaa946. 3736.
44. Zhao Q, Eichten A, Parveen A, Adler C, Huang Y, Wang W, et al. Single-cell transcriptome analyses reveal endothelial cell heterogeneity in tumors and changes following antiangiogenic treatment. Cancer research. 2018;78(9):2370-82.
45. Navab M, Hough GP, Stevenson LW, Drinkwater DC, Laks H, Fogelman AM. Monocyte migration into the subendothelial space of a coculture of adult human aortic endothelial and smooth muscle cells. The Journal of clinical investigation. 1988;82(6):1853-63.
46. Genomics x. Chromium Next GEM Single Cell Multiome ATAC + Gene Expression. Revision F edAugust 2022.
47. Hao Y, Hao S, Andersen-Nissen E, Mauck III WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87. e29.
48. Dahal S, Huang P, Murray BT, Mahler GJ. Endothelial to mesenchymal transformation is induced by altered extracellular matrix in aortic valve endothelial cells. Journal of Biomedical Materials Research Part A. 2017;105(10):2729-41.
49. Kovacic JC, Dimmeler S, Harvey RP, Finkel T, Aikawa E, Krenning G, et al. Endothelial to mesenchymal transition in cardiovascular disease: JACC state-of-the-art review. Journal of the American College of Cardiology. 2019;73(2):190-209.
50. Bondareva O, Rodríguez-Aguilera JR, Oliveira F, Liao L, Rose A, Gupta A, et al. Single-cell profiling of vascular endothelial cells reveals progressive organ-specific vulnerabilities during obesity. Nature Metabolism. 2022:1-20.
51. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. Nature methods. 2021;18(11):1333-41.
52. Fornes O, Castro-Mondragon JA, Khan A, Van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic acids research. 2020;48(D1):D87-D92.
53. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. Nature medicine. 2019;25(8):1280-9.
54. Örd T, Õunap K, Stolze LK, Aherrahrou R, Nurminen V, Toropainen A, et al. Single-cell epigenomics and functional fine-mapping of atherosclerosis GWAS loci. Circulation research. 2021;129(2):240-58.

55. Zhang L, Tang C, Zhang M, Tong X, Xie Y, Yan R, et al. Single cell meta-analysis of EndMT and EMT state in COVID-19. Frontiers in immunology. 2022;13.
56. Deng G, Zhang L, Wang C, Wang S, Xu J, Dong J, et al. AGEs-RAGE axis causes endothelial-to-mesenchymal transition in early calcific aortic valve disease via TGF-β1 and BMPR2 signaling. Experimental Gerontology. 2020;141:111088.
57. Chowdhury RR, D'Addabbo J, Huang X, Veizades S, Sasagawa K, Louis DM, et al. Human Coronary Plaque T Cells Are Clonal and Cross-React to Virus and Self. Circulation Research. 2022;130(10):1510-30.
58. Alsaigh T, Evans D, Frankel D, Torkamani A. Decoding the transcriptome of calcified atherosclerotic plaque at single-cell resolution. Communications biology. 2022;5(1):1-17.
59. Pan H, Xue C, Auerbach BJ, Fan J, Bashore AC, Cui J, et al. 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. 2020;142(21):2060-75.
60. Ekiz HA, Conley CJ, Stephens WZ, O'Connell RM. CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments. BMC bioinformatics. 2020;21(1):1-15.
61. Ricciotti E, FitzGerald GA. Prostaglandins and inflammation. Arteriosclerosis, thrombosis, and vascular biology. 2011;31(5):986-1000.
62. Levi M, van der Poll T, Büller HR. Bidirectional relation between inflammation and coagulation. Circulation. 2004;109(22):2698-704.
63. Ni C, Buszczak M, editors. The homeostatic regulation of ribosome biogenesis. Seminars in Cell & Developmental Biology; 2022: Elsevier.
64. Suárez Y, Fernández-Hernando C, Pober JS, Sessa WC. Dicer dependent microRNAs regulate gene expression and functions in human endothelial cells. Circulation research. 2007;100(8):1164-73.
65. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 1995;57(1):289-300.
66. Drobni ZD, Kolossvary M, Karady J, Jermendy AL, Tarnoki AD, Tarnoki DL, et al. Heritability of coronary artery disease: Insights from a classical twin study. Circulation: Cardiovascular Imaging. 2022;15(3):e013348.
67. McPherson R, Tybjaerg-Hansen A. Genetics of coronary artery disease. Circulation research. 2016;118(4):564-78.
68. Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. Nature medicine. 2022;28(8):1679-92.
69. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science. 2018;361(6409):1380-5.
70. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature communications. 2019;10(1):1-10.
71. Liu T, Zou X-Z, Huang N, Ge X-Y, Yao M-Z, Liu H, et al. miR-27a promotes endothelial-mesenchymal transition in hypoxia-induced pulmonary arterial hypertension by suppressing BMP signaling. Life sciences. 2019;227:64-73.
72. Yang W, Ng FL, Chan K, Pu X, Poston RN, Ren M, et al. Coronary-heart-disease-associated genetic variant at the COL4A1/COL4A2 locus affects COL4A1/COL4A2 expression, vascular cell survival, atherosclerotic plaque stability and risk of myocardial infarction. PLoS genetics. 2016;12(7):e1006127.
73. Woodfin A, Voisin M-B, Nourshargh S. PECAM-1: a multi-functional molecule in inflammation and vascular biology. Arteriosclerosis, thrombosis, and vascular biology. 2007;27(12):2514-23.
74. Rodor J, Chen SH, Scanlon JP, Monteiro JP, Caudrillier A, Sweta S, et al. Single-cell RNA sequencing profiling of mouse endothelial cells in response to pulmonary arterial hypertension. Cardiovascular Research. 2022;118(11):2519-34.
75. Stolze LK, Conklin AC, Whalen MB, Rodríguez ML, Õunap K, Selvarajan I, et al. Systems Genetics in Human Endothelial Cells Identifies Non-coding Variants Modifying Enhancers, Expression, and Complex Disease Traits. The American Journal of Human Genetics. 2020;106(6):748-63.
76. Toropainen A, Stolze LK, Örd T, Whalen MB, Torrell PM, Link VM, et al. Functional noncoding SNPs in human endothelial cells fine-map vascular trait associations. Genome Research. 2022;32(3):409-24.

77. Genomics x. Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing. Revision C ed2022.
78. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics. 2009;5(6):e1000529.
79. Consortium GP, Auton A, Brooks L, Durbin R, Garrison E, Kang H. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
80. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Briefings in bioinformatics. 2013;14(2):144-61.
81. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8.
82. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
83. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nature biotechnology. 2018;36(1):89-94.
84. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular systems biology. 2019;15(6):e8746.
85. Heidecker B, Lamirault G, Kasper EK, Wittstein IS, Champion HC, Breton E, et al. The gene expression profile of patients with new-onset heart failure reveals important gender-specific differences. European heart journal. 2010;31(10):1188-96.
86. Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome biology. 2020;21:1-16.
87. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):1-21.
88. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nature methods. 2017;14(10):975-8.
89. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature communications. 2017;8(1):14049.
90. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell systems. 2019;8(4):329-37. e4.
91. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? Journal of classification. 2014;31(3):274-95.
92. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 2007;81(3):559-75.
93. Chen L, Qin ZS. traseR: an R package for performing trait-associated SNP enrichment analysis in genomic intervals. Bioinformatics. 2016;32(8):1214-6.
94. Villanueva RAM, Chen ZJ. ggplot2: elegant graphics for data analysis. Taylor & Francis; 2019.
95. Conway EM, Collen D, Carmeliet P. Molecular mechanisms of blood vessel growth. Cardiovascular research. 2001;49(3):507-21.
96. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, et al. Gplots: various R programming tools for plotting data. 2016. R package version. 2014;2(0).

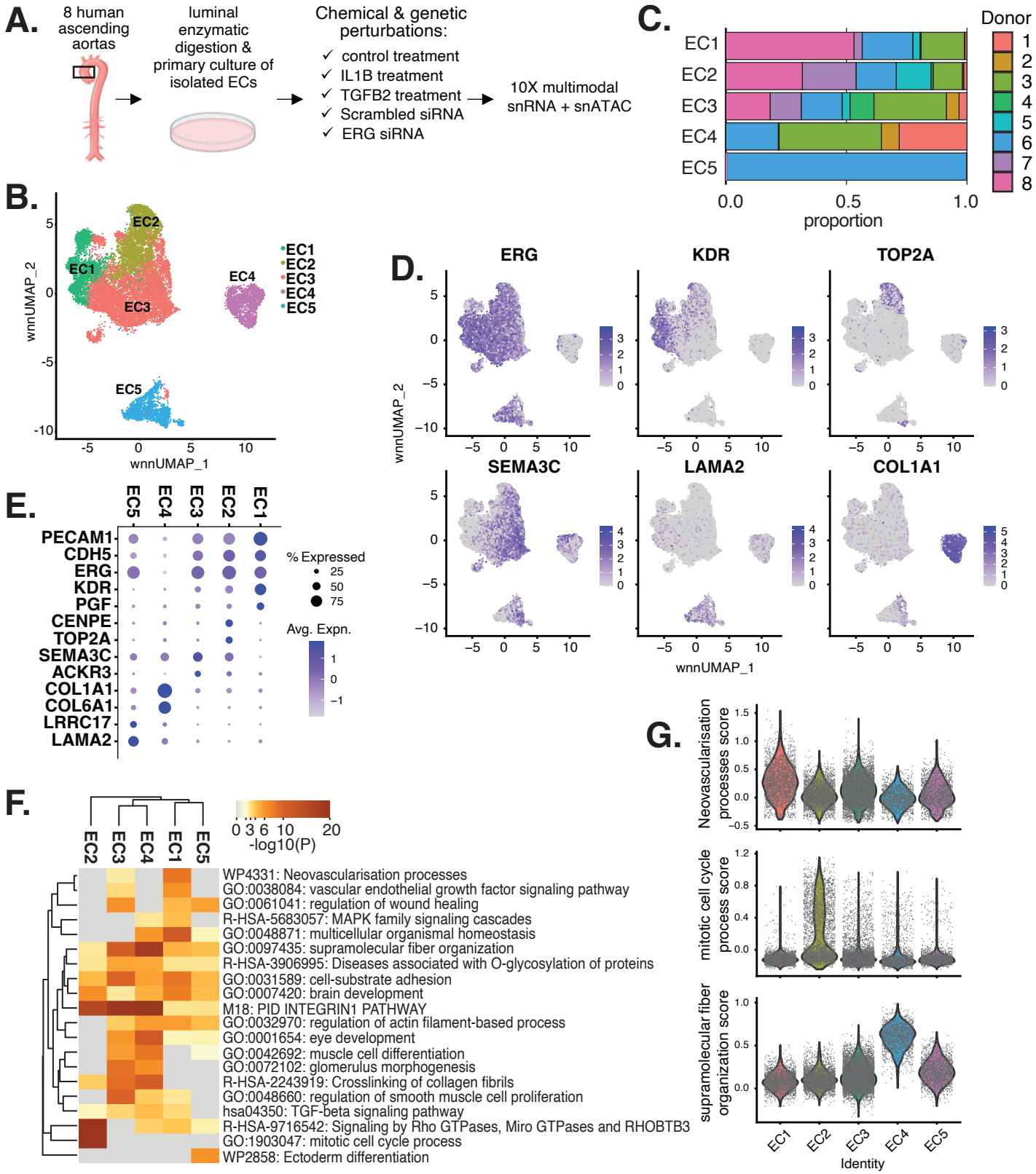

Figure 1 | EC transcriptomic profiles reveal a heterogenous population. (A), Schematic diagram of the experimental design. ECs are isolated from eight donor human ascending aortic trimmings and treated with IL1B, TGFB2, or siERG (ERG siRNA) to induce EndMT. **(B)**, Weighted Nearest Neighbor UMAP (WNNUMAP) of aggregate cells from all perturbations and donors. Briefly, the WNN procedure implemented in Seurat v4 is designed to integrate multiple types of data that are collected in the same cells to define a single unified representation of single-cell multimodal data. Each dot represents a cell, and the proximity between each cell represents the similarity between both transcriptional and epigenetic profiles. Colors denote different clusters. **(C)**, Stacked bargraph indicating proportion of cells from each donor for each EC subtype. **(D)**, Feature plot displaying gene expression across cells of top markers for pan EC (*ERG*), EC1 (*KDR*), EC2 (*TOP2A*), EC3 (*SEMA3C*), EC4 (*LAMA2*), and EC4 (*COL1A1*). **(E)**, Dot plot of top markers for pan EC (*PECAM1*, *CDH5*, *ERG*), EC1 (*KDR*, *PGF*), EC2 (*CENPE*, *TOP2A*), EC3 (*SEMA3C*, *ACKR3*), EC4 (*COL1A1*, *COL6A1*), and EC5 (*LRRC17*, *LAMA2*). The size of the dot represents the percentage of cells within each EC subtype that express the given gene, while the shade of the dot represents the level of average expression ("Avg. Expn." in the legend). **(F)**, Heatmap of pathway enrichment analysis (PEA) results from submitting top 200 differentially expressed genes (DEGs; by ascending p-value) between EC subtypes. Rows (pathways) and columns (EC subtypes) are clustered based on -Log_{10}(P) **(G)**, Violin plots of top Metascape pathway module scores across EC subtypes. Module scores are generated for each cell barcode with Seurat function AddModuleScore.

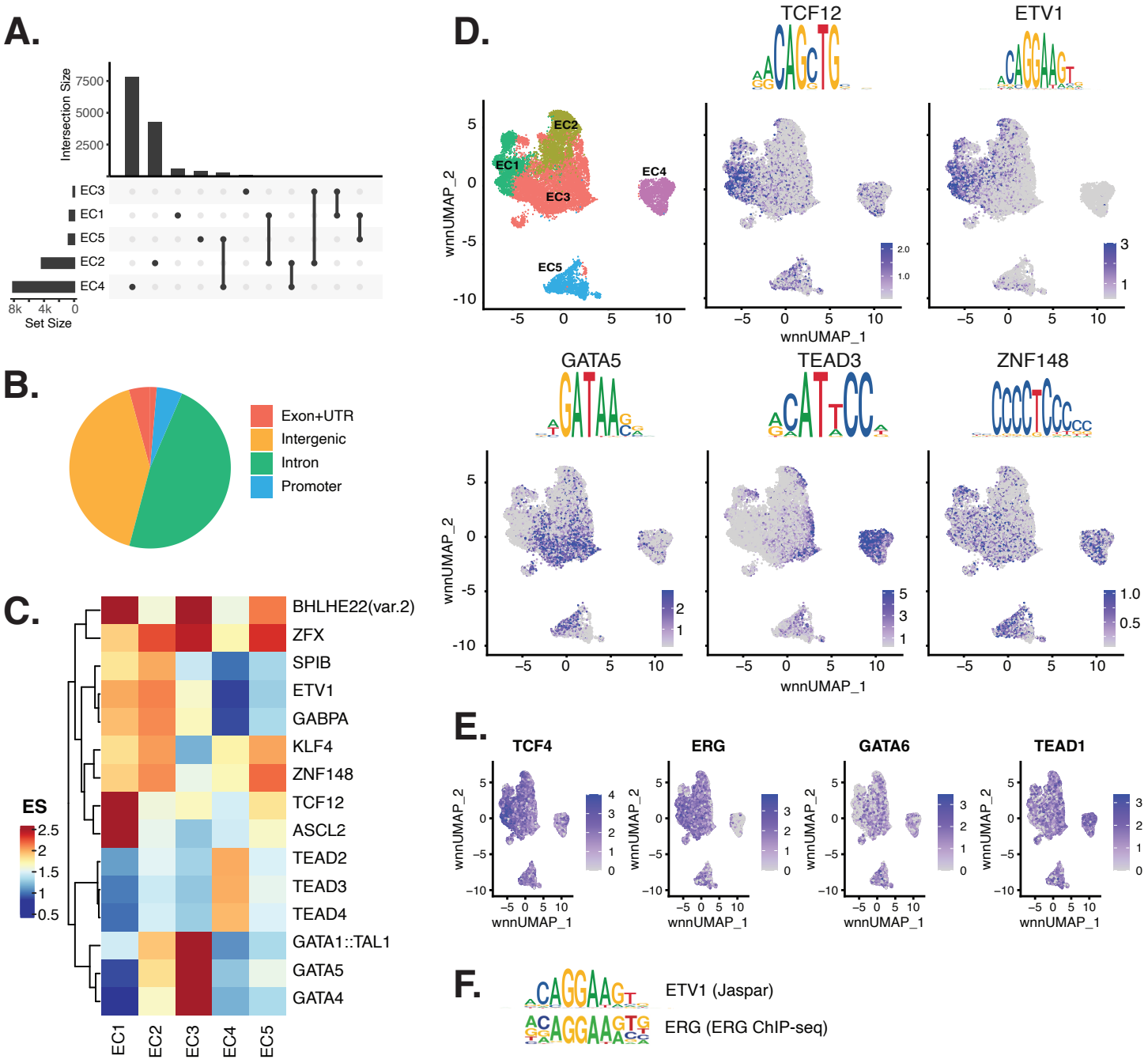

Figure 2 | ECs have epigenetically distinct cell states.
(A), Upset plot of differential peaks across EC subtypes. Upset plots are a data visualization method for showing set data with more than three intersecting sets. Intersection size represents the number of genes at each intersection, while set size represents the number of genes for each EC subtype. **(B)**, Pie chart of complete peak set indicating peaks located in exon or UTR, intergenic, intronic or promoter regions. **(C)**, Heatmap of top transcription factors (TFs) from motif enrichment analysis for each EC subtype. Top TFs for each EC subtype are selected based on ascending p-value. Rows (TFs) and columns (EC subtype) are clustered based on enrichment score (ES). **(D)**, Feature plots and position weight matrices (PWMs) for top TF binding motifs for EC1 (TCF12), EC2 (ETV1), EC3 (GATA5), and EC4 (TEAD3). Per-cell motif activity scores are computed with chromVAR, and motif activities per cell are visualized using Signac function FeaturePlot. **(E)**, Feature plots of gene expression profiles from snRNA data of candidate TFs binding to the respective motifs in D. **(F)**, PWMs comparing Jaspar 2020 ETV1 motif to ERG motif reported in Hogan et al.

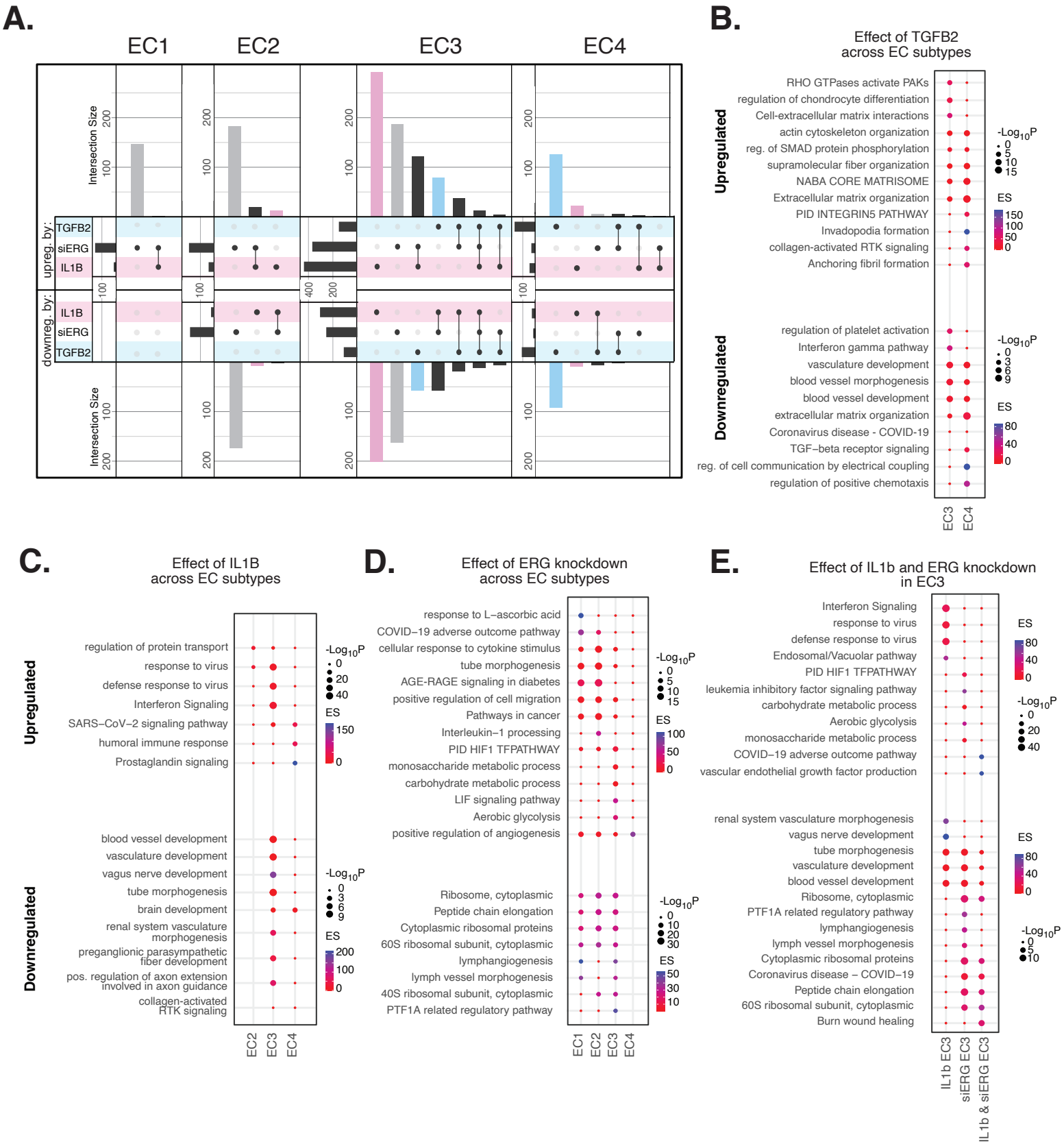

Figure 3 | EndMT perturbations *in vitro* elicit EC subtype-specific transcriptional responses. (A), Upset plots of up- and down-regulated DEGs across EC subtypes with siERG (grey), IL1B (pink), and TGFB (blue). Intersection size represents the number of genes at each intersection. **(B)**, PEA for EC3-4 up- and down-regulated DEGs with TGFB2 compared to control media. **(C)**, PEA for EC2-4 up- and down-regulated DEGs with IL1B compared to control media. **(D)**, PEA for EC1-4 up- and down-regulated DEGs with siERG compared to siSCR. **(E)**, PEA comparing up- and down-regulated DEGs that are mutually exclusive and shared between IL1B and siERG in EC3.

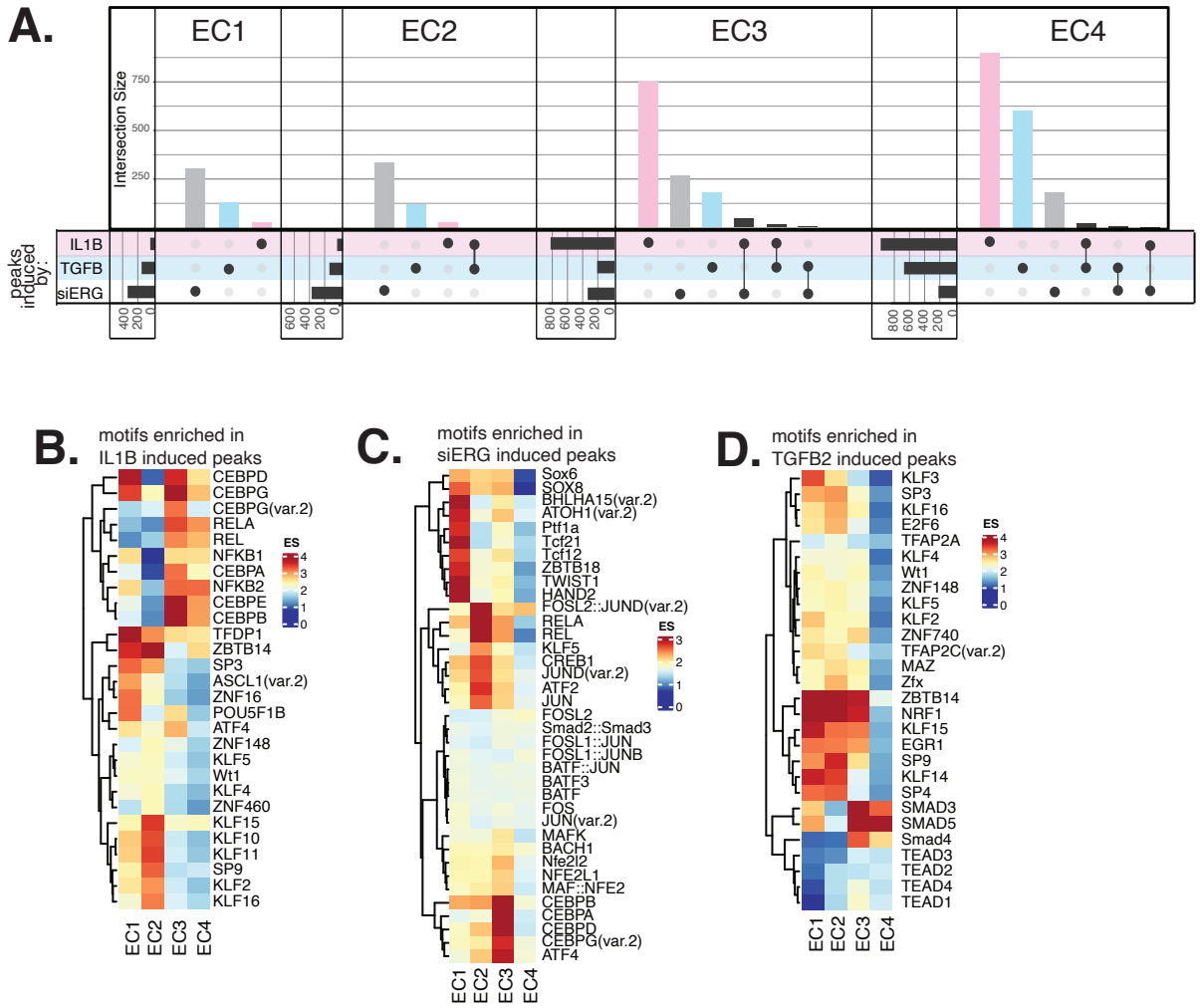

Figure 4 | *In vitro* EC EndMT models reorganize active epigenetic landscapes with subtype specificity. (A), Upset plot of induced peaks for siERG (grey), IL1B (pink), and TGFB2 (blue) across EC1-4. Intersection size represents the number of genes at each intersection. **(B)**, Heatmap of top motifs enriched in IL1B-affected peaks. **(C)**, Heatmap of top motifs enriched in siERG-affected peaks. **(D)**, Heatmap of top motifs enriched in TGFB2-affected peaks. For B-D, top TFs for each EC subtype are selected based on ascending p-value. Rows (TFs) and columns (EC subtypes) are clustered based on enrichment score (ES).

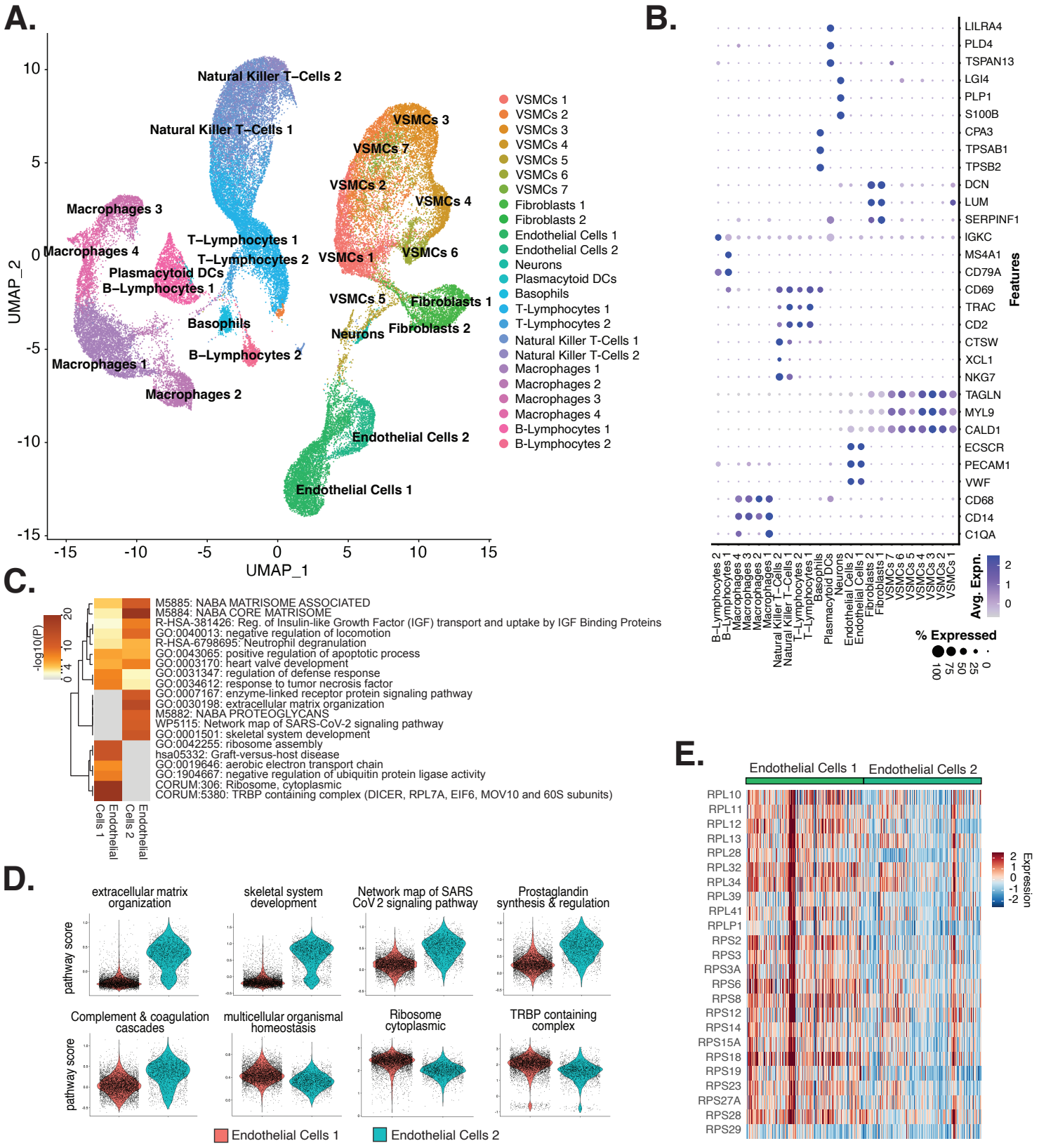

Figure 5 | ECs *ex vivo* human atherosclerotic plaques are distinguished by EndMT. (A), scRNA-seq UMAP of different cell subtypes across 17 samples of *ex vivo* human atherosclerotic plaques. **(B)**, Dot plot of top markers for each cell type. The size of the dot represents the percentage of cells within each cell subtype that express the given gene, while the shade of the dot represents the level of average expression ("Avg. Expn." in the legend). **(C)**, Heatmap of pathway enrichment analysis (PEA) results generated from submitting 200 differentially expressed genes (DEGs) between Endothelial Cells 1 (Endo1) and Endothelial Cells 2 (Endo2). Rows (pathways) and columns (cell subtypes) are clustered based on -Log_{10}(P). **(E)**, Heatmap displaying expression of genes belonging to ribosome cytoplasmic pathway for Endo1 and Endo2.

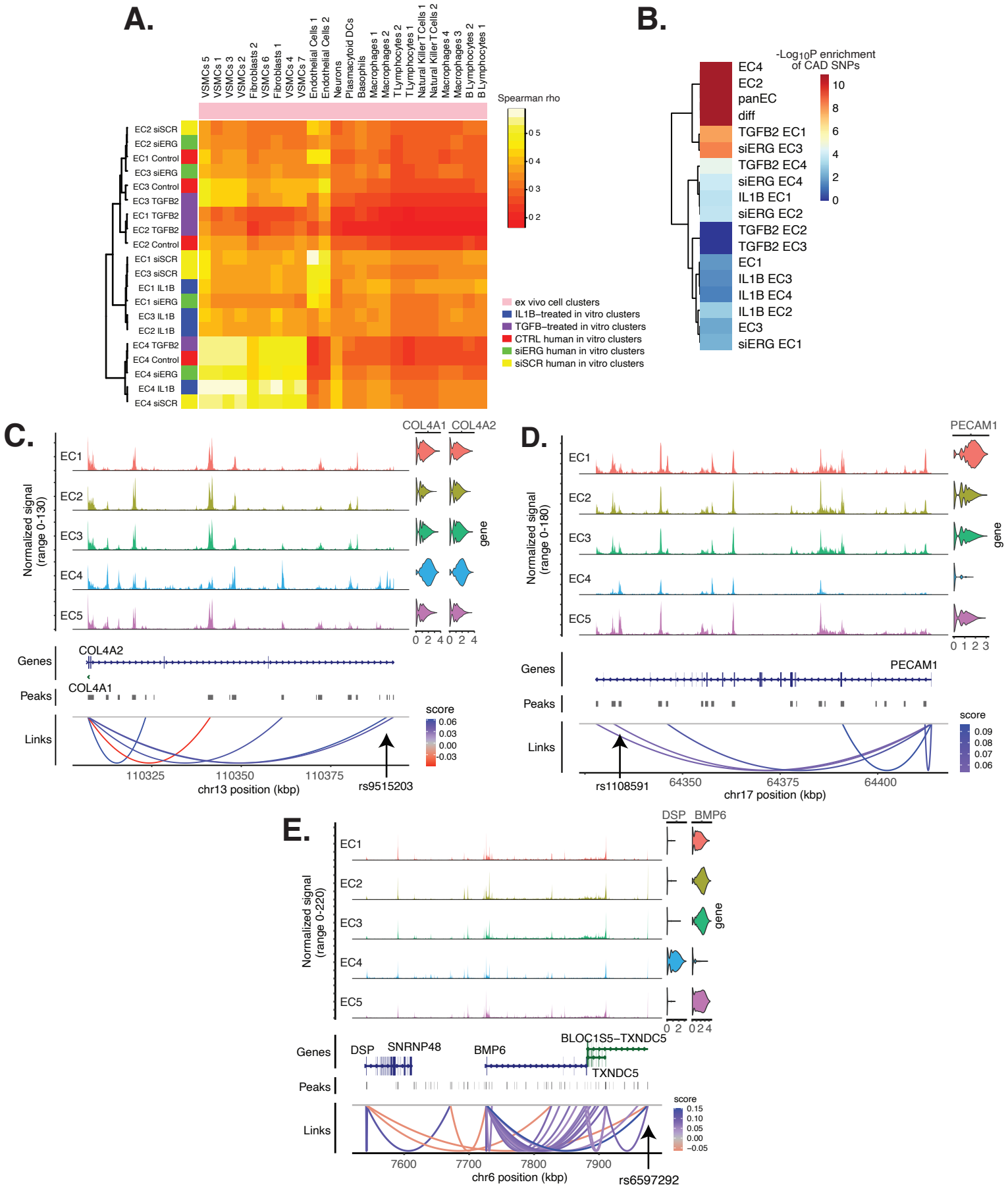

Figure 6 | EC subtype is a major determinant in the ability to recapitulate 'omic profiles seen in atherosclerosis. (A), Spearman correlation heatmap displaying average expression between *in vitro* perturbation-subtype combinations and *ex vivo* cell subtypes using all up- and down-regulated genes between IL1B, TGFB2, or siERG versus control. Spearman correlation was used as the distance metric. Rows (*in vitro* EC subtypes) and columns (*ex vivo* cell subtypes) are clustered using all significant genes (adjusted p-value < 0.05) induced and attenuated across all *in vitro* EC subtypes for each perturbation versus its respective control. **(B)**, Heatmap of CAD-associated SNP enrichments across *in vitro* EC subtypes and perturbation-subtype combinations. Rows (EC subtypes and perturbation-subtype combinations) are clustered using $-\text{Log}_{10}(P)$ for enrichment in significant CAD-associated SNPs (p-value $< 5\text{x}10^{-8}$). Note that "diff" represents peaks common to more than one EC subtype; it is found by subtracting EC1-5 subtype-specific peaks from the entire *in vitro* peak set (termed "panEC"). **(C)**, Coverage plots displaying links for COL4A1/COL4A2 genes to EC4-specific peaks, including one overlapping with CAD-associated SNP rs9515203. **(D)**, Coverage plot showing links for PECAM1 gene to EC4-specific peaks, including one overlapping with CAD-associated SNP rs1108591. **(E)**, Coverage plot showing links for DSP and BMP6 genes to EC4-specific peaks, including one overlapping with CAD-associated SNP rs6597292.