Schulz *et al. Genome Medicine* (2026) 18:53
https://doi.org/10.1186/s13073-026-01638-6

Genome Medicine

RESEARCH Open Access

ERG is a regulator of dynamic and reversible endothelial plasticity

Kristen Schulz[1,2†], Steven R. Botts[2,3†], Kai Ellis[4,5†], Corey A. Scipione[2], Nadiya Khyzha[1,6], Christy Ho[1,2], Rathnakumar Kumaragurubaran[2,7], Kyoko E. Yuki[4], Joshua D. Wythe[8,9,10,11], Clint L. Miller[9,12,13], Michael D. Wilson[4,5], Kathryn L. Howe[1,2,3,14,15,16*†] and Jason E. Fish[1,2,3,15,17*†]

Abstract

Background Endothelial cells (ECs) orchestrate vascular homeostasis and resilience but can undergo reprogramming into a mesenchymal-like phenotype through an endothelial-to-mesenchymal transition (EndMT). Crucially, EndMT is a linchpin underlying several cardiometabolic diseases, but is almost universally studied as an endpoint. The transcription factor ERG (ETS-related gene) is critical to the maintenance of EC identity and function, yet the dynamic transcriptional and functional consequences of ERG loss on EndMT programs, and whether this can be reversed, has not been explored.

Methods We modeled both acute and chronic ERG loss in human aortic ECs using siRNA knockdown and CRISPR/Cas9-mediated *ERG* deletion. We profiled temporal changes in chromatin accessibility (ATAC-seq), transcriptomic responses (RNA-seq), and endothelial phenotypes, including migration and barrier integrity. The temporal kinetics of ERG loss and restoration was assessed by comparing stable *ERG* knockout to transient *ERG* knockdown and recovery over time. The implications to human disease were deciphered by examining ERG gene regulatory networks in human atherosclerosis and linkage with genetic variation associated with human cardiovascular disease.

Results Analysis of gene regulatory networks revealed profound and dynamic rewiring of endothelial and mesenchymal transcriptional programs upon loss of ERG. While endothelial identity was rapidly lost by 24 h of *ERG* knockdown, acquisition of mesenchymal identity, barrier dysfunction, and enhanced cell migration required 72 h to manifest. Loss of ERG was accompanied by a rapid reduction in accessibility of ETS motifs and an extensive gain in open chromatin containing AP1 motifs. Disease-relevant endothelial dysfunction programs were associated with dynamically reorganized transcriptional networks. Importantly, restoration of ERG expression reversed EndMT gene regulatory networks and phenotypes.

†Kristen Schulz, Steven R. Botts and Kai Ellis contributed equally to this work.

†Kathryn L. Howe and Jason E. Fish are co-senior and co-corresponding authors.

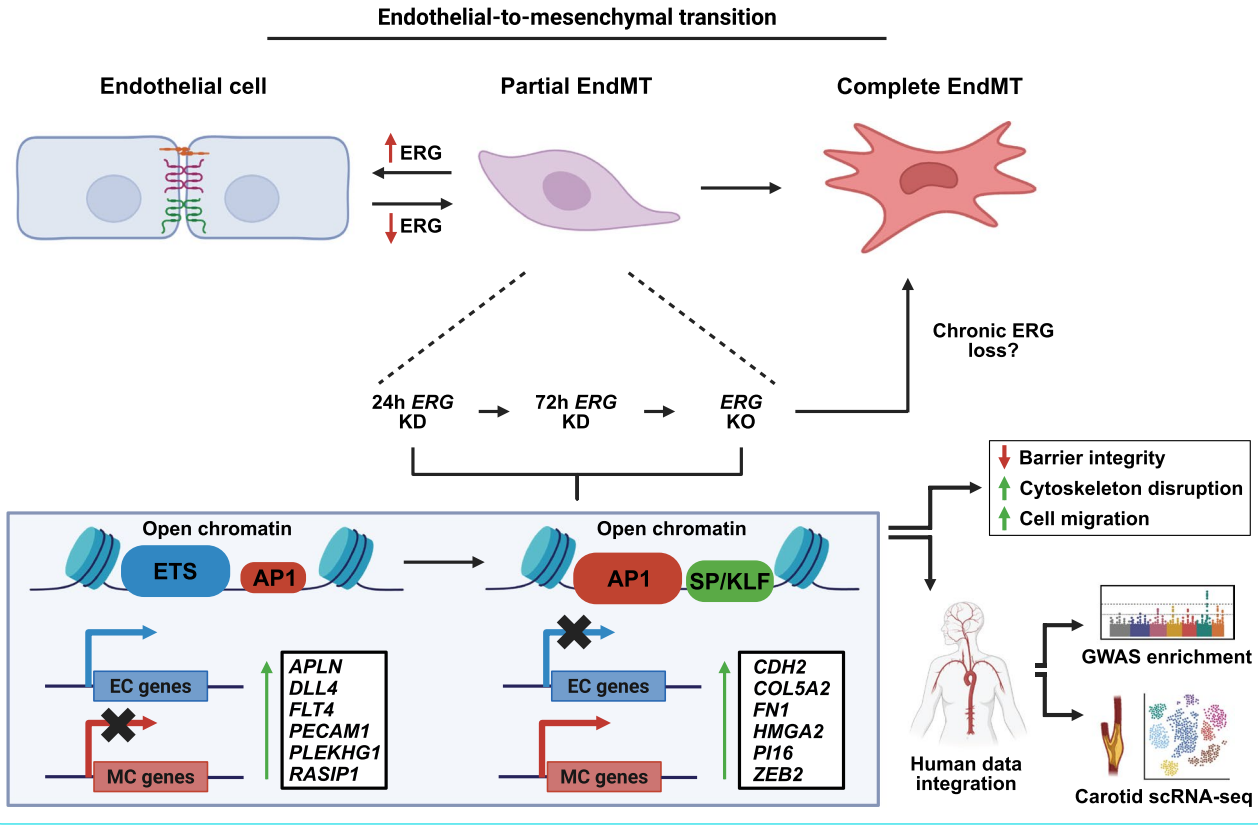
*Correspondence:
Kathryn L. Howe
kathryn.howe@uhn.ca
Jason E. Fish
jason.fish@utoronto.ca

Full list of author information is available at the end of the article

Conclusions Overall, this study highlights the ETS factor, ERG, as an essential transcriptional safeguard of endothelial identity and function, and demonstrates that ERG loss initiates a progressive, yet reversible, EndMT program with EC identity loss preceding a gain of mesenchymal gene regulatory networks and phenotypes. This study establishes loss of ERG as an early initiating event in EndMT and suggests that ERG-targeted therapies may hold promise for promoting endothelial resilience.

Keywords Endothelial-to-mesenchymal transition, Endothelial function, Endothelial cell identity, Gene regulatory networks, Cardiovascular disease

Graphical Abstract



Background

Endothelial cells (ECs) form a monolayer lining the vessel wall and act as a highly resilient interface and physiologic barrier between flowing blood and vascular tissue. However, studies have shown that ECs can lose their protective properties when they become activated in response to pathological insults such as disturbed flow [1], dyslipidemia [2], hyperglycemia [3], smoking [4], or other pro-inflammatory signals (e.g., TNF-α and IL-1β) [5]. While it was once thought that adult ECs are terminally differentiated, recent lineage tracing studies have revealed that ECs can undergo endothelial-to-mesenchymal transition (EndMT), delaminate from the lumen, and migrate into the vessel wall [6]. Importantly, the full extent of potential phenotypes that these EC-derived cells adopt is not fully known and it remains unclear whether the process of EndMT is detrimental or beneficial in the context of arterial diseases [7]. Conflicting studies report that cells that have undergone EndMT can stabilize the fibrous cap of an atherosclerotic plaque or adopt a pro-inflammatory or immune cell-like phenotype [8, 9]. The kinetics and reversibility of EndMT is also unclear as both partial and full EndMT have been reported [10, 11]. Understanding the mechanisms governing dynamic regulation of endothelial phenotypes is key to implementing strategies to maintain or restore vascular resilience to combat cardiovascular disease [12].

The endothelial-enriched transcription factor E-26 transformation specific (ETS)-related gene (ERG) is a key regulator of EC identity and protective functions [13–16]. ERG maintains the expression of canonical endothelial marker genes, such as VE-Cadherin (*CDH5*) and PECAM-1 (*CD31*), while suppressing NF-κB mediated pro-inflammatory pathways [17]. Recently, exogenous

ERG and SOX17 expression in fibroblasts was found to promote an endothelial-like state, inducing EC marker expression and EC function [18, 19]. Loss of endothelial ERG promotes TGF-β/SMAD3 signaling, which drives EC dysfunction and EndMT [20]. Notably, ERG is downregulated in several chronic inflammatory and cardiometabolic diseases such as atherosclerosis, liver fibrogenesis and pulmonary hypertension [13, 20, 21]. Given that ERG loss may be a linchpin event in cardiovascular disease pathogenesis, there is a need to define how ERG prevents EndMT to maintain endothelial identity and function, and whether EndMT induced by ERG loss can be reversed in human disease models.

Herein, we investigate the role of ERG in maintaining endothelial identity by examining dynamic changes in EC chromatin accessibility and transcriptional landscapes and their impact on endothelial phenotypes and function following ERG loss. We identify a progressive transition of ECs towards a mesenchymal phenotype, which involves an acute loss of endothelial identity markers 24 h after *ERG* knockdown (KD) accompanied by a robust increase in mesenchymal gene expression, barrier dysfunction, and cell migration 72 h after *ERG* KD. Mechanistically, ATAC-seq reveals that ERG loss induces a prominent closure of chromatin regions containing ETS motifs and a gain in accessibility of chromatin containing activator protein 1 (AP1) motifs. Integration of transcriptional profiling data with genome-wide association studies (GWAS) and single-cell RNA sequencing (scRNA-seq) data from human atherosclerotic plaque samples reveals that ERG-mediated EndMT programs are prominent in diseases associated with endothelial dysfunction. Importantly, we show that restoring ERG expression reverses mesenchymal gene regulatory networks and phenotypes induced by *ERG* KD and resets EC identity programs. Collectively, these findings implicate ERG in the maintenance of EC identity and homeostasis in healthy vessels and highlight a novel role for ERG in reversing the pathological consequences of EndMT and EC dysfunction in the setting of cardiovascular disease.

Methods

Cohorts used in the study

For phenotypic assays, western blots and immunofluorescence, three wild-type (WT) and three Δ*ERG* knockout (KO) telomerase-immortalized human aortic endothelial cell (teloHAEC) clones were used. For ATAC-seq, two WT and two Δ*ERG* KO clones were used. For RNA-seq, three WT and three Δ*ERG* KO clones were used with biological replicates performed on separate days. After removing one sample because of poor quality, a total of five replicates for WT and six replicates for Δ*ERG* KO were included for RNA-seq analysis. Within each biological replicate of the control and *ERG* knockdown (KD) experiments in teloHAECs, all three readouts (protein, RNA-seq and ATAC-seq) were collected from different wells of a 12-well plate of cells treated identically. The number of replicates for protein, RNA-seq and ATAC-seq were five, four and three, respectively. For phenotypic assays, western blots and immunofluorescence, three

replicates were performed on different days. For ERG recovery experiments all three readouts (protein, RNA-seq and ATAC-seq) were collected from different wells of a 12-well plate of cells treated identically. The number of replicates for protein, RNA-seq and ATAC-seq were five, four and three, respectively. For phenotypic assays, western blots and immunofluorescence, three replicates were performed on different days. For qPCR, five replicates were performed on different days. Single cell RNA-seq (scRNA-seq) data was obtained from a published study of carotid artery atherosclerotic core and proximal adjacent regions from 3 patients from the Gene Expression Omnibus (GEO) database at https://www.ncbi.nlm.nih.gov/geo/ (GEO accession: GSE159677) [22, 23]. H3K27ac and input ChIP-seq data was obtained from a published study in a single replicate of human umbilical vein endothelial cells (HUVEC) treated for 24 h with *ERG* or control siRNA (GEO accession: GSE124892) [15, 24]. Human fibroblast ATAC-seq data (five replicates) was also obtained (GEO accession: GSE121053) [25, 26].

Cell culture

All experiments were performed in teloHAECs (ATCC 50–238–1690) grown at 37 °C and 5% CO_2 in Endothelial Cell Growth Media-Microvascular 2 (ECGM-MV2) (PromoCell C-22121) on rat tail collagen type 1-coated plates (Sigma-Aldrich C3867). ECGM-MV2 media was supplemented with 5% fetal calf serum, 5 ng/mL recombinant human epidermal growth factor, 0.5 ng/mL recombinant human vascular endothelial growth factor 165, 10 ng/mL recombinant human basic fibroblast growth factor, 20 ng/mL long R3 insulin-like growth factor-1, 1 µg/mL ascorbic acid, and 0.2 µg/mL hydrocortisone (from ECGM-MV2 Supplement Pack; PromoCell C-39221). Experiments using CRISPR/Cas9-edited cells were performed on cells between passage 20 and 30. Experiments using unedited teloHAECs were performed on cells between passage 11 and 20.

CRISPR/Cas9 deletion of *ERG* exon 6

For the CRISPR/Cas9 deletion experiment, gRNAs (Additional File 1: Table S1) targeting the 5′ and 3′ boundaries of *ERG* exon 6 were designed using the MIT CRISPR design tool (http://crispr.mit.edu/). The gRNAs were purchased from Integrated DNA Technologies (IDT) as standard DNA oligomers. The gRNA oligomers were then annealed, phosphorylated, and cloned into pSpCas9(BB)–2A-GFP (PX458) under the control of the

U6 promoter (Addgene Plasmid ID 48138). TeloHAECs were transfected with 2.5 μg of each 5′ and 3′ gRNA-PX458 plasmids or two scramble control gRNA-PX458 plasmids using Lipofectamine 2000 (2μL/μg DNA, Invitrogen) in 100 mm dishes. The GFP^+ cell populations, representing cells transfected with the GFP-tagged Cas9 construct, were isolated using fluorescence-activated cell sorting with a Becton Dickinson FACSAria sorter after 48 h (SickKids-UHN Flow Cytometry Facility). Non-transfected cells were utilized to set the sorting gates. After recovery, single cells were seeded in 96-well plates and inspected for colonies after 1–2 weeks. Upon reaching 60% confluency, the colonies were split into replicate 96-well plates. Genomic DNA was isolated from one of the replicate plates by lysing cells in 96-well plates for 3 h at 65 °C in a lysis buffer (100 mM NaCl, 10 mM Tris–HCl pH 8, 25 mM EDTA pH 8, 0.5% SDS, 0.2 mg/mL) followed by 15 min inactivation at 95 °C. Genotyping was performed using Taq Polymerase (Invitrogen) with primers flanking the deletion sites or the motif mutation sites. The PCR products were run on gel electrophoresis and visualized using MiniBIS Pro (DNR Bio-Imaging Systems). Gel bands were extracted and sent for Sanger sequencing. Sanger sequencing was confirmed by aligning to the human genome (hg19) using the Blat function in the UCSC Genome Browser. WT clones were treated the same as above but did not receive gDNA. Given the potential of generating undesired large deletions using CRISPR/Cas9, multiple individual WT and homozygous *ERG* KO clones were tested in all experiments to diminish the likelihood of basing our conclusions on one clone with spurious off-target deletions.

ERG small interfering RNA inhibition

ERG KD was performed in teloHAECs. At 60–70% confluence, cells were serum starved for 1 h in Opti-MEM reduced serum medium (ThermoFisher 31985–070) and transfected in ECGM-MV2. Lipofectamine RNAiMAX (Invitrogen 56532) was used to transfect cells. A final concentration of 40 nM Ambion™ Silencer™ Select *ERG* siRNA (ThermoFisher 4392421 – s4813) or negative control siRNA (ThermoFisher 4390844) was used. Media was changed 4 h following transfection. KD was confirmed by western blot using collected cell lysates 24 h and 72 h post-transfection. For recovery experiments, media was changed every 48 h, with collection for RNA-seq, ATAC-seq, western blotting and functional experiments at specified timepoints. Biological replicates were performed on separate days.

Cell collection for downstream experiments

Cell lysates were collected as described below, using 100μL of 2 × RIPA lysis buffer (Millipore Sigma; 20–188) and 1 × Protease inhibitor (Sigma Aldrich; 11836170001). To collect cells for ATAC-seq, cells in each well were lifted with TrypLE, quenched in cell media, and counted. 50,000–70,000 cells from each well were aliquoted and spun down for 5 min at 500 g. The cell pellet was resuspended in 100μL BAM Banker and frozen in a freezing chamber at −80 °C. Cell lysates for RNA-seq were prepared as follows: cells were washed three times with room temperature PBS, 350μL of room temperature Buffer RLT was added, cells were then scraped, and the lysates transferred to a 1.5 mL tube.

EC migration

Ibidi culture-inserts (Ibidi 80209) were adhered to a #1.5 chambered coverglass (Lab-Tek 155409) plate. Cells were seeded on each side of the insert and grown to confluence (1–2 days depending on initial seeding density). For cells imaged with fluorescent confocal microscopy, ECs were stained with sterile Hoechst (20 mg/mL stock diluted 1:500 in media) and incubated for 15 min at 37 °C. Hoechst stain was removed and Spy650 fast actin labelling dye (Cytoskeleton CY-SC505) diluted 1:1000 in media was added and incubated at 37 °C for 30 min. The insert was then removed and media containing Spy650 fast actin labelling dye at 1:2000 was added. Cells were then transported for live cell imaging on the Nikon AR1 Confocal microscope or the Zeiss Axio Observer (10 × magnification was used for both). Image locations and image parameters were set while cells were incubating (approximately 30–45 min). 1–5 sets of time-lapse images were taken along each insert gap. Images were taken every 5 min for approximately 6.5 h (WT or *ERG* KO) or 4 h (control [Ctrl] or *ERG* KD). Image analysis was performed using the methodology described by Jonkman et al. [27].

xCelligence real-time cell analysis

16-well E-plates were coated with attachment factor (Gibco S-006–100) and left at 37 °C for 1 h. Baseline resistance measurements were then taken using 200μL of ECGM-MV2 for 1 h before cell seeding using xCelligence Real-Time Analyzer (RTCA, ACEA Biosciences, San Diego, CA). *ERG* KD or Ctrl teloHAECs were then seeded at a density of 3×10^4 cells/well in 200μL of medium. Cells were monitored using the xCelligence RTCA for 72 h post-ERG loss. Normalized cell index was generated by normalizing cell index values to the cell index 3 h post-seeding.

Western blot

ERG WT or KO and control or *ERG* KD cells (for CD31, ERG and CDH5) were harvested using 2 × RIPA lysis buffer (Millipore Sigma 20–188) with 1 × protease inhibitor (Sigma Aldrich 11836170001) and 1 × phosphatase inhibitor (Thermo Scientific A32957). Protein

concentration was measured using a Pierce BCA Protein Assay Kit (Thermo Scientific 23228). Samples were diluted with 2 × RIPA and 1 × protease inhibitor to ensure equal concentration of protein in each sample. Samples were prepared for loading using 4 × Laemmli loading buffer (final concentration 1x; BioRad 1610747) and β-mercaptoethanol (final concentration 2.5%) and boiled for 10 min at 95 °C. Control or *ERG* KD cells (for FN1 and VIM) were harvested using 4 × Laemmli loading buffer with 1 × protease inhibitor and 1 × phosphatase inhibitor. Approximately 3–8 μg of protein was loaded (depending on expression levels of the protein of interest) into a polyacrylamide gel (SDS-PAGE; Biorad 4561094). A BLUeye prestained protein ladder (FroggaBio PM007-0500F) was used to assess molecular weight. Loaded gels were run in running buffer at 80 V for 30 min (or until loading front had entered stacking gel) and then at 110 V for 1 h. PVDF membranes (Cytiva 10600023) were prepared for wet transfer by soaking the membrane in 100% methanol followed by transfer buffer for 10 min. The membrane and gel sandwich were run in cold transfer buffer next to an ice pack at 110 V for 90 min. PVDF membranes were then blocked in 5% skim milk diluted in 0.2% Tween-20 (TBST) for 1 h on a rocking plate. Membranes were incubated overnight at 4 °C with primary antibody diluted (1:1000 unless otherwise indicated) in 2% milk. The following primary antibodies were used: mouse anti-CDH5 (SantaCruz SC-9989), mouse anti-CD31 (Abcam 9498), rabbit anti-ERG (Abcam 133264), 1:2000 rabbit anti-GAPDH (Cell Signaling 5174), rabbit anti-FN1 (Abcam 2413), rabbit anti-Vimentin (Cell Signaling 7441S). Following primary antibody incubation, membranes were washed for 3 × 5 min and secondary antibody was added (1:5000 dilutions) in 2% milk for 1 h at room temperature. The following secondary antibodies were used: anti-rabbit HRP (Cell Signaling 7074) and anti-mouse HRP (Cell Signaling 7076). Membranes were exposed using ECL reagent (Thermo Scientific 34580) for 5 min, then transferred to a film for imaging. Membranes were imaged with a BioRad ChemiDoc Imaging system. Membranes were then stripped with stripping buffer (Thermo Scientific 46430) for 10 min and re-probed up to 3 times.

Immunocytochemistry

Cells were seeded either on rat-tail collagen-coated coverslips (Sarstedt 83.1840.002) in 24-well plates (Control or *ERG* KD) or rat-tail collagen-coated 8-well chamber slides (Thermo Scientific 177445) (WT or *ERG* KO). Cells were washed 1 × with PBS and fixed with 4% paraformaldehyde for 20 min. After fixation, cells were washed 1 × with PBS and permeabilized with 0.5% Triton-X-100 for 10 min. Cells were washed with 1% bovine serum albumin (BSA) in PBS and then blocked with 5% BSA for 1 h at room temperature. Primary antibody was then diluted in 1% BSA in PBS and cells were incubated with the dilutions at 4 °C overnight. The following primary antibody dilutions were used: 1:400 sheep anti-vWF (Abcam 133264), 1:300 rabbit anti-FN1 (Abcam 2413), 1:200 mouse anti-CDH5 (SantaCruz SC-9989), 1:300 rabbit anti-Col1A1 (Cell Signaling 72026), 1:300 mouse anti-CD31 (Abcam 9498), 1:300 rabbit anti-ERG (Abcam 133264). The following day, primary antibody was removed, cells were washed 3 × with PBS and the following secondary antibodies diluted in 1% BSA were added (1:300 dilution for all): anti-Rabbit 488 (Invitrogen A11008), anti-mouse 568 (Invitrogen A11004), anti-sheep 647 (Jackson 713–606-147). Cells were incubated with secondary antibody for 2 h at room temperature. In wells assessing F-actin, cells were incubated with a 1:600 dilution of Alexa Fluor 647 Phalloidin (Invitrogen A22287) for 30 min at room temperature. Following 3 × 5 min PBS washes, cells were stained with Hoechst (1:500 dilution of 20 mg/mL stock; Thermo Scientific 3342) for 20 min at room temperature. To mount cells, mounting media (Invitrogen P36930) and microscope slides (VWF International CA 48323–180) were used. Cell preparations were sealed with CoverGrip Coverslip Sealant (Biotium REF23005) and stored at 4 °C until imaging. Slides were imaged on a Leica SP8 Confocal microscope using a 63 × objective.

Quantitative reverse transcriptase polymerase chain reaction

RNA was extracted using the RNeasy® Micro Kit (Qiagen 74004) according to the manufacturer's instructions or Trizol (ThermoFisher 15596018). Concentration and purity of RNA was measured with nanodrop and 1 μg of RNA was taken for cDNA synthesis. RNA samples were diluted to a concentration of 1 ng/10μL. Reverse transcription kit (ThermoFisher 4368813) was used to reverse transcribe RNA into cDNA. Quantitative reverse transcriptase PCR was performed using SYBR Green (Applied Biosystems A46109) and forward and reverse primers are detailed in Additional file 1: Table S4. Data was normalized to the TATA-Box Binding protein (*TBP*) using the delta-delta Ct Method.

Statistical analysis (non-omics data)

Statistical analysis for non-omics data was performed using GraphPad Prism version 10.1.1 (GraphPad Software Inc., La Jolla, CA, USA). The number of independent biological replicates is indicated in the figure legends for all experiments. Bar and line plots were composed of 3–8 independent samples per group and were presented as mean ± standard error of the mean (SEM). Data were assessed with a Shapiro–Wilk test for normality and F test (2 groups) or Spearman's test for equality of variance

(> 2 groups). If normality and equality of variance were assumed, comparisons between groups were made using an unpaired t-test (2 groups) or a two-way ANOVA with Fisher's LSD multiple comparisons (> 2 groups). Comparisons between groups that were normally distributed but had unequal variances were made using Welch's t-tests. If data were not normally distributed, comparisons were made using non-parametric Mann–Whitney tests. For protein quantification of replicates assessed on separate days, comparisons were made using non-parametric paired samples Wilcoxon tests. For comparisons between two groups, a one-tailed unpaired t-test, Welch's t-test, or Mann–Whitney test was used in cases where a priori hypotheses for directionality of phenotype were made.

Bulk RNA sequencing

RNA extraction RNA was extracted from cell lysates in 350µL Buffer RLT using an RNeasy® Micro Kit (Qiagen 74004) according to manufacturer's instructions with the following modifications: i) after adding 350µL of 70% ethanol and centrifuging samples in RNeasy MinElute Spin Columns for 15 s at 8,000 g, the flowthrough was reapplied to the spin columns and samples were centrifuged for an additional 15 s at 8,000 g; ii) when adding 350µL of Buffer RW1, samples were incubated at room temperature for 5 min, centrifuged for 15 s at 8,000 g, and flowthrough was discarded; and iii) after adding 14µL of RNase-free water to the spin columns, samples were incubated at room temperature for 1 min and centrifuged for 1 min at 8,000 g. RNA flowthrough was reapplied to spin columns, samples were incubated at room temperature for an additional 1 min, and centrifuged for 1 min at 8,000 g to elute RNA.

Library preparation and sequencing RNA-seq libraries for CRISPR/Cas9 WT and *ERG* KO teloHAECs were prepared using CATS mRNA-seq kit v2 (with poly(A) selection) with dual indexing (Diagenode, Belgium) as per the manufacturer's protocol. To generate cDNA, 250 ng of total RNA was spiked-in with SIRV Set 3 (Lexogen GmbH, Austria), calculated based on 1% mRNA and 1% spike-in. The resulting libraries were quantified with Qubit™ DNA High Sensitivity assay (ThermoFisher, USA). Fragment sizes were analyzed on the Agilent Bioanalyzer using the High Sensitivity DNA assay prior to sequencing. Paired-end 50 bp sequencing was performed at The Centre for Applied Genomics (TCAG, Canada) on an Illumina HiSeq2500 Rapid Run (Illumina, USA) with cycles extended to 68 bp using a custom paired-end sequencing primer for read 2 (Diagenode) and the following dark cycle recipe 5DC + 58(R1) + 8(Index1) + 8(Index2) + 5DC + 58(R2). Paired end sequencing produced 59,000,000 to 87,000,000 reads per sample.

RNA extracted from *ERG* and control KD teloHAECs was sent to Novogene (Sacramento, CA) for library preparation and sequencing. Samples were assessed for quantity, integrity, and purity using an Agilent 5400 Fragment Analyzer System (Santa Clara, CA). During library preparation, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, first strand cDNA was synthesized using random hexamer primers followed by second strand cDNA synthesis. Libraries were complete after end repair, A-tailing, adapter ligation, size selection, amplification, and purification. Each library was quantified with Qubit™ DNA High Sensitivity assay (ThermoFisher, USA) and fragment sizes were analyzed on the Agilent Bioanalyzer using the High Sensitivity DNA assay prior to sequencing. Paired-end 150 bp sequencing was performed on an Illumina NovaSeq 6000 System (Illumina, USA) and produced 38,000,000–59,000,000 raw reads per sample.

Read preprocessing and differential expression analysis Raw RNA-seq reads were processed with Trimmomatic (v0.32; using recommended parameters) to remove adaptor sequences and preserve high-quality reads. The trimmed reads were then aligned to the human genome reference assembly (GRCh37/hg38) using STAR (v2.5.1b; default parameters). Quality control metrics for raw and aligned RNA-seq reads were analyzed using FastQC, MultiQC (v1.3), and Picard. One WT CRISPR/Cas9-treated teloHAEC sample was discarded due to poor quality. After validating sample quality, feature quantification for each sample was performed using Salmon (v1.09) against a transcriptome derived from hg38.

Differential expression (DE) analysis was performed using DESeq2 (v1.44.0) in R (v4.4.1). Three separate analyses were conducted on the RNA-seq datasets. The first analysis compared expression between *ERG* KO and WT CRISPR/Cas9-treated teloHAECs (2 groups). The second integrated the original CRISPR data with expression data from the *ERG* and control siRNA-treated (KD) teloHAECs at 24 h and 72 h post transfection (6 groups). Finally, the third analysis examined the *ERG* and control siRNA treated teloHAECs at 72 h of KD (time 0), and 3, 5 and 7 days later (8 groups). For all comparisons, DESeq2 objects were filtered to retain transcripts with at least 10 counts per sample in the number of samples equal to the size of the smallest group (5 samples for the KO analysis and 4 samples for the other two analyses). DE comparisons between Δ*ERG* teloHAECs and their corresponding controls were conducted using default parameters (p.adj value < 0.05 denoting significance). PCA analysis was performed on variance stabilizing-transformed data with visualization of the top 10–11 transcript loadings using the R package PCAtools (v2.16.0). Row-scaled heatmaps with Euclidean clustering of samples and 100 K-means

clusters were generated using variance stabilizing-transformed data and the R package pheatmap (v1.0.12).

Generation of cell type composite scores Composite scores for EC and EndMT transcripts were produced with a gene signature scoring method based on the Mann–Whitney U statistic using the R package Ucell (v2.8.0) [28]. EC or EndMT gene lists were generated by obtaining total elevated single cell markers for ECs or fibroblasts from the Human Protein Atlas (v23.0; proteinatlas.org) [29], respectively, and supplementing lists with manually curated genes known to be involved in EC function or EndMT. Complete EC and EndMT gene lists are available in Additional file 1: Table S2. All DE genes within each gene list were used to generate composite scores. Composite scores between groups were compared using an unpaired t-test or Welch's t-tests with Holm-Šídák multiple comparisons correction.

Cluster and pathway enrichment analyses Gene set co-regulation analysis (GESECA) was performed on Log2 and quantile normalized data (normalizeBetweenArrays (log2(dataset), method = "quantile") using the R package fgsea (v1.3.0). GESECA gene sets were assessed using the Msigdb human hallmark geneset (date of download: September 21, 2022). Additional pathway enrichment analysis was conducted using the R package clusterProfiler (v4.12.6) to predict GO biological process pathways associated with DE genes (org.Hs.eg.db database; log2 fold change > 0.5 or < −0.5). Pathways were assigned to categories based on the inclusion of curated pathway terms. Complete lists of terms for each pathway category are available in Additional file 1: Table S3. Pathway category radar plots were generated using the R package ggradar (v0.2). Disease enrichment was performed with DE genes (log2 fold change > 0.5 or < −0.5) using the R package enrichR (v3.4) and DisGeNET (v24.3) knowledge database containing 9,828 terms, 17,464 genes, and 63 terms per gene.

Integration of RNA-sequencing with genome-wide association study data GWAS single nucleotide polymorphisms (SNPs) associated with diseases of interest were obtained from the NHGRI-EBI GWAS Catalog [30] for integration with DE genes between Δ*ERG* KO and WT teloHAECs. For cases where multiple SNPs mapped to the same gene, the SNP with the minimum p-value was plotted. *ERG* KO- and WT-enriched genes associated with disease SNPs were highlighted in each plot.

Single-cell RNA sequencing analysis

Gene expression matrices generated from scRNA-seq of carotid artery atherosclerotic core and proximal adjacent regions were obtained for 3 patients (GEO accession GSE159677) [22, 23]. Demultiplexing, alignment, and gene counting were performed with Cell Ranger (v3.0.2 with default parameters; 10 × Genomics) using the GRCh38 transcriptome. Matrices were filtered, normalized, and clustered using R (v4.4.1) and Seurat (v5.1.0). Briefly, features that were detected in less than 3 cells and cells with less than 500 counts or less than 200 features or more than 4000 features were excluded. Matrices were additionally filtered to remove cells with a mitochondrial transcriptome of greater than 10% and each sample was then randomly downsampled to the minimum sample size of 2,813 cells, which resulted in a total of 16,878 cells for analysis. Normalization was performed using the LogNormalize method with a scale factor of 100,000. PCA was performed on the top 2000 variable features that were identified using the variance stabilizing-transformation selection method. Cells were clustered using the Louvain algorithm based on the first 15 PC dimensions with a resolution of 0.4. Cell cluster identities were manually assigned based on expression of classical marker genes. Cells were then subsetted to retain EC clusters and reclustered using the Louvain algorithm with the first 15 PC dimensions and a resolution of 0.1. Uniform manifold approximation and projection (UMAP) was used for non-linear reduction and two-dimensional data visualization. Differential gene expression analysis between ECs obtained from atherosclerotic core versus proximal adjacent regions was performed using the Wilcoxon Rank Sum test in Seurat (p.adj value < 0.05 denoting significance). A Pearson correlation coefficient between plaque-enriched transcripts from carotid samples and *ERG* KO-enriched transcripts from teloHAECs was obtained using the R package ggpubr (v0.6.0).

Assay for transposase-accessible chromatin with sequencing

Sample preparation and sequencing For ATAC-seq on WT and *ERG* KO CRISPR lines, approximately 50,000 live cells were lysed in non-ionic detergent lysis buffer (10 mM Tris–HCl, pH 7.4, 3 mM $MgCl_2$, 10 mM NaCl, 0.1% NP-40). Nuclei were pelleted and resuspended in 50μL transposition reaction solution containing 2.5μL Tn5 transposase attached to sequencing adapters (Illumina Nextera Tagment DNA enzyme TDE1, Nextera® DNA Sample Preparation Kit FC-121–1030), 25μL Tagment DNA buffer (20 mM Tris–HCl pH = 7.6, 10 mM $MgCl_2$, 20% Dimethyl Formamide), 16.5μL PBS, 5μL NFH_2O, 0.1% Tween, and 0.01% digitonin. Samples were resuspended in Tn5 mixture and incubated at 37 °C for 30 min under gentle mixing conditions (300 rpm) using a thermomixer (Eppendorf). The transposed and adapter-ligated DNA fragments were purified using the Zymo Research DNA Clean and Concentrator purification kit and PCR amplified using barcoded primers (Nextera Index Kit FC-121–1011) using cycle numbers determined by RT-

PCR. A SageSciences PippinPrep was used to size select DNA between 125 and 600 bp using a 2% agarose gel. The Centre for Applied Genomics (TCAG, Canada) performed paired-end 150 bp sequencing on the final ATAC libraries to a depth of ~50 million reads per sample using an Illumina HiSeq2500.

For *ERG* and control siRNA ATAC-seq, approximately 250,000 cells were harvested from each biological replicate of teloHAECs treated with *ERG* or control siRNA for 24 h or 72 h. Separate experiments were performed at 72 h of knockdown (time 0), and 3, 5 and 7 days later for recovery experiments. Cells were cryopreserved in BAMbanker freezing media (Fujifilm cellular dynamics CS-02–001) using a Mr. Frosty (Thermo Scientific™ N51000001). On the day of ATAC-seq, cells were washed with 1 mL of 37 °C ECGM-MV2 (PromoCell C-22221) and pelleted at 500 g. Free DNA from cells lysed during the cryopreservation process was removed through a 30 min incubation in ECGM-MV2 supplemented with 200 units of DNaseI (Thermo scientific EN0521) at 37 °C. Approximately 50,000 live cells were isolated from the recovered population, and ATAC-seq was performed as described above.

Read preprocessing and differential peaks analysis The quality of raw ATAC-seq data was assessed with FastQC (v0.11.8) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed of adapter sequences using Trimmomatic (v0.32; using recommended parameters). Burrows-Wheeler Aligner (BWA) (v0.7.8, default parameters) was used to align the trimmed reads to the hg38 (GRCh38) genome reference assembly. Further QC, including assessments of the library complexity, read size distribution, and read density at the TSS were performed using ATACseqQC. Peaks were called with Genrich (https://github.com/jsh58/Genrich) using its multiple replicates, deduplication and ATAC-seq modes.

In preparation for differential peaks analysis, a counts matrix was constructed in R (Version 4.4.1) using the featureCounts function from the R package Rsubread. A list of BAM files was provided as input, and genomic features were defined by a union peaks file generated using the bedops –merge tool. Deseq2 was used for normalization and to call differential peaks. ChIP-seeker was used to annotate differential peaks to genes defined by the UCSC hg38 TxDb. Genomic Regions of Annotations Tool (GREAT) [31] was used to call GO term enrichments of differential peaks using the union peaks file as a background. HOMER was used to call motif enrichments of differential peaks against hg38, using the size given and –nomotif settings.

For footprinting analysis using TOBIAS (0.14.0) [32], JASPAR's non-redundant motif set was downloaded (date of download: 26 November, 2021), subsetted to include only transcription factors that were expressed at an average of > 10 transcripts per sample in our RNA-seq, and formatted using TOBIAS formatMotifs. ATAC-seq replicates were merged using Samtools (v 1.21), and reads were corrected for Tn5 bias using TOBIAS ATACorrect. Footprinting scores were generated using TOBIAS FootprintScores, and finally footprinting was compared across conditions with TOBIAS BINDetect using the reduced transcription factor set.

Chromatin immunoprecipitation sequencing

Read preprocessing and differential peaks analysis H3K27ac and input ChIP-seq FASTQs from HUVECs treated for 24 h with *ERG* or control siRNA were downloaded from GSE124892 [15, 24]. Preprocessing was performed as described in the ATAC-seq section above. Peaks were called using HOMER (http://homer.ucsd.edu/homer/), TAG files were created for each condition using HOMER makeTagDirectory, and peaks were identified using HOMER findPeaks in histone mode against input DNA. Peaksets from both conditions were merged using HOMER mergePeaks to create a union peakfile. HOMER getDifferentialPeaks was run to identify differential peaks between conditions.

Integrated sequencing analysis to identify high evidence genes (HEGs)

To create a HEG list, ChIP-seeker was used to annotate ATAC-seq and ChIP-seq peaks to their nearest genes. Gene fold changes were calculated by summing the fold changes of all associated peaks. The minimum *p.adj* value (matching the fold change direction) was used to represent peak significance. Genes with consistent fold change direction and $p.adj < 0.05$ across RNA-seq, ATAC-seq, and ChIP-seq datasets were identified as HEGs. To determine which transcription factor (TF) families may be regulating expression of HEGs, TOBIAS differential footprinting log2FCs for each transcription factor were annotated to nearest gene using ChIPseeker [33] and were subsetted for regions that mapped to upregulated or downregulated HEGs. If no footprints from a transcription factor mapped to a HEG it was assigned a fold change of 0. TOBIAS footprinting log2FC values for each transcription factor were correlated with the ATAC-seq log2FC values for each HEG and were rank ordered by R value. To identify transcription factors whose expression changes may be driving global changes in transcription factor family binding, RNA-seq expression for all transcription factors across all conditions (72 h control siRNA, 72 h *ERG* KD, 24 h control siRNA, 24 h *ERG* KD), was correlated with the global footprinting score for that transcription factor across those conditions. Transcription factors were then grouped by family (AP1, ETS, etc.) and plotted by R value.

Results

Chronic loss of ERG drives robust changes in the endothelial gene regulatory landscape and promotes expression of a network of genes related to mesenchymal identity

To investigate the role of ERG in maintaining identity and function of the endothelium, we generated three independent *ERG* knockout (KO; Δ*ERG*) clones and three matched wild-type (WT) controls in telomerase-immortalized human aortic endothelial cells (teloHAECs) using CRISPR/Cas9-mediated deletion of *ERG* exon 6 (Fig. 1A). Loss of ERG in each clone was confirmed by western blot, revealing a 94% reduction in protein expression in KO cells (Fig. 1B). Δ*ERG* teloHAECs were characterized by a prominent disruption of EC morphology, with loss of their classic cobblestone monolayer morphology and development of a mesenchymal-like cell phenotype (Fig. 1C). To identify the molecular mechanisms mediating these morphologic alterations, we compared the transcriptional landscape of Δ*ERG* and WT teloHAECs using RNA sequencing (RNA-seq). Dimensionality reduction and visualization of these data using principal component analysis (PCA) revealed that Δ*ERG* and WT samples clustered distinctly from one another, with over 70% of the variance in gene expression attributable to genotype (Fig. 1D). Notably, the top PCA loadings driving clustering of Δ*ERG* cells included mesenchymal cell transition genes such as fibulin-2 (*FBLN2*), fibulin-5 (*FBLN5*), and versican (*VCAN*), while those that supported WT cell clustering included pro-angiogenic genes such as apelin (*APLN*) and cytokine-like 1 (*CYTL1*). Remarkably, differential gene expression analysis identified 3,090 differentially expressed (DE) genes enriched in WT cells and 3,409 DE genes enriched in Δ*ERG* cells that accounted for 20.2% and 22.3% of the total transcriptome, respectively (Fig. 1E). Thus, > 40% of the transcriptome in aortic ECs is impacted by long-term loss of ERG. To profile the global effect of ERG loss on EC identity, we leveraged a composite gene signature scoring package [28] to assess expression of EC- and EndMT-enriched transcripts in Δ*ERG* and WT teloHAECs (Fig. 1F). We observed downregulation of 162 of 212 EC-enriched genes in Δ*ERG* cells, including transcripts associated with cell adhesion (*CDH5*; claudin 5, *CLDN5*), angiogenesis (*APLN*; Fms-related tyrosine kinase 4, *FLT4/VEGFR3*; kinase insert domain receptor, *KDR/VEGFR2;* tyrosine kinase with immunoglobulin-like and EGF-like domains 1, *TIE1*), and EC identity (BMX non-receptor tyrosine kinase, *BMX*; von Willebrand factor, *VWF*). In contrast, Δ*ERG* teloHAECs had an upregulation of EndMT-enriched genes (115 of 197 transcripts), including transforming growth factor beta 1 (*TGFB1*), ZEB and TWIST transcription factors, and the extracellular matrix proteins collagen type VIII alpha 1 (*COL8A1*), *FBLN2/5*, and fibronectin (*FN1*).

To examine the chromatin accessibility changes that contribute to the marked alteration of the EC transcriptome, we performed ATAC-seq on Δ*ERG* and WT teloHAECs. We detected ~ 90,000 unique peaks across samples with discrete clustering by genotype (88% variance attribution in PCA) (Fig. 1G and Additional File 1: Fig. S1A). Nearly 20% of these peaks were identified as differentially accessible regions (DARs), where 8,828 DARs were enriched in WT cells and 8,684 DARs were enriched in Δ*ERG* teloHAECs (Fig. 1G and Additional File 1: Fig. S1B). Most DARs occurred in distal intergenic and intronic regions of the genome (Fig. 1H), which is consistent with the established role of ERG in regulating endothelial enhancers [15, 34]. Motif enrichment analysis revealed that loss of ERG resulted in increased accessibility of motifs associated with the specificity protein/Krüppel-like factor (SP/KLF) (i.e., KLF2 and KLF4) and activator protein-1 (AP1) families (Fig. 1I and Additional File 1: Fig. S1C,D), the latter of which shapes the EC enhancer landscape and mediates EC activation [34, 35]. Despite potential redundancy, as ERG is one of at least 19 ETS factors expressed in ECs [36], loss of ERG alone resulted in a profound decrease in ETS motif accessibility. No changes were noted in GATA or SMAD motif accessibility. Pathway analysis of WT-enriched DARs highlighted core EC functions including proliferation and neovascularization (Fig. 1J), which were derived from accessibility changes in EC genes such as *CYTL1* (Fig. 1K). In contrast, Δ*ERG*-enriched DARs were associated with EndMT-related pathways including ECM organization, abnormal cell adhesion, and SMAD binding (Fig. 1J), which resulted from altered accessibly of ECM genes including *FBLN5* (Fig. 1K). Together, these data highlight a prominent role for ERG in governing transcriptional and chromatin accessibility programs that enforce EC identity and suppress EndMT in the arterial endothelium.

ERG knockout induces endothelial-to-mesenchymal transition, disrupts endothelial cell–cell junctions, and drives increased cell migration

Next, we sought to confirm if these increases in EndMT-related transcriptional networks were accompanied by changes at the protein and phenotypic levels. Assessment of EC markers revealed a robust decrease in CD31, CDH5, and vWF, as determined by quantifying normalized fluorescence intensity and western blotting (Fig. 2A, B and Additional file 1: Fig. S2A-D). An increased expression (COL1A1) and altered distribution (FN1) of mesenchymal cell markers was also observed in Δ*ERG* cells (Fig. 2A, C). Reduced localization of junctional proteins (CD31, CDH5) in confluent cells suggested a loss of barrier integrity in Δ*ERG* cells (Additional file 1: Fig. S2A).

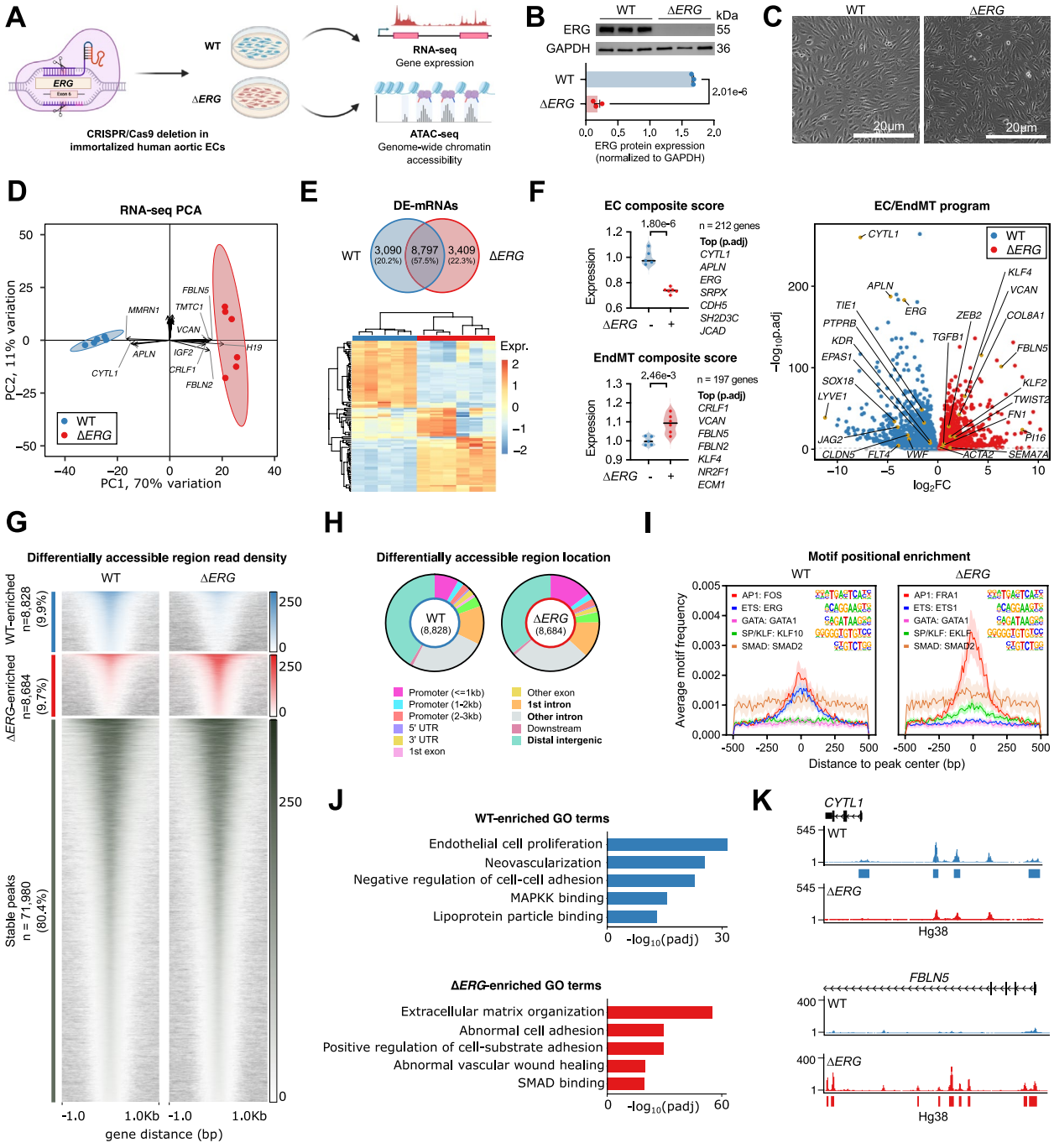


Fig. 1 *ERG* knockout alters the transcriptional landscape of human aortic endothelial cells and drives gene expression programs for mesenchymal transition. **A** Schematic for CRISPR/Cas9-mediated deletion of *ERG* exon 6 and profiling of Δ*ERG* and wild-type (WT) telomerase-immortalized human aortic endothelial cells (teloHAECs) with RNA sequencing (RNA-seq) and assay for transposase-accessible chromatin with sequencing (ATAC-seq). Quantification of ERG protein expression **(B)** and brightfield imaging **(C)** for Δ*ERG* and WT teloHAEC clones ($n = 3$; unpaired t-test). **D** Principal component analysis (PCA) of RNA-seq in Δ*ERG* and WT teloHAECs with top transcript loadings indicated ($n = 5$–6). **E** Hierarchical clustering analysis of Δ*ERG* and WT transcriptomes with shared and differentially expressed (DE) mRNAs indicated (p.adj < 0.05). **F** Composite scores for DE endothelial cell (EC) and endothelial-to-mesenchymal transition (EndMT) transcripts in Δ*ERG* and WT teloHAECs with top transcripts by p-value highlighted (left) ($n = 5$–6; unpaired t-tests). Scores are normalized relative to WT teloHAECs. DE analysis of EC and EndMT programs shown in volcano plot (right) (p.adj < 0.05). **G** Read density heatmap of accessible chromatin regions from ATAC-seq in Δ*ERG* and WT teloHAECs ($n = 2$; p.adj < 0.05). **H** Ring plots depicting chromosome locations for differentially accessible regions (DARs) with locations of interest shown in bold text. **I** Positional enrichment plots for motif families of interest in Δ*ERG* and WT teloHAECs. Top motifs by p-value are depicted with consensus sequences. **J**, **K** Gene Ontology (GO) pathway enrichment of DARs with select genes associated with reduced accessibility (*CYTL1*) and increased accessibility (*FBLN5*) in Δ*ERG* teloHAECs highlighted (p.adj < 0.05). Bar graphs are presented as mean ± standard error of the mean (SEM). Violin plots indicate density of data points at corresponding values

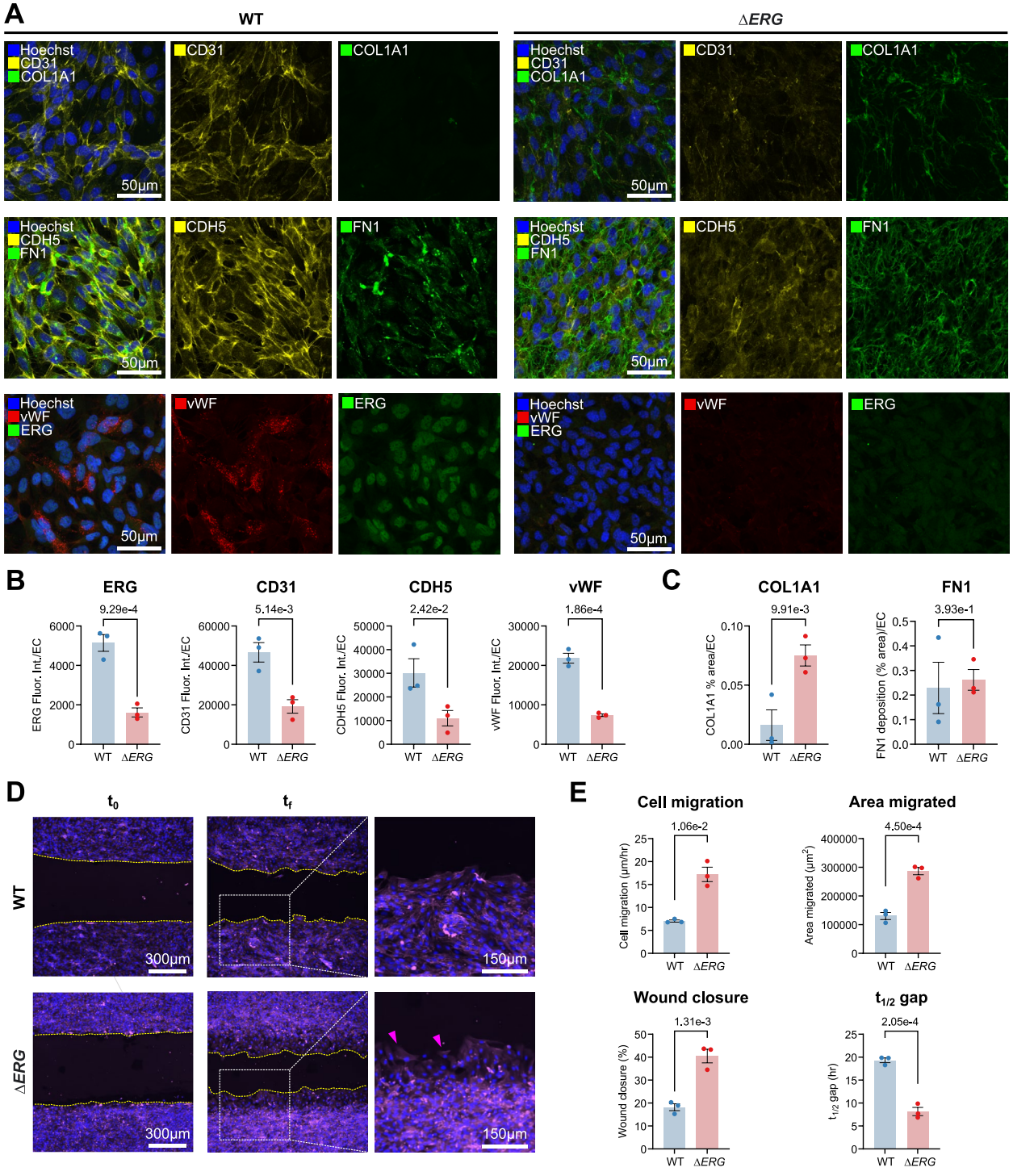


Fig. 2 (See legend on next page.)

Cytoskeletal remodeling was also evident in Δ*ERG* cells, with an altered F-actin network observed (Additional file 1: Fig. S2B). Given the differences in ECM organization, adhesion, and abnormal wound response pathways identified by Gene Ontology (GO) analysis of the transcriptional profiling data (Fig. 1J), we assessed the migratory capacity of Δ*ERG* cells. Δ*ERG* cells migrated more rapidly than WT teloHAECs in a scratch wound assay, and while WT cells largely maintained their cell–cell interactions and migrated collectively, Δ*ERG* cells displayed single cell migration (Fig. 2D, E inset; Additional file 2: Video S1 and Additional file 3: Video S2).

(See figure on previous page.)

Fig. 2 Chronic loss of *ERG* promotes endothelial-to-mesenchymal transition, disrupts cell–cell junctions, and increases migration of aortic endothelial cells. **A** Representative confocal microscopy images of *ERG* knockout (ΔERG) and wild-type (WT) telomerase-immortalized human aortic endothelial cells (teloHAECs). Images depict EC marker staining (ETS related gene, ERG; cluster of differentiation 31, CD31; vascular endothelial-cadherin, CDH5; von Willebrand factor, vWF) and mesenchymal cell marker staining (collagen type 1 alpha 1, COL1A1; fibronectin, FN1). Quantification of EC markers (**B**) and mesenchymal markers (**C**) ($n = 3$; unpaired t-tests). ERG – quantification of nuclear fluorescence intensity normalized to cell number. CD31, CDH5, vWF – quantification of image fluorescence intensity normalized to cell number. COL1A1, FN1 – quantification of the percentage of image area with signal present normalized to cell number. Each datapoint is the average of at least 3 fields of view in each biological replicate. **D** Representative images of live-cell imaging experiment depicting migration of Δ*ERG* and WT teloHAECs. Image post-barrier removal (left); image post-migration assay (right). Dashed yellow lines denote migration front and magenta arrows denote single migrating cells. **E** Quantification of cell migration rate ($n = 3$; Welch's t-test) and area migrated, wound closure, and the ½ gap closure time ($t_{1/2}$ gap) ($n = 3$; unpaired t-tests). Values determined from plot of gap area over time. Each datapoint is the average of 1–5 fields of view along the length of the wound in each biological replicate. Bar graphs are presented as mean ± standard error of the mean (SEM)

Transient ERG loss reveals dynamic and sequential alterations in endothelial and mesenchymal cell identity transcriptional pathways

To determine the temporal relationship and kinetics of EC identity loss and EndMT acquisition, we performed siRNA-mediated KD of *ERG* in teloHAECs. Timepoints for *ERG* KD were selected based upon ERG protein expression (Fig. 3A) and EC morphology data (Fig. 3B). We designated 24 h post-transfection as "early silencing" based on a 75% reduction in ERG protein without discernable morphological changes in teloHAECs. Conversely, we denoted 72 h post-transfection as "late silencing" since cells maintained a comparable reduction in ERG expression but adopted a spindle-like cell morphology indicative of a mesenchymal-like phenotype (Fig. 3B).

We next used RNA-seq and ATAC-seq to identify the gene regulatory programs underlying the transition from an EC identity to a mesenchymal cell fate following ERG loss. PCA of a composite dataset including siRNA-treated teloHAECs (24 and 72 h) and *ERG* KO teloHAECs revealed a progressive transition in cell transcriptome from early to chronic stages of EndMT (red data points along PC2 axis), which was most evident when resolving samples using the second and third principal components (Additional file 1: Fig. S3A). This was also highlighted by hierarchical clustering of RNA-seq data (Additional file 1: Fig. S3B). Differential gene expression analysis similarly identified an increasing divergence in transcriptomes across timepoints, where 21% of the transcriptome (3,194 genes) was differentially regulated after 24-h *ERG* KD, 37% of the transcriptome (5,655 genes) was differentially regulated after 72-h *ERG* KD and 43% of the transcriptome (6,499 genes) was differentially regulated with *ERG* KO (Fig. 3C). Aggregate gene signature scoring of EC-enriched transcripts revealed a progressive loss of endothelial identity that was already evident 24 h after *ERG* KD (92 transcripts; p.adj = 3.61e-4) and was driven by downregulation of EC genes including *APLN*, *CLDN5*, and *CYTL1* (Fig. 3D, E). Notably, the EC score was progressively reduced over time, as seen from 24-h *ERG* KD to *ERG* KO (Fig. 3D; left panel). Consistent with our observation that mesenchymal-like morphology is not yet evident at 24 h of *ERG* KD, there was no significant increase in composite EndMT score during early *ERG* silencing, although several individual EndMT transcripts were modestly upregulated in *ERG* KD teloHAECs at this timepoint (e.g., *TGFB1*, *SNAI1*, *COL8A1*). However, by 72 h, EndMT was significantly induced with upregulation of 91 associated transcripts, and this was further enhanced in chronic *ERG* KO ECs with an increase in 115 EndMT-related transcripts (Fig. 3D; right panel). Overall, these transcriptomic data support a sequential and progressive model of EndMT where the loss of endothelial identity precedes the acquisition of an established mesenchymal cell state in aortic ECs (Fig. 3D, E).

ERG KD teloHAECs were subsequently profiled using ATAC-seq to identify early chromatin accessibility changes that accompany the loss of EC identity and induction of EndMT. In line with our transcriptomic results, we observed a marked and progressive increase in differential chromatin accessibility between early and late *ERG* silencing, where *ERG* KD and control ECs each contained ~ 3,000 DARs at 24 h and > 13,000 DARs at 72 h (Fig. 3F). Pathway analysis of both early and late EndMT accessibility profiles highlighted vascular developmental and morphogenic processes in control cells and mesenchymal cell-associated and migratory pathways in *ERG* KD cells (Fig. 3G). Additionally, gene set co-regulation analysis (GESECA) revealed a steady increase in chromatin accessibility near genes associated with mesenchymal transition over time (Fig. 3H, denoted by the 'Epithelial to mesenchymal transition' pathway term). In contrast to RNA-seq data, ATAC-seq results showed that mesenchymal, extracellular matrix, and migration pathways were significantly enriched as early as 24 h of *ERG* KD (Fig. 3G, H), which suggested a rapid and progressive opening of chromatin associated with these sites. To characterize the transcription factor (TF) activity underlying this mesenchymal cell signal, we performed a motif enrichment analysis on a composite ATAC-seq dataset across *ΔERG* timepoints (i.e., early, late, and chronic) (Fig. 3I). We observed a rapid and increasing enrichment of ETS family motifs (ERG, ETS1, ETV2, FLI1) across early to

chronic stages of EndMT in downregulated ATAC-seq peaks (Fig. 3I; top panel). Interestingly, downregulated peaks in *ERG* KO cells were also enriched in AP1 motifs, perhaps driven by a long-term failure of AP1 to compensate for a loss of ERG at this subset of regulatory elements [34]. Notably, we identified a strong association of upregulated peaks with a subset of AP1 family motifs including AP1, JUN, and FOS that demonstrated a stable enrichment from early to chronic loss of ERG and expanded to include additional related bZIP family members, including BACH1 and MAF factors in the late setting (Fig. 3I; bottom panel). Similar changes in ETS and AP1 motif dynamics were observed for the subset of 'EC composite' (n = 270) and 'EndMT composite' (n = 242) genes, with an increasing enrichment in AP1 and other bZIP family motifs from early to late stages for EndMT composite genes (Additional file 1: Fig. S3C). We next mapped the altered ETS or AP1 motifs to lost or gained peaks, respectively. Peaks which were significantly smaller upon *ERG* KD and contain an ERG motif were increasingly associated with genes involved in endothelial GO terms (e.g., angiogenesis, blood vessel development) over time, while peaks that were significantly larger in *ERG* KD and contain a JUN motif were increasingly associated with genes involved in cellular locomotion and altered cell adhesion (Additional file 1: Fig. S3D).

To delineate genes of interest in early cell identity changes, we integrated a published H3K27Ac chromatin immunoprecipitation (ChIP-seq) dataset (which marks active enhancer elements) in HUVECs treated with control or *ERG* siRNA for 24 h (GSE124892) [15, 24] with our 24 h ATAC-seq and RNA-seq data to identify high evidence genes (HEGs) with commonly enriched accessibility, enhancer activity, and expression (Fig. 3J). H3K27ac ChIP-seq changes were significantly associated ($p < 10^{-16}$) with corresponding changes in our ATAC-seq and RNA-seq data (Additional file 1: Fig. S4A,B). Moreover, we identified 4,613 unique genes that were altered by loss of ERG in at least one dataset, with 335 genes that were significantly downregulated across all three 24-h datasets (downregulated HEGs) and 174 genes significantly upregulated (upregulated HEGs) (Fig. 3J and Additional file 1: Fig. S4B,C). Both upregulated and downregulated HEGs were associated with cytoskeletal processing (Additional file 1: Fig. S4B,C), with many of the downregulated HEGs consisting of RhoGEFs such as pleckstrin homology and RhoGEF domain containing G1 (*PLEKHG1*; previously associated with mechanosensitive EC cytoskeletal remodeling) [37], and upregulated genes consisting of cytoskeleton-associated proteins such as FN1, collagen type V alpha 2 (*COL5A2*) and prolyl 3-hydroxylase 2 (*P3H2*). We also observed a shift away from vascular development towards cell adhesion and TGF-β pathways.

To better understand which TFs may be driving the regulation of HEGs, we correlated ATAC-seq footprinting scores for each TF with accessibility changes at sites near each HEG. Ordering these correlations by rank revealed that TOBIAS differential footprinting scores for the AP1 family strongly correlated with ATAC-seq fold changes in the upregulated HEG set, while ETS factor differential footprinting scores strongly correlated with ATAC-seq changes in the downregulated HEGs (Fig. 3K). As ETS and AP1 are large protein families whose members share a similar motif sequence, we dissected which TFs may be driving the observed changes in footprinting by correlating the global footprinting score with RNA-seq expression for individual TF family members across all siRNA-treated and CRISPR teloHAEC datasets (Additional file 1: Fig. S4D). As expected, this analysis revealed that ERG expression was strongly correlated with differential ETS binding scores (R = 0.96), while JUN, FOS, and FOSL1 expression correlated most strongly with differential AP1 binding (R > 0.8). Together, these data suggest an early transition from ERG to a JUN/FOS/FOSL1 transcriptional network that may drive loss of endothelial identity and gain of mesenchymal and migratory phenotypes in aortic ECs during the early phases of EndMT.

Endothelial dysfunction induced by *ERG* silencing mirrors the kinetics of transcriptional programs for mesenchymal transition

Pathway enrichment from RNA-seq of teloHAECs upon loss of ERG revealed early and progressive shifts in adhesion and migration pathway categories (Fig. 4A). Notably, transcript contributions (gene ratios) for both cell adhesion and migration pathways increased from 24 to 72 h of *ERG* silencing to match that of *ERG* KO ECs. To investigate the protein-level changes that underlie this rapid transition to a mesenchymal phenotype, we quantified expression of the classical EC markers CD31 and CDH5 and the mesenchymal markers FN1 and vimentin (VIM). While EC identity-associated chromatin accessibility and transcriptional programs were disrupted after only 24 h of *ERG* KD (Fig. 3), this did not manifest at the protein level until late *ERG* silencing (72 h post-knockdown) (Fig. 4B,C). Similarly, increased FN1 and VIM expression was not observed until 72 h of *ERG* KD (Fig. 4D, E), which was consistent with the delayed acquisition of mesenchymal gene expression networks with loss of ERG in teloHAECs (Fig. 3).

We next determined whether alterations in core EC functions including establishment of barrier function and cell migration followed a similar trajectory to gene and protein changes following ERG loss. Immunostaining revealed that *ERG* KD reduced expression of junctional proteins (CDH5) and disrupted cytoskeletal organization at 24 h, which was more pronounced at 72 h of silencing

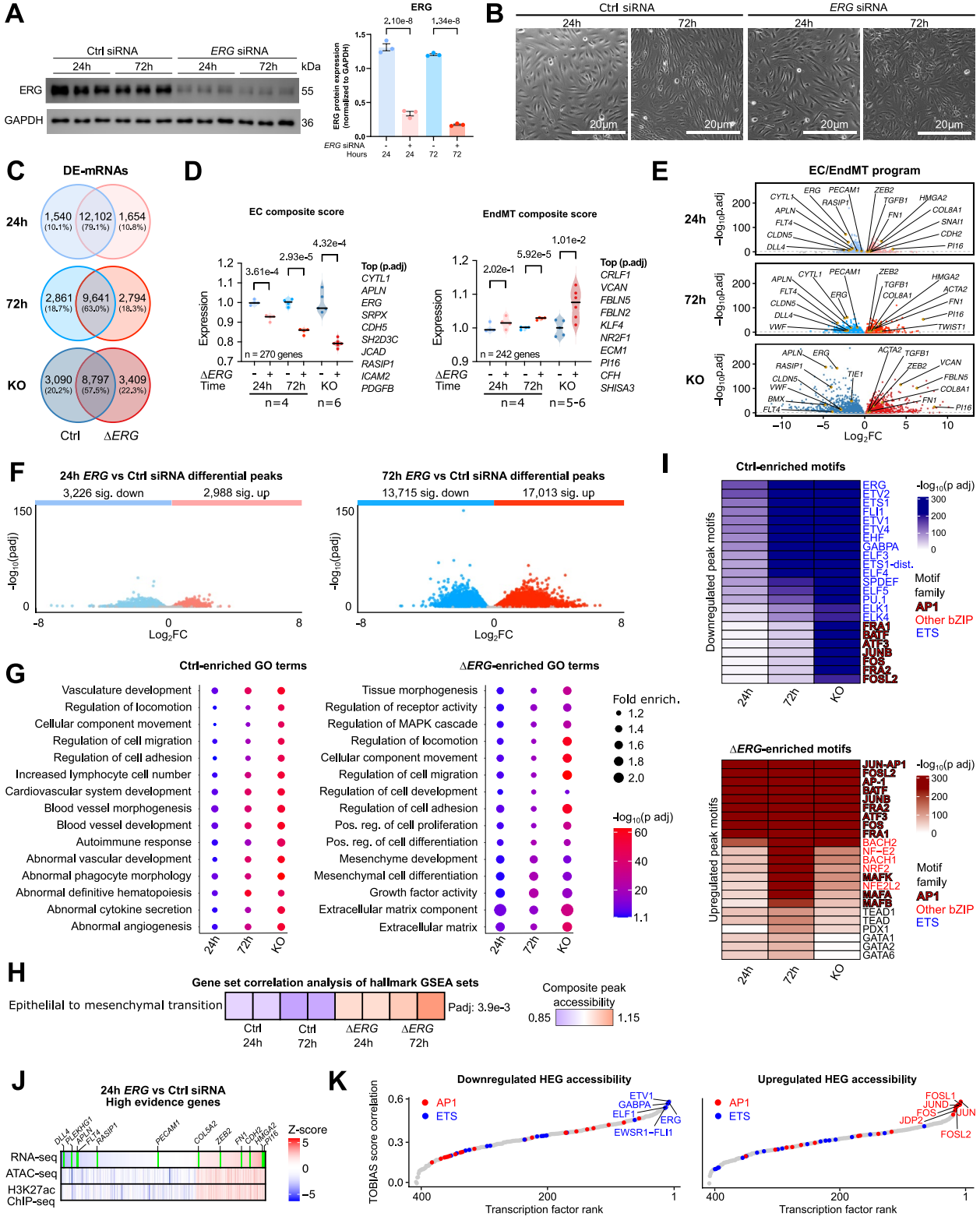


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Early induction of endothelial-to-mesenchymal transition is accompanied by activation of the AP1 transcription factor family. **A** Western blot and densitometry quantification of ERG protein expression in teloHAECs treated with control (Ctrl) or *ERG* small interfering RNA (siRNA) for 24 or 72 h ($n = 3$; two-way ANOVA). **B** Representative brightfield images of *ERG* knockdown (KD) versus control teloHAECs at the indicated times. **C** Differentially expressed (DE) mRNA plots displaying the total number of Δ*ERG* and control (Ctrl)-enriched transcripts at 24-h and 72-h *ERG* KD (top and middle) and *ERG* knockout (KO; bottom) (p.adj < 0.05). **D** Composite scores for DE EC and endothelial-to-mesenchymal transition (EndMT) transcripts in *ERG* KD and *ERG* KO teloHAECs with top transcripts by p-value highlighted ($n = 4$–6; Welch's t-tests). Scores are normalized relative to control teloHAECs in each group. **E** Volcano plots of DE genes in teloHAECs (p.adj < 0.05). **F** Volcano plots of assay for transposase-accessible chromatin with sequencing (ATAC-seq) derived differentially accessible regions (DARs) for teloHAECs exposed to control or *ERG* siRNA for 24 (left) or 72 h (right) (p.adj < 0.05; $n = 3$). **(G)** Genomic regions enrichment of annotations tool (GREAT) analysis of DARs enriched in control (left) or Δ*ERG* samples (right). **H** Gene set correlation analysis of ATAC-seq time course data against the 'Epithelial to mesenchymal transition' pathway in the hallmark gene set enrichment analysis (GSEA) set. **I** HOMER motif enrichment analysis of DARs associated with control (top) or Δ*ERG* teloHAECs (bottom). **J** Z-score normalized heat maps of ATAC-seq, H3K27ac chromatin immunoprecipitation (ChIP-seq), and RNA sequencing (RNA-seq) data for 509 high evidence genes (HEGs), which are differentially and uniformly regulated at 24-h *ERG* KD across all datasets (p.adj < 0.05). Select EC and EndMT genes are depicted. **K** Transcription factors rank-ordered by the correlation (R-value) of TOBIAS footprinting fold change with ATAC-seq fold change for regions that annotated to HEGs with ChIP-seeker. The rank 1 transcription factor has the highest correlation between binding (footprinting) and accessibility (ATAC-seq) at these regions. Bar graphs are presented as mean ± standard error of the mean (SEM). Violin plots indicate density of data points at corresponding values. FC, fold change

(Fig. 4F, G). Decreased CD31 junctional localization was only evident in late *ERG* silencing (Fig. 4F). To further evaluate the functional consequences of ERG loss over time, EC barrier function was assessed using a real-time cellular impedance assay (Fig. 4H). Cell index was used as a surrogate measure for the electrical resistance of the monolayer and was measured up to 72 h post-*ERG* KD. These experiments revealed that barrier integrity was maintained until 24 h after siRNA treatment, when a sharp decline in integrity was observed in *ERG* KD teloHAECs. Alterations in migratory phenotypes were also assessed in a scratch wound assay. Although no changes in migration were observed 24 h after *ERG* KD, migration was significantly increased 72 h after *ERG* KD (Fig. 4I, J). Collectively, these findings indicate that ERG loss triggers early structural changes in ECs, but functional alterations in barrier integrity and acquisition of a migratory, EndMT-like phenotype require more time.

Assessing ERG loss and restoration as a window to cardiovascular disease progression and regression

To determine how ERG loss in aortic ECs induces transcriptional programs associated with pathology, we performed a pathway enrichment analysis using the DisGeNET database, which contains associations between 17,464 genes and 9,828 disease pathways [38]. We predicted disease associations using RNA-seq data obtained from *ERG* siRNA and CRISPR/Cas9 *ERG* deletion teloHAECs and identified a prominent enrichment of cancer pathways (neoplasm metastasis, tumor progression), which supported our observation of dedifferentiation, mesenchymal transformation and increased migratory capacity in ERG-deficient cells (Fig. 5A). We also noted a robust and progressive enrichment of cardiovascular disease pathways (atherosclerosis, myocardial infarction, aortic aneurysm) from early *ERG* KD to *ERG* KO, which highlighted the role of ERG in protecting against arterial endothelial dysfunction. Associations were also noted with fibrotic disease (liver cirrhosis). To further examine the relevance of ERG loss in human disease and identify candidate causative genes, we integrated our RNA-seq data with the NHGRI-EBI GWAS Catalog [30] and observed genome-wide-significant single-nucleotide polymorphisms within EC genes (*APLN*, *CYTL1*) and EndMT genes (*FBLN2*, *VCAN*) that were linked to GWAS for cancer, cardiovascular disease, coronary artery disease, and aortic aneurysm (Fig. 5B-E). We next assessed associations with atherosclerosis by analyzing a public scRNA-seq dataset from human carotid endarterectomy samples (Additional file 1: Fig. S5A,B) [22], where we identified notable clustering of ECs according to atherosclerotic plaque versus non-plaque adjacent regions (Fig. 5F). We then compared the transcriptional profiles of plaque-enriched ECs with *ERG* KO teloHAECs (Fig. 5G) and found that while the overall correlation between the two datasets was not significant, there was an overlap of EndMT genes that were upregulated in both atherosclerosis and *ERG*-deficient cells, including *COL8A1*, *FBLN5*, *FN1*, and S100 calcium binding protein A4 (*S100A4*) (Fig. 5H; right panel). Expression of EC transcripts including ADAM metallopeptidase with thrombospondin type 1 motif 9 (*ADAMTS9*), *CYTL1*, Frizzled-4 (*FZD4*), and *PLEKHG1*, which have been implicated in angiogenesis and alignment of ECs in the endothelium [37], were reduced in both datasets (Fig. 5H; left panels). Together, these findings highlight the potential role of ERG in protecting against transcriptional programs that induce endothelial dysfunction and are associated with diseases of the arterial vasculature. Future work will be needed to directly link ERG activity to the control of these disease-associated genes.

The extent to which EndMT is reversible remains a key unanswered question of relevance to the design of therapeutic strategies that enforce EC identity, maintain EC resilience, and mitigate or reverse cardiovascular disease progression. Long-term *ERG* KD (i.e., 72 h) in siRNA-treated teloHAECs closely recapitulated the robust alterations in chromatin accessibility, mesenchymal

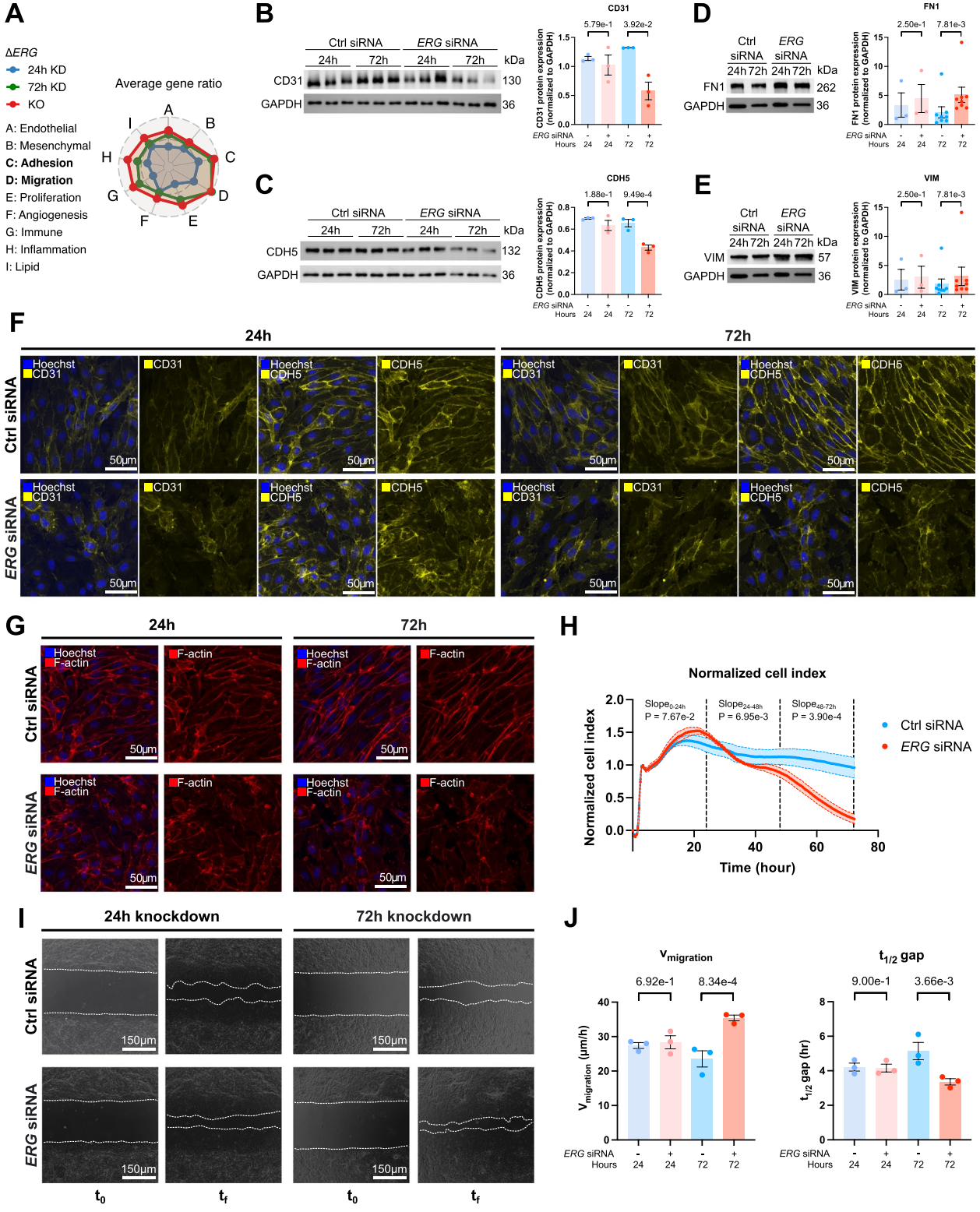


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Disruption of endothelial function mirrors the temporality of endothelial-to-mesenchymal transition following ERG loss. **A** Radar plot depicting enrichment of Gene Ontology Biological Process pathway categories from transient *ERG* knockdown (KD) to *ERG* knockout (KO) in telomerase-immortalized human aortic endothelial cells (teloHAECs) (p.adj < 0.05). Western blot and densitometry quantification of the EC proteins CD31 (**B**; $n = 3$; Welch's t-tests) and CDH5 (**C**; $n = 3$; two-way ANOVA) and the mesenchymal proteins FN1 and VIM (**D**, **E**; $n = 3$–8; Wilcoxon tests) in teloHAECs treated with control (Ctrl) or *ERG* small interfering RNA (siRNA) for 24 or 72 h. **F** Representative confocal microscopy images of *ERG* KD versus control teloHAECs depicting the junctional markers CD31 and CDH5 (representative of 3 independent experiments). **G** Representative confocal microscopy images of *ERG* KD versus control teloHAECs with F-actin visualized (representative of 3 independent experiments). **H** Time-course monitoring of barrier function (cell index) post-*ERG* KD ($n = 3$; two-way ANOVA). Values are normalized to 3 h post-cell seeding. p-values associated with differences in average slope between *ERG* KD and control teloHAECs are indicated for 0–24 h, 24–48 h, and 48–72 h post-cell seeding. **I** Representative brightfield images depicting live-cell migration of *ERG* KD versus control teloHAECs. Images post-barrier removal (t_0) and post-migration assay are shown (t_f). **J** Quantification of cell migration rate ($v_{migration}$) and ½ gap maximal area ($t_{1/2}$; $n = 3$; two-way ANOVA). Values determined from plot of gap area over time. Each dot is the average of 1–5 fields of view along the length of the wound in each biological replicate. Bar and line graphs are presented as mean ± standard error of the mean (SEM)

transition, and migratory capacity observed in *ERG* KO, which prompted us to investigate whether these changes in cell state persisted when ERG expression was restored. To this end, we treated teloHAECs with *ERG* or control siRNA and assessed the resolution of EndMT beginning at 72 h post-transfection (day 0), during which ERG expression was expected to recover due to dilution of siRNA from cell proliferation (Fig. 6A). When compared with control cells, *ERG* siRNA-treated teloHAECs possessed a spindle-shaped fibroblast-like morphology at 72 h of KD (day 0) and progressively transitioned towards a cobblestone, endothelial-like arrangement during siRNA wash-out (Fig. 6B). ERG protein expression gradually recovered with levels returning to normal by Day 5 (Fig. 6B). We next assessed global gene regulatory networks by performing RNA-seq and ATAC-seq at Day 0 (72 h of KD) and then at Days 3, 5 and 7 of recovery (Fig. 6A). At the transcript level, there were 4,243 DE genes at Day 0, and this rapidly decreased to 1956, 634 and 67 differential genes at Days 3, 5 and 7 of recovery, respectively (Fig. 6C). The EC identity and EndMT composite scores returned to normal by 5 days of recovery (Fig. 6D), with the recovery of constituent EC identity genes (*APLN*, *DLL4*, *ERG*, *VWF*) and EndMT genes (*ACTA2*, *FN1)* validated via qPCR (Fig. 6E, Additional file 1: Fig. S6A). A similar resetting of the regulatory landscape was observed in ATAC-seq data with 14,757 DARs at Day 0, followed by a progressive decrease to 5858, 476 and 0 DARs at Days 3, 5 and 7, respectively (Fig. 6F). The EC identity gene, *PLEKHG1,* exemplifies the re-establishment of open chromatin at an ERG-bound regulatory element during recovery (Fig. 6G). Analysis of motif usage during recovery revealed that loss of ETS motif usage was primarily associated with downregulated peaks at Day 0, and accessibility of these motifs was progressively regained during recovery (Fig. 6H). In contrast, a gain of AP1 motif usage was associated with upregulated peaks at Day 0, and these became less accessible across the recovery period (Fig. 6H).

We were surprised that the regulatory landscape appeared to fully reset following ERG recovery (Fig. 6F). Further analysis revealed that peaks that were downregulated in ECs following *ERG* KD retained some accessibility relative to publicly available fibroblast ATAC-seq data (n = 5) (GSE121053) [25, 26], suggesting that EC identity genes may retain a 'bookmark' to facilitate re-expression upon ERG induction (Additional file 1: Fig. S6B).

Finally, we asked whether restoration of ERG was sufficient to resolve the barrier dysfunction and increased migration of *ERG* siRNA-treated teloHAECs. Immunofluorescence staining revealed that CD31 and VE-Cadherin were localized to cell–cell junctions, and cortical F-actin staining was regained following 8 days of recovery of ERG expression (Additional file 1: Fig. S6C). Migration speed and gap closure also returned to baseline following recovery of ERG expression (Fig. 6I). Collectively, these data demonstrate the inherent plasticity and resilience of ECs, as EndMT induced by ERG loss – including gene regulatory networks, cellular morphology, barrier integrity, and migratory responses – recover following restoration of ERG.

Discussion

Despite extensive evidence of EndMT occurring in numerous cardiovascular and cardiometabolic diseases [3, 7, 39], there remains uncertainty regarding the dynamics and mechanisms of this process. This could be related to the challenges in assessing EndMT in animal models or human samples as an end-state [40, 41], where the stepwise progression of cellular identity changes are difficult to study. Accordingly, the dynamic transcriptional regulatory networks controlling these transitions remain poorly characterized. Here, we identified a set of endothelial and EndMT identity markers that are regulated by ERG in aortic ECs, and monitored their chromatin, transcriptional, and protein-level changes, as well as the functional consequences of these alterations in ECs over time. By leveraging these gene sets in a time course analysis, we found that ERG loss compromised endothelial identity as early as 24 h post gene silencing at the epigenetic and transcriptional level, with expression of EC genes progressively diminishing over time. Phenotypic changes in EC morphology and behavior,

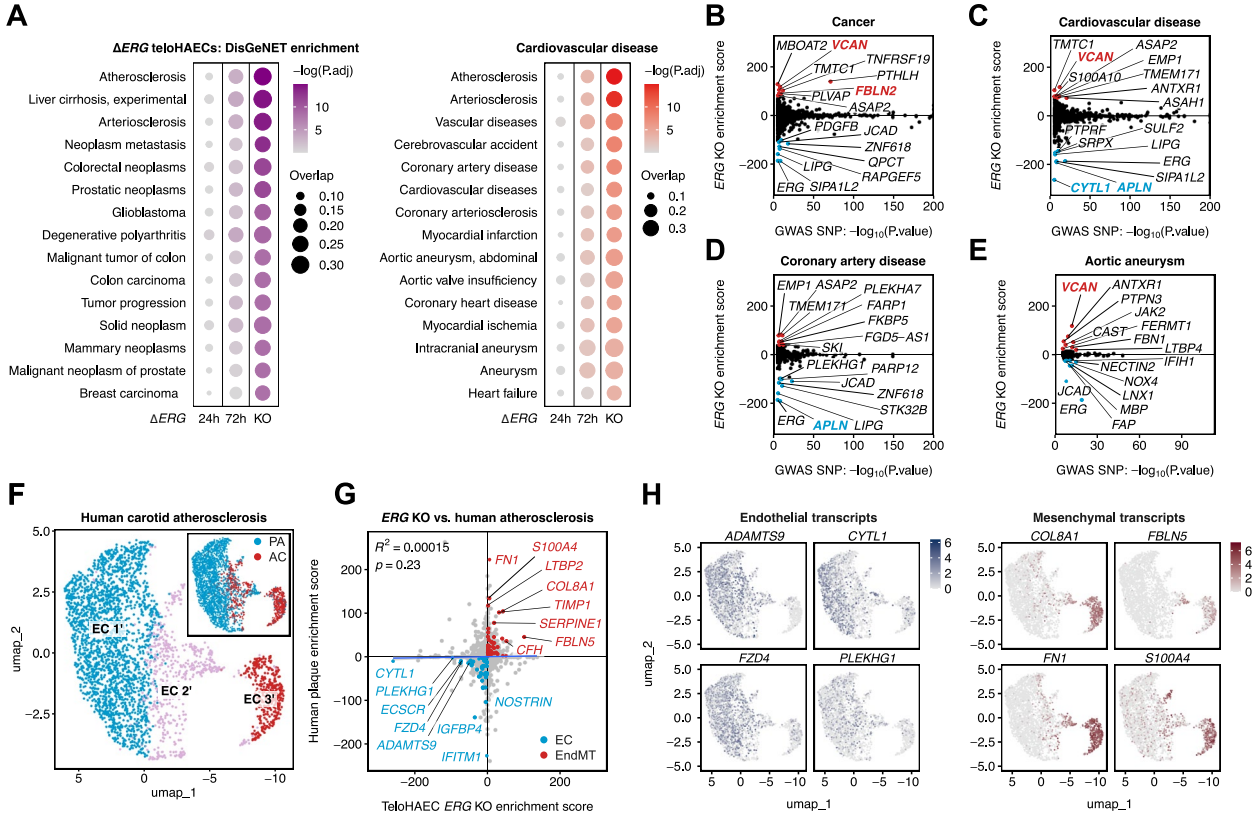


Fig. 5 Gene networks dysregulated by loss of ERG are associated with diseases of the arterial vasculature. **A** Disease enrichment of differentially expressed (DE) transcripts from *ERG*-deficient teloHAECs from *ERG* siRNA knockdown (KD) at 24 and 72 h to *ERG* knockout (KO) (n = 4–6; p.adj < 0.05). Top pathways by p-value are shown. **B**-**E** Genome-wide association study (GWAS) enrichment of DE transcripts from *ERG* KO teloHAECs and single nucleotide polymorphisms (SNPs) associated with cancer (B), cardiovascular disease (C), coronary artery disease (D), and aortic aneurysm (E) (p.adj < 0.05). *ERG* KO- and wild-type (WT)-enriched genes are denoted with red and blue text, respectively. **F** Uniform manifold approximation and projection (UMAP) plot of ECs from human carotid atherosclerotic core (AC) and proximal adjacent regions (PA). **G** Correlation plot between *ERG* KO-enriched teloHAEC transcripts and AC-enriched EC transcripts with select EC and endothelial-to-mesenchymal transition (EndMT) genes highlighted (p.adj < 0.05). **H** UMAP plots for DE endothelial and mesenchymal transcripts of interest

cytoskeletal rearrangement, and increased mesenchymal marker transcripts manifested at a later stage, only becoming predominant 72 h after loss of ERG. Our work suggests a sequential process of EndMT following ERG loss, whereby loss of EC identity precedes the acquisition of mesenchymal-like fate. Our work implies that ERG-dependent transcriptional programs preserve EC identity and may serve as a barrier to the acquisition of EndMT phenotypes.

Leveraging ATAC-seq and RNA-seq allowed us to identify the gene regulatory networks underpinning the progressive EndMT following compromised ERG function. The rapid rewiring of the transcriptional landscape upon loss of ERG was profound. More than 20% of the EC transcriptome was modulated by just 24 h of *ERG* KD, and > 40% of genes were altered in *ERG* KO ECs, with a roughly equal number of up- and downregulated genes, suggesting that ERG function is a critical regulator of EC transcriptional identity. Similarly, alterations in accessible chromatin regions were also pervasive, with > 6,000 DARs evident 24 h after *ERG* KD, increasing to > 17,000 differential regions in *ERG* KO ECs. Further analysis revealed that downregulated DARs showed a profound loss of ETS factor motif accessibility over time. Surprisingly, other ETS factors that are highly expressed in ECs, such as FLI1, were not able to compensate for ERG loss. In chronic *ERG* KO cells, AP1 motifs were also lost in downregulated peaks. In contrast, regions of increasing accessibility upon loss of ERG consistently showed a strong enrichment for AP1 motifs. SP/KLF motifs were also enriched in chronic *ERG* KO cells. This dual association of AP1 motifs with regions that are persistently upregulated alongside those lost with long-term ERG loss suggests AP1 may simultaneously compensate for ERG loss in the short term, while also driving EndMT via activating mesenchymal and ECM-related genes. We speculate that the collapse of EC identity regulatory regions may allow for AP1 factors to redistribute to mesenchymal and ECM-related genes. While AP1 factors have transcriptional pioneering properties [42], upregulated

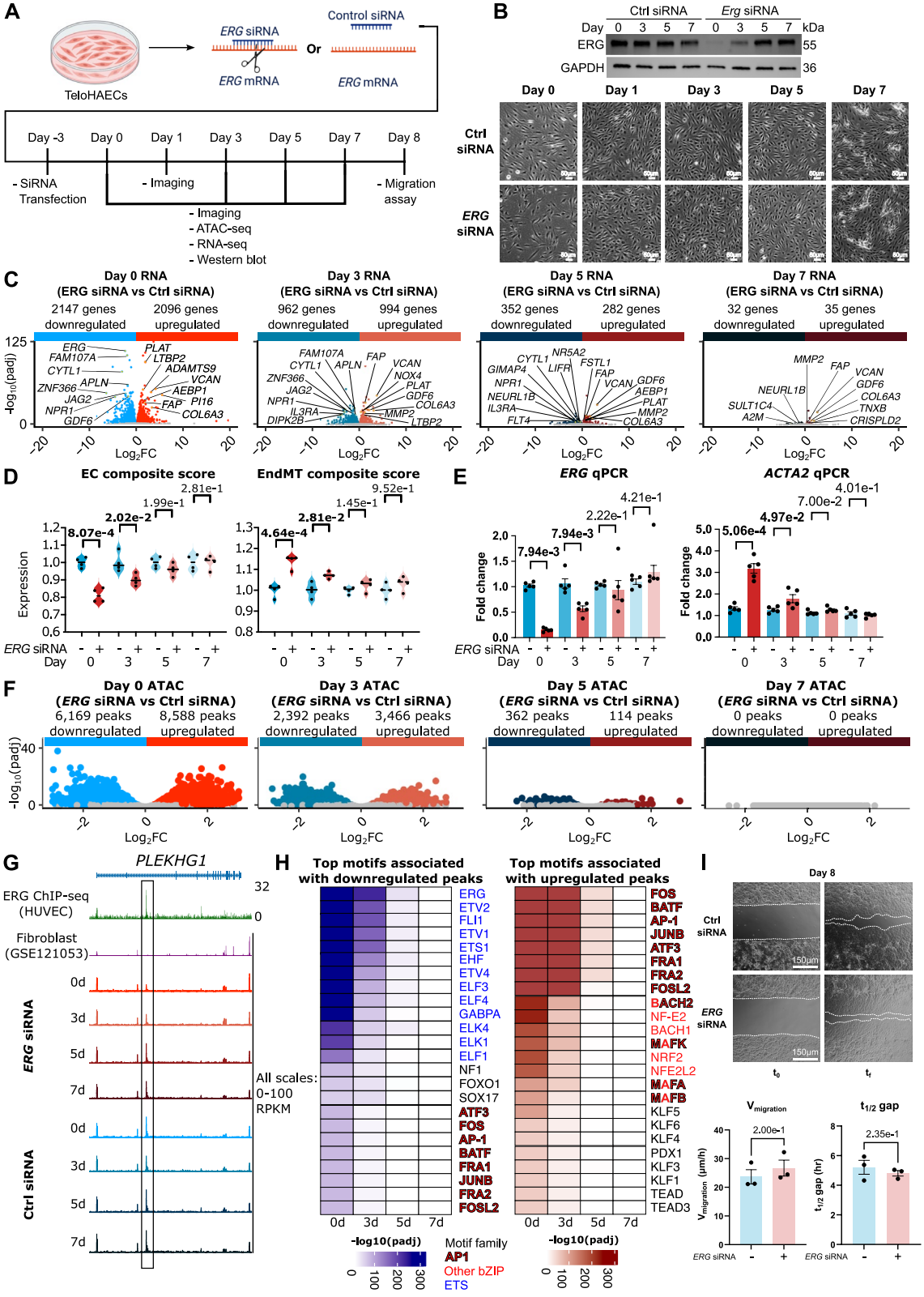


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Recovery of ERG abolishes EndMT gene regulatory networks and restores endothelial cell identity and function. **A** Schematic of ERG recovery experiments. ATAC-seq, assay for transposase-accessible chromatin with sequencing; RNA-seq, RNA sequencing; siRNA, small interfering RNA; TeloHAECs, telomerase-immortalized human aortic endothelial cells. **B** Representative western blot and brightfield imaging of *ERG* or Control (Ctrl) siRNA treated teloHAECs. **C** Volcano plots of RNA-seq data, with differentially expressed (DE) genes highlighted (p.adj < 0.05; $n = 4$). **D** Composite scores for DE genes associated with endothelial identity or endothelial-to-mesenchymal transition. **E** Quantitative PCR (qPCR) validation of the expression of *ERG* ($n = 5$; Mann–Whitney test) and the EndMT gene, *ACTA2* ($n = 5$; Welch's t-test). **F** Volcano plots of ATAC-seq data, with differentially accessible regions (DARs) highlighted (p.adj < 0.05; $n = 3$). **G** Genome browser track of an ERG-bound enhancer upstream of *PLEKHG1* that sees reduced accessibility upon *ERG* KD and is absent in fibroblasts (GSE121053). **H** HOMER motif enrichment analysis of DARs associated with downregulated (left) or upregulated (right) peaks in *ERG* siRNA treated teloHAECs. **I** Representative live-cell imaging depicting migration of Ctrl or *ERG* siRNA treated teloHAECs (top). Image post-barrier removal (left); image post-migration assay (right). Dashed lines denote migration front. Quantification of cell migration rate ($v_{migration}$; $n = 3$; Mann–Whitney test) and ½ gap maximal area ($t_{1/2}$; $n = 3$; unpaired t-test) (bottom). Values determined from plot of gap area over time. Each dot is the average of 1–5 fields of view along the length of the wound in each biological replicate. Bar graphs are presented as mean ± standard error of the mean (SEM). Violin plots indicate density of data points at corresponding values. Significant p-values are denoted with bold text. FC, fold change

peaks tend to have some modest level of accessibility in WT cells. This suggests that a mesenchymal gene program competes with an EC identity regulatory network for transcription factor recruitment.

Given the profound changes in the transcriptional and regulatory landscape upon loss of ERG, the reversibility of cellular identity changes was questionable. Indeed, there are reports in the literature of partial or complete EndMT. For example, in the setting of acute myocardial infarction, a partial EndMT phenotype was noted, which resolved over time without any discernable long-term effects on EC identity or function [43]. However, in the setting of atherosclerosis, where ECs undergoing EndMT delaminate from the lumen and migrate into the intima [6], it is not known if this cell state can be reversed. We therefore tested the permanence of EndMT by restoring native ERG expression. Impressively, recovery of endothelial gene regulatory networks, morphology, and function occurred following restoration of ERG. While the EC and EndMT composite scores were partially restored after 3 days, they were fully recovered by 5 days. Notably, the regulatory landscape was fully reset by 7 days. We found that EC identity regulatory elements maintained some degree of accessibility even in the absence of ERG, suggesting that these elements may be bookmarked to provide epigenetic memory of their EC origin and to facilitate their reinstatement upon ERG re-expression. The mechanisms that maintain this accessibility remain to be uncovered. It is also unknown if there is a 'point-of-no-return' where this accessibility might be lost, preventing reversal of an EndMT state. Our data provide insights into windows of opportunity where novel therapeutic strategies may be deployed to restore EC identity to combat EC dysfunction and cardiovascular disease.

Contextualizing our data with human disease datasets highlighted a strong role for ERG in cardiovascular disease. This analysis also highlighted enrichment with cancer pathways, which may be related to the loss of identity and enhanced mesenchymal and migratory phenotypes that occur during oncogenesis [44], as well as fibrotic liver disease. Indeed, ERG is downregulated in the setting of multiple cardiovascular and fibrotic diseases, including pulmonary hypertension [21], liver fibrosis [20], and atherosclerosis [13]. Our findings demonstrate the rapid molecular and phenotypic alterations that accompany ERG downregulation in ECs. They also suggest that restoration of ERG expression may reverse EC dysfunction by reestablishing EC identity while simultaneously suppressing EndMT phenotypes. The dynamic EC plasticity observed in our study implicates ERG as a fulcrum about which EC state transitions may tilt towards homeostasis or pathological remodeling, and sheds new light on the timing and reversible nature of this critical disease process.

Conclusions

Taken together, this study reveals the dynamic molecular underpinnings of EndMT and demonstrates that ERG regulates a network of EC identity genes while suppressing mesenchymal gene programs that are regulated by AP1 transcription factors. The acquisition of EndMT was found to involve a sequential loss of EC identity followed by gain of mesenchymal identity and was reversible upon restoration of ERG expression. The strong associations of ERG gene regulatory networks with transcriptional changes that occur in human atherosclerosis and linkage with genetic variants associated with cardiovascular disease, demonstrate that ERG may be a potential target to restore EC identity and to promote vascular resilience to combat cardiovascular disease.

Abbreviations

AP1	Activator protein 1
ATAC-seq	Assay for transposase-accessible chromatin with sequencing
ChIP-seq	Chromatin immunoprecipitation sequencing
Ctrl	Control
DAR	Differentially accessible region
DE	Differentially expressed
EC	Endothelial cell
ECGM-MV2	Endothelial cell growth media-microvascular 2
EndMT	Endothelial-to-mesenchymal transition
ERG	ETS-related gene
ETS	E-26 transformation specific
GESECA	Gene set co-regulation analysis
GO	Gene Ontology
GWAS	Genome-wide association study
HEG	High evidence gene
KD	Knockdown

KO	Knockout
PCA	Principal component analysis
RNA-seq	RNA sequencing
scRNA-seq	Single-cell RNA sequencing
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
teloHAEC	Telomerase-immortalized human aortic endothelial cell
TF	Transcription factor
WT	Wild-type

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-026-01638-6.

Additional file 1: Table S1. Guide RNA sequences for stable *ERG* gene disruption. Table S2. EC and EndMT gene lists. Table S3. Pathway category terms. Table S4. Primers for qPCR. Fig. S1. Chromatin accessibility profiling of *ERG* knockout in human aortic endothelial cells. Fig. S2. Chronic loss of *ERG* promotes endothelial-to-mesenchymal transition, disrupts junctional arrangement, and increases migration of aortic endothelial cells. Fig. S3. Induction of endothelial-to-mesenchymal transition is accompanied by activation of the AP1 transcription factor family. Fig. S4. Multiomic integration identifies conserved endothelial responses to loss of ERG. Fig. S5. Characterizing cell type composition of a publicly available human atherosclerosis single-cell RNA sequencing dataset. Fig. S6. Gene regulatory networks and endothelial function are restored upon ERG re-expression.

Additional file 2: Video S1. Live-cell migration assay in wild-type (WT) teloHAECs. Time-lapse confocal imaging of WT teloHAECs following removal of an Ibidi cell culture insert revealed that the cells exhibited collective migration into the gap area. Nuclear (blue) and F-actin (magenta) staining demonstrated that WT teloHAECs maintained an organized cytoskeletal structure during gap closure and that cell-cell junctions were largely preserved. As quantified in Figure 2, WT teloHAECs migrated at a slower rate than Δ*ERG* teloHAECs.

Additional file 3: Video S2. Live-cell migration assay in *ERG* knockout (Δ*ERG*) teloHAECs. Time-lapse confocal imaging of Δ*ERG* teloHAECs following removal of an Ibidi cell culture insert demonstrated that Δ*ERG* cells had increased migration into the wound area as compared to WT teloHAECs (quantified in Figure 2). Δ*ERG* cells had increased single-cell motility and decreased coordinated movement, leading to an irregular migration front. Nuclear (blue) and F-actin (magenta) staining revealed reduced cell-cell interactions.

Acknowledgements
Not applicable.

Authors' contributions
KS, SRB, KE, CAS, KLH, and JEF designed the research study and analyzed and interpreted results. JEF and KLH provided overall supervision and acquired funding for the study. MDW supervised the bioinformatic analyses. KS and CH conducted the cell culture experiments. RK and KEY acquired the RNA-seq and ATAC-seq data for the CRISPR/Cas9 ***ERG*** knockout experiments. SRB and KE acquired the RNA-seq and ATAC-seq data for the ***ERG*** knockdown experiments and analyzed the RNA-seq, ATAC-seq, and ChIP-seq data. NK generated and validated the CRISPR/Cas9 ***ERG*** knockout cell line. JDW, CLM, and MDW provided experimental and analytical expertise during the acquisition of experimental data and manuscript writing. KS, SRB, KE, CAS, KLH, and JEF wrote the manuscript. All authors provided feedback on the manuscript and approved the submission. The order of co-first authors was determined by time contributed to project conception and conducting experiments.

Funding
This work was supported by Medicine by Design, which received funding from the Canada First Research Excellence Fund (J.E.F.), the Canadian Institutes of Health Research (PJT-191915; K.L.H. and J.E.F. and PJT-173489; J.E.F.), the National Institutes of Health (R01HL164577 and R01HL148239; C.L.M.) and a Single-Cell Data Insights award from the Chan Zuckerberg Initiative, LLC, and Silicon Valley Community Foundation (C.L.M.). Studentship funding was received from CIHR (K.S., S.R.B. and K.E.), the Ontario Graduate Studentship program (S.R.B. and K.S.) and the Natural Sciences and Engineering Research Council (N.K.). Support was also received from a New Investigator Award from the Heart and Stroke Foundation of Canada (K.L.H.), the Wylie Scholar Award from Vascular Cures (K.L.H.) and the Blair Early Career Professorship in Vascular Surgery (K.L.H.). Infrastructure funding was obtained from the John R. Evans Leaders Fund from the Canada Foundation for Innovation (J.E.F.). The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Data availability
All raw and processed sequencing datasets generated in this study are available at the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/): RNA-Seq data (GSE292727; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292724) [45]; CATS-Seq CRISPR RNA-seq (GSE292725; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292725) [46]; ATAC-seq data (GSE292724; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292727) [47]. Code for data processing can be found at: https://github.com/kellisfm/Schulz_et_al_2025. Published human scRNA-seq datasets from human carotid atherosclerosis were from Alsaigh et al. (2022) and can be found at: GSE159677 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159677) [22, 23]. Published ATAC-seq from human pulmonary fibroblasts is available at: GSE121053 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121053) [25, 26].

Declarations

Ethics approval and consent to participate
This study did not involve human participants. Publicly available, deidentified scRNA-seq data from patients with carotid atherosclerosis were used (GSE159677) [22, 23]. Ethics approval was not required for utilizing this data and therefore the Helsinki Declaration was not relevant to this research.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada
[2]UHN Research Institutes, University Health Network, Toronto, Canada
[3]Institute of Medical Sciences, University of Toronto, Toronto, Canada
[4]Genetics and Genome Biology, SickKids Research Institute, Toronto, Canada
[5]Department of Molecular Genetics, University of Toronto, Toronto, Canada
[6]Basic Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA
[7]Department of Pediatrics, Division of Developmental Biology, University of Cincinnati College of Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[8]Department of Cell Biology, University of Virginia School of Medicine, Charlottesville, VA, USA
[9]Robert M. Berne Cardiovascular Research Center, University of Virginia School of Medicine, Charlottesville, VA, USA
[10]Brain, Immunology, and Glia (BIG) Center, University of Virginia School of Medicine, Charlottesville, VA, USA
[11]Department of Neuroscience, University of Virginia School of Medicine, Charlottesville, VA, USA
[12]Department of Genome Sciences, University of Virginia, Charlottesville, VA, USA
[13]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA
[14]Department of Surgery, Division of Vascular Surgery, University of Toronto, Toronto, Canada
[15]Peter Munk Cardiac Centre, Toronto General Hospital, Toronto, Canada
[16]Peter Munk Cardiac Centre, UHN Research Institutes, University Health Network, PMCRT 3-704, 101 College Street, Toronto, ON M5G 1L7, Canada
[17]Peter Munk Cardiac Centre, UHN Research Institutes, University Health Network, PMCRT 3-309, 101 College Street, Toronto, ON M5G 1L7, Canada

Received: 8 July 2025 / Accepted: 25 March 2026
Published online: 01 May 2026

References

1. Tamargo IA, Baek KI, Kim Y, Park C, Jo H. Flow-induced reprogramming of endothelial cells in atherosclerosis. Nat Rev Cardiol. 2023;20:738–53. https://doi.org/10.1038/s41569-023-00883-1.
2. Higashi Y. Endothelial function in dyslipidemia: roles of LDL-cholesterol, HDL-cholesterol and triglycerides. Cells. 2023;12:1293. https://doi.org/10.3390/cells12091293.
3. Yang D-R, Wang M-Y, Zhang C-L, Wang Y. Endothelial dysfunction in vascular complications of diabetes: a comprehensive review of mechanisms and implications. Front Endocrinol. 2024;15:1359255. https://doi.org/10.3389/fendo.2024.1359255.
4. Messner B, Bernhard D. Smoking and cardiovascular disease. Arterioscler Thromb Vasc Biol. 2014;34:509–15. https://doi.org/10.1161/ATVBAHA.113.300156.
5. Libby P. Inflammation in atherosclerosis. Nature. 2002;420:868–74. https://doi.org/10.1038/nature01323.
6. 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. Nat Commun. 2016;7:11853. https://doi.org/10.1038/ncomms11853.
7. 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. J Am Coll Cardiol. 2019;73:190–209. https://doi.org/10.1016/j.jacc.2018.09.089.
8. Newman AAC, Serbulea V, Baylis RA, Shankman LS, Bradley X, Alencar GF, et al. Multiple cell types contribute to the atherosclerotic lesion fibrous cap by PDGFRβ and bioenergetic mechanisms. Nat Metab. 2021;3:166–81. https://doi.org/10.1038/s42255-020-00338-8.
9. Park C, Baek KI, Hung R-C, Choi L, Jeong K, Kim P, et al. Disturbed flow induces reprogramming of endothelial cells to immune-like and foam cells under hypercholesterolaemia during atherogenesis. Cardiovasc Res. 2025;121:2679–99. https://doi.org/10.1093/cvr/cvaf233.
10. Helmke A, Casper J, Nordlohne J, David S, Haller H, Zeisberg EM, et al. Endothelial-to-mesenchymal transition shapes the atherosclerotic plaque and modulates macrophage function. FASEB J. 2019;33:2278–89. https://doi.org/10.1096/fj.201801238R.
11. Li X, Souilhol C, Canham L, Jia X, Diagbouga M, Ayllon BT, et al. Dll4 promotes partial endothelial-to-mesenchymal transition at atherosclerosis-prone regions of arteries. Vascul Pharmacol. 2023;150:107178. https://doi.org/10.1016/j.vph.2023.107178.
12. Gao Y, Galis ZS. Exploring the role of endothelial cell resilience in cardiovascular health and disease. Arterioscler Thromb Vasc Biol. 2021;41:179–85. https://doi.org/10.1161/ATVBAHA.120.314346.
13. 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. Arterioscler Thromb Vasc Biol. 2011;31:142–50. https://doi.org/10.1161/ATVBAHA.110.216473.
14. 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;144:2428–44. https://doi.org/10.1242/dev.146050.
15. Kalna V, Yang Y, Peghaire CR, Frudd K, Hannah R, Shah AV, et al. The transcription factor ERG regulates super-enhancers associated with an endothelial-specific gene expression program. Circ Res. 2019;124:1337–49. https://doi.org/10.1161/CIRCRESAHA.118.313788.
16. Wythe JD, Dang LTH, Devine WP, Boudreau E, Artap ST, He D, et al. Ets factors regulate Vegf-dependent arterial specification. Dev Cell. 2013;26:45–58. https://doi.org/10.1016/j.devcel.2013.06.007.
17. Shah AV, Birdsey GM, Randi AM. Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG. Vascul Pharmacol. 2016;86:3–13. https://doi.org/10.1016/j.vph.2016.05.003.
18. Farber G, Dong Y, Wang Q, Rathod M, Wang H, Dixit M, et al. Direct conversion of cardiac fibroblasts into endothelial-like cells using Sox17 and Erg. Nat Commun. 2024;15:4170. https://doi.org/10.1038/s41467-024-48354-6.
19. Farber G, Takasugi P, Ricketts S, Wang H, Xie Y, Farber E, et al. Sox17 and Erg synergistically activate endothelial cell fate in reprogramming fibroblasts. J Mol Cell Cardiol. 2025;199:33–45. https://doi.org/10.1016/j.yjmcc.2024.11.012.
20. 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. Nat Commun. 2017;8:895. https://doi.org/10.1038/s41467-017-01169-0.
21. Shinohara T, Moonen J-R, Chun YH, Lee-Yow YC, Okamura K, Szafron JM, et al. High shear stress reduces ERG causing endothelial-mesenchymal transition and pulmonary arterial hypertension. Arterioscler Thromb Vasc Biol. 2025;45:218–37. https://doi.org/10.1161/ATVBAHA.124.321092.
22. Alsaigh T, Evans D, Frankel D, Torkamani A. Decoding the transcriptome of calcified atherosclerotic plaque at single-cell resolution. Commun Biol. 2022;5:1084. https://doi.org/10.1038/s42003-022-04056-7.
23. Alsaigh T, Evans D, Frankel D, Torkamani A. Decoding the transcriptome of calcified atherosclerotic plaque at single-cell resolution. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse159677.
24. Birdsey GM, Kalna V. The transcription factor ERG regulates super-enhancers in endothelial cells [ERG siRNA]. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124892.
25. Markov GJ, Mai T, Nair S, Shcherbina A, Wang YX, Burns DM, et al. AP-1 is a temporally regulated dual gatekeeper of reprogramming to pluripotency. Proc Natl Acad Sci U S A. 2021;118:e2104841118. https://doi.org/10.1073/pnas.2104841118.
26. Markov GJ, Mai T, Nair S, Shcherbina A, Wang YX, Burns DM, et al. AP-1 is a temporally regulated dual gatekeeper of reprogramming to pluripotency. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121053.
27. Jonkman JEN, Cathcart JA, Xu F, Bartolini ME, Amon JE, Stevens KM, et al. An introduction to the wound healing assay using live-cell microscopy. Cell Adhes Migr. 2014;8:440–51. https://doi.org/10.4161/cam.36224.
28. Andreatta M, Carmona SJ. UCell: robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J. 2021;19:3796–8. https://doi.org/10.1016/j.csbj.2021.06.043.
29. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single–cell type transcriptomics map of human tissues. Sci Adv. 2021;7:eabh2169. https://doi.org/10.1126/sciadv.abh2169.
30. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51:D977–85. https://doi.org/10.1093/nar/gkac1010.
31. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28:495–501. https://doi.org/10.1038/nbt.1630.
32. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun. 2020;11:4267. https://doi.org/10.1038/s41467-020-18035-1.
33. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31:2382–3. https://doi.org/10.1093/bioinformatics/btv145.
34. 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. https://doi.org/10.7554/eLife.22536.
35. Ahmad M, Theofanidis P, Medford RM. Role of activating protein-1 in the regulation of the vascular cell adhesion molecule-1 gene expression by tumor necrosis factor-alpha. J Biol Chem. 1998;273:4616–21. https://doi.org/10.1074/jbc.273.8.4616.
36. Hollenhorst PC, Jones DA, Graves BJ. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic Acids Res. 2004;32:5693–702. https://doi.org/10.1093/nar/gkh906.
37. Abiko H, Fujiwara S, Ohashi K, Hiatari R, Mashiko T, Sakamoto N, et al. Rho guanine nucleotide exchange factors involved in cyclic-stretch-induced reorientation of vascular endothelial cells. J Cell Sci. 2015;128:1683–95. https://doi.org/10.1242/jcs.157503.
38. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48:D845–55. https://doi.org/10.1093/nar/gkz1021.
39. Alvandi Z, Bischoff J. Endothelial-Mesenchymal Transition in Cardiovascular Disease. Arterioscler Thromb Vasc Biol. 2021;41:2357–69. https://doi.org/10.1161/ATVBAHA.121.313788.

40. Chen P-Y, Qin L, Baeyens N, Li G, Afolabi T, Budatha M, et al. Endothelial-to-mesenchymal transition drives atherosclerosis progression. J Clin Invest. 2015;125:4514–28. https://doi.org/10.1172/JCI82719.
41. Ranchoux B, Antigny F, Rucker-Martin C, Hautefort A, Péchoux C, Bogaard HJ, et al. Endothelial-to-Mesenchymal Transition in Pulmonary Hypertension. Circulation. 2015;131:1006–18. https://doi.org/10.1161/CIRCULATIONAHA.114.008750.
42. Vierbuchen T, Ling E, Cowley CJ, Couch CH, Wang X, Harmin DA, et al. AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol Cell. 2017;68:1067-1082.e12. https://doi.org/10.1016/j.molcel.2017.11.026.
43. Tombor LS, John D, Glaser SF, Luxán G, Forte E, Furtado M, et al. Single cell sequencing reveals endothelial plasticity with transient mesenchymal activation after myocardial infarction. Nat Commun. 2021;12:681. https://doi.org/10.1038/s41467-021-20905-1.
44. Youssef KK, Narwade N, Arcas A, Marquez-Galera A, Jiménez-Castaño R, Lopez-Blau C, et al. Two distinct epithelial-to-mesenchymal transition programs control invasion and inflammation in segregated tumor cell populations. Nat Cancer. 2024;5:1660–80. https://doi.org/10.1038/s43018-024-00839-5.
45. Schulz K, Botts SR, Ellis K, Wilson MD, Fish JE, Howe KL. Dynamic rewiring of cell identity transcriptional programs by loss of ERG in aortic endothelial cells [RNA]. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292727.
46. Schulz K, Botts SR, Ellis K, Wilson MD, Fish JE, Howe KL. Dynamic rewiring of cell identity transcriptional programs by loss of ERG in aortic endothelial cells [CATS]. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292725.
47. Schulz K, Botts SR, Ellis K, Wilson MD, Fish JE, Howe KL. Dynamic rewiring of cell identity transcriptional programs by loss of ERG in aortic endothelial cells [ATAC-Seq]. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292724.

Publisher's Note