

The Transcription Factor ERG Regulates Super-Enhancers Associated with an Endothelial-Specific Gene Expression Program

Viktoria Kalna¹, Youwen Yang^{1&}, Claire R. Peghaire¹, Karen Frudd¹, Rebecca Hannah², Aarti V. Shah¹, Lourdes Osuna Almagro^{1#}, Joseph J. Boyle¹, Berthold Göttgens², Jorge Ferrer³, Anna M. Randi^{1*}, Graeme M. Birdsey¹

¹National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, London, UK; ²Department of Haematology, Wellcome Trust and MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK; ³Epigenomics and Disease, Department of Medicine, Imperial College London, London, UK; [&] Current address: Department of Cell and Developmental Biology, University College London, London, UK, and; [#] Current address: Brain Metastasis Group, Centro Nacional de Investigaciones Oncológicas, Madrid, Spain.

Running title: ERG Regulates Endothelial Super-Enhancers



Circulation Research

Subject Terms:

Angiogenesis
Computational Biology
Endothelium/Vascular Type/Nitric Oxide
Epigenetics
Vascular Biology

Address correspondence to:

Dr. Anna M. Randi
National Heart and Lung Institute
Imperial College London
Hammersmith Hospital
Du Cane Road
London, W12 0NN
United Kingdom
a.randi@imperial.ac.uk

Dr. Graeme M. Birdsey
National Heart and Lung Institute
Imperial College London
Hammersmith Hospital
Du Cane Road
London, W12 0NN
United Kingdom
g.birdsey@imperial.ac.uk

ABSTRACT

Rationale: The ETS transcription factor (TF) ERG is essential for endothelial homeostasis, driving expression of lineage genes and repressing pro-inflammatory genes. Loss of ERG expression is associated with diseases including atherosclerosis. ERG's homeostatic function is lineage-specific, since aberrant ERG expression in cancer is oncogenic. The molecular basis for ERG lineage-specific activity is unknown. Transcriptional regulation of lineage specificity is linked to enhancer clusters (super-enhancers).

Objective: To investigate whether ERG regulates endothelial-specific gene expression via super-enhancers.

Methods and Results: Chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) in human umbilical vein endothelial cells (HUVEC) showed that ERG binds 93% of super-enhancers ranked according to H3K27ac, a mark of active chromatin. These were associated with endothelial genes such as *DLL4*, *CLDN5*, *VWF* and *CDH5*. Comparison between HUVEC and prostate cancer TMPRSS2:ERG fusion-positive VCaP cells revealed distinctive lineage-specific transcriptome and super-enhancer profiles. At a subset of endothelial super-enhancers (including *DLL4* and *CLDN5*), loss of ERG results in significant reduction in gene expression which correlates with decreased enrichment of H3K27ac and Mediator subunit MED1, and reduced recruitment of acetyltransferase p300. At these super-enhancers, co-occupancy of GATA2 and AP-1 is significantly lower compared to super-enhancers that remained constant following ERG inhibition. These data suggest distinct mechanisms of super-enhancer regulation in EC and highlight the unique role of ERG in controlling a core subset of super-enhancers. Most disease-associated single nucleotide polymorphisms (SNPs) from genome-wide association studies (GWAS) lie within noncoding regions and perturb TF recognition sequences in relevant cell types. Analysis of GWAS data shows significant enrichment of risk variants for CVD and other diseases, at ERG endothelial enhancers and super-enhancers.

Conclusions: The TF ERG promotes endothelial homeostasis via regulation of lineage-specific enhancers and super-enhancers. Enrichment of CVD-associated SNPs at ERG super-enhancers suggests that ERG-dependent transcription modulates disease risk.

Keywords:

Epigenetics, gene expression and regulation, endothelium, endothelial cell, transcription factors, super-enhancers.

Nonstandard Abbreviations and Acronyms:

BRD4	Bromodomain-containing protein 4
CAD	Coronary artery disease
ChIP-seq	Chromatin immunoprecipitation with deep sequencing
CVD	Cardiovascular disease
EC	Endothelial cell
ERG	Ets-Related Gene
ETS	E-26 transformation specific transcription factor
ENCODE	Encyclopedia of DNA Elements
GO	Gene ontology
GSEA	Gene set enrichment analysis
GWAS	Genome-wide association study
HAEC	Human aortic endothelial cell
HUVEC	Human umbilical vein endothelial cell
LDTF	Lineage determining transcription factor
MED1	Mediator complex subunit 1
SE	Super-enhancers
SNP	Single nucleotide polymorphism
TF	Transcription factor
TMPRSS2	Transmembrane protease, serine 2
TSS	Transcription start site
VCaP	Human prostate epithelial cancer cell line



INTRODUCTION

Maintenance of endothelial homeostasis is essential for vascular health. Disruption of endothelial homeostasis, as observed in atherosclerosis and in chronic inflammatory diseases, leads to profound changes in the phenotype of endothelial cells (EC) with upregulation of pro-inflammatory pathways and loss of anti-inflammatory pathways. Loss of endothelial lineage identity is associated with endothelial-to-mesenchymal transition (EndMT), a process implicated in multiple diseases¹. The transcriptional mechanisms regulating EC lineage identity and maintenance of endothelial homeostasis are also areas of immense interest for vascular regenerative therapies, but remain poorly understood.

The ETS transcription factor ERG is a critical regulator of endothelial homeostasis (reviewed in ²). In the endothelium, ERG expression appears around developmental day E8.5 and is maintained into adulthood³. ERG is required for endothelial lineage specification⁴, vascular development and angiogenesis⁵; endothelial-specific deletion of *Erg* in mouse results in embryonic lethality due to vascular defects^{6,7}. ERG drives expression of lineage-specific genes such as VE-cadherin (*CDH5*), *DLL4*, claudin-5 (*CLDN5*) and von Willebrand factor (*VWF*) and controls processes including survival, permeability and cytoskeletal dynamics (reviewed in ²). Molecular pathways through which ERG promotes vascular stability and angiogenesis include Wnt/ β -catenin signaling⁶ and angiopoietin-1-dependent Notch signaling⁸. ERG maintains vascular homeostasis also by repressing expression of pro-inflammatory genes such as *ICAM1* and *IL8*^{9,10} and by protecting from EndMT¹¹. In line with its homeostatic role, ERG's expression is lost in vascular diseases such as the activated endothelium overlying human atherosclerotic plaques¹⁰. However, aberrant expression of ERG in non-EC can be detrimental. ERG overexpression as the result of chromosomal translocations in prostate cancer correlates with malignancy and invasiveness, poor prognosis

and shorter survival times¹². In these circumstances ERG acts as an oncogene. Since the first report of gene fusions between ERG and the regulatory region of the androgen-dependent TMPRSS2 gene¹³, the molecular mechanisms through which ERG aberrant expression is oncogenic have been investigated in detail (reviewed in¹⁴). The striking difference between the homeostatic versus oncogenic roles of ERG has important implications for the possible therapeutic potential of this pathway in pathologies associated with cardiovascular disease (CVD).

The mechanistic basis for the difference in ERG's lineage specific activity may lie in the chromatin landscape. Genetic regulatory elements, called enhancers, play a key role in mediating the transcriptional regulation of lineage-specific gene expression¹⁵. Enhancer activation by transcription factors (TF) is cooperative and hierarchical¹⁵⁻¹⁷. Members of the ETS, AP-1 and GATA families have been shown to functionally interact at enhancers associated with lineage-specific genes¹⁶. Recently, clusters of enhancer elements variably termed super-enhancers, stretch enhancers or enhancer clusters, have been described in numerous cell types, including EC¹⁸⁻²¹. These regulatory elements are associated with an extremely high abundance of TFs, H3K27ac-modified nucleosomes and Mediator complex and drive cell type-specific gene expression^{18, 19}.

Furthermore, super-enhancers are enriched in single nucleotide polymorphisms (SNPs) associated with specific diseases, in a cell type-specific manner¹⁸. The majority of SNPs identified by genome-wide association studies (GWAS) associated with human disease traits are localized to non-coding regions of the genome²², including promoters and enhancers, and frequently perturb TF recognition sequences. Thus SNPs associated with CVD risk may affect TF binding sites within super-enhancers in EC and other cell types relevant to CVD.

In this study, we show that ERG regulates a subset of endothelial super-enhancers, and that CVD-associated SNPs are enriched at ERG enhancers and super-enhancers. The association of ERG-bound loci with CVD risk variants provides candidate SNPs for future studies on the epigenomic pathways underlying cardiovascular pathologies.

METHODS

All supporting data are available within the article and its Online Data Supplement. Sequencing data generated in this study (ChIP-Seq datasets for ERG, H3K27ac and MED1 in HUVEC) have been made publicly available at the NCBI Gene Expression Omnibus (GEO), and can be retrieved through accession number GSE124893.

An expanded Methods section is available in the Online Supplement.

RESULTS

Integrated genomic analysis reveals an ERG-regulated transcriptional program for the vascular endothelium.

ChIP-seq analysis in HUVEC identified 40821 genomic ERG binding sites associated with 14786 genes. Selected ERG peaks in the promoters of known ERG target genes *CDH5* and *ICAM1* were validated by ChIP-qPCR (Online Figure IA). As expected, the canonical ERG motif (C/a/g)(A/C)GGAA(G/A)²³ was the most represented at ERG sites in HUVEC (Online Figure IB). Globally, analysis of the distribution of

ERG binding sites relative to annotated transcription start sites (TSS) revealed that most ERG-bound regions were intragenic and intergenic sites located distally from the promoter (+/- 2kb from TSS) (Figure 1A).

Integrated analysis of global expression profiling²⁴ with ERG ChIP-seq showed that ERG binds 85% (1232/1454) of its activated targets and 80% (939/1180) of its repressed targets (Online Figure IC). Gene ontology (GO) pathway analysis revealed that directly activated ERG targets clustered in functions related to angiogenesis, blood vessel morphogenesis and hemostasis, while repressed genes associated with TGF β /SMAD signaling and stress pathways (Figure 1B), in line with its known roles.

ERG-bound enhancers drive endothelial gene expression.

We next examined the relationship between ERG binding and chromatin states in HUVEC using data from the Encyclopedia of DNA Elements (ENCODE) Consortium²⁵. 97% of ERG peaks mapped to regions of DNase I hypersensitivity, a marker of accessible open chromatin²⁶. Analysis of known ERG target genes showed ERG genomic loci overlapping histone marks of active promoters (H3K4me3 and H3K27ac) and enhancers (H3K4me1 and H3K27ac) at sites of DNase I hypersensitivity (Figure 1C and Online Figure ID). Globally, ERG binding in HUVEC was greatest at active enhancers (Figure 1D).

ERG-bound enhancers were identified in known ERG targets, including *CDH5*. Four ERG-bound active enhancers, named E1 to E4, were selected based on H3K27ac and H3K4me1 enrichment in a 23kb region along the *CDH5* locus either side of the TSS (Online Figure IIA). The E1, E2 and E4 enhancers, but not the repeat DNA-containing E3 enhancer, were individually cloned into luciferase reporter vectors containing the *CDH5* promoter, previously characterized as transactivated by ERG²⁷. In HeLa cells (which do not express endogenous ERG), all enhancers increased ERG-dependent transactivation of the *CDH5* promoter, with E4 being the most active (data not shown). This was then validated in HUVEC, where the E4 enhancer increased basal *CDH5* promoter activity >4-fold (Figure 1E); moreover, this region was responsive to ERG transactivation, which further increased luciferase activity by 9-fold compared to *CDH5* promoter alone (Figure 1E). Mutation of the nine AGGAA putative ERG-binding motifs in region E4 (Online Figure IIB, C) completely abolished enhancer activity and the response to ERG (Figure 1E). These findings support a key role for ERG-mediated transactivation of gene expression through EC enhancers.

ERG binds to HUVEC super-enhancers.

Multiple ERG-bound enhancers were found in close proximity with each other in ERG activated genes, such as *CDH5* (see Figure 1C). Clusters of enhancers, known as super-enhancers can be distinguished from isolated typical enhancers by enrichment in lineage-specific TF, co-activators such as Mediator complex subunit 1 (MED1) and the histone modification mark H3K27ac^{18, 19, 28}. Super-enhancers preferentially associate with genes that define cell lineage identity^{18, 19}. To define endothelial super-enhancers, we identified enhancer regions by co-occupancy of H3K27ac and H3K4me1 in HUVEC²⁵. Enhancers within 12.5kb of each other were then concatenated to define a single entity, and ranked by increasing H3K27ac enrichment, as described²⁸. The analysis identified 917 super-enhancers (Figure 2A and Online Table I) that mapped to 822 genes, including ERG activated targets *VWF*, *CDH5*, *ICAM2*, *SOX17*, *DLL4*, as well as *ERG* itself (Figure 2A, B). A similar super-enhancer profile was obtained when ranked by MED1 enrichment (Online Figure IIIA).

Genes associated with endothelial super-enhancers were found to have significantly higher mean expression levels compared to those associated with typical enhancers (Figure 2C). Gene Set Enrichment Analysis (GSEA) showed significant enrichment of ERG driven genes with the top 500 ranked super-

enhancer genes (Figure 2D). These data suggest that ERG regulates endothelial gene expression via super-enhancers.

Remarkably, the vast majority of super-enhancers (93%) were bound by ERG, compared to only 34% of typical enhancers (Figure 3A). In keeping with higher TF occupancy at super-enhancers¹⁹, the canonical ERG motif was significantly more bound by ERG in super-enhancers compared to typical enhancers (35% vs 25%, respectively) (Online Figure IIIB). Furthermore, within the 917 super-enhancers, H3K27ac and ERG binding signal significantly correlated (Figure 3B). We thus tested whether ERG itself could be used to identify super-enhancers in EC. Using ERG enrichment at active enhancers as the ranking parameter, we identified 1125 super-enhancers in HUVEC (Figure 3C), associated with a similar gene set as the H3K27ac super-enhancers (see Figure 2A). Indeed, GSEA demonstrated a strong positive correlation between super-enhancers defined by ERG and those by H3K27ac (Figure 3D). Moreover, functional clustering of H3K27ac super-enhancers and ERG super-enhancers revealed shared pathways essential to EC identity and function (Figure 3E).

Thus, ERG binding identifies super-enhancers in differentiated endothelial cells and supports the prominent role of ERG as a lineage-determining transcription factor for the vascular endothelium.

Differential super-enhancers binding underlies the lineage-specific activity of ERG.

At odds with its homeostatic role in EC, aberrant ERG expression due to chromosomal translocations, such as the TMPRSS2:ERG gene fusions in prostate cells, is a hallmark of cancer (reviewed in ¹⁴). To exploit ERG's therapeutic potential in the vasculature it is crucial to understand the molecular basis of its lineage-specific role. Therefore, we compared ERG bound gene targets in HUVEC with those from VCaP prostate cancer cells carrying the TMPRSS2:ERG gene fusion. Two publicly available ERG ChIP-seq datasets in VCaP cells^{29, 30} were found to correlate highly (Online Figure IVA); we used the data in Chng *et al.*²⁹ for further analysis. Integration of this ChIP-seq data with transcriptome profiling of ERG-depleted VCaP cells³¹ identified 584 genes bound and activated by ERG and 589 genes bound and repressed by ERG (Online Figure IVB). GO pathway analysis showed regulation of genes involved in DNA replication, cell proliferation, cytoskeletal remodeling and apoptosis (Figure 4A). We next examined the overlap between genes directly bound by ERG in VCaP cells versus HUVEC (Online Figure IVC). Interestingly, only 249 ERG bound target genes (8%) were found in common between HUVEC and VCaP cells (Figure 4B and Online Table II); of these only 58% were activated or repressed concurrently in both cell types (Figure 4C). Interestingly, even pathways controlled by ERG in both HUVEC and VCaP cells (such as cell migration, adhesion and Notch signaling) are regulated in a lineage-specific manner (Online Figure IVD). Selected ERG transcriptional targets were validated in HUVEC and VCaP cells by RT-qPCR (Online Figure IVE).

Comparison of ChIP-seq datasets between HUVEC and VCaP cells showed that only 23% of ERG bound sites are in common between the HUVEC and VCaP genomes (Online Figure IVF). Interestingly, the majority (70%) of these shared sites are located close (± 1 kb) to the TSS (Figure 4D). In contrast, the proportion of ERG binding sites unique to HUVEC or VCaP cells are preferentially located at sites distal to the TSS (Figure 4D). Mapping of ERG genomic binding with histone modification marks confirmed a significant overlap of ERG binding to promoters in HUVEC and VCaP cells (Figure 4E, left). However, no significant overlap was observed between ERG binding at enhancers in HUVEC and VCaP cells, with only 18% of ERG binding sites found at enhancers in VCaP cells (Figure 4E, right). We next asked whether ERG binding in VCaP cells was also associated with super-enhancers. Using ranked enrichment of H3K27ac from ChIP-seq in VCaP cells³², we identified 208 super-enhancers (Figure 5A). Genes associated with super-enhancers in VCaP had significantly higher average expression levels compared to those associated with isolated typical enhancers (Online Figure VA), as expected. The vast majority of VCaP super-enhancers were bound by ERG (91%), compared to 48% of typical enhancers (Figure 5B).

Interestingly, GSEA showed no significant relationship between super-enhancer-associated genes in VCaP cells and HUVEC (Online Figure VB). GO analysis revealed different pathways regulated by super-enhancer-associated genes in HUVEC compared to VCaP cells (Figure 5C). Finally, super-enhancers occupied by ERG showed cell-lineage specificity: endothelial ERG-regulated genes (such as *CDH5*) are associated with super-enhancers in HUVEC but not VCaP cells, and *vice versa* for VCaP cell genes (such as *TMPRSS2*) (Figure 5D and Online Figure VC).

The difference between super-enhancer profiles in HUVEC versus VCaP cells suggests that the chromatin landscape is unique to the particular cell type. Analysis of histone modifications associated with active (H3K27ac and H3K4me1) or repressed (H3K27me3) chromatin in HUVEC and VCaP super-enhancer regions supports this hypothesis. Genomic regions corresponding to HUVEC SE were enriched in active histone marks in HUVEC but not VCaP (Figure 5E, top). Conversely, genomic regions corresponding to VCaP SE were enriched in active marks in VCaP but not HUVEC (Figure 5E, bottom). The reverse pattern was observed for the repressive mark H3K27me3 (Online Figure VD).

To further define the lineage-specific machinery of super-enhancers, we focused on TFs that may collaborate with ERG to regulate gene transcription in a cell type-specific manner. We searched for TF DNA motifs located +/-200bp from the ERG binding sites in HUVEC and VCaP cells. Interestingly, different TF motifs are enriched in ERG binding sites in the two cell types. In HUVEC, these include AP-1, FOXO1, GATA2 and SOX3, TFs known to be important in endothelial gene expression^{16,33} (Figure 5F, top). In VCaP cells, the enriched motifs included TFs FOXA1 and HOXB13, TFs previously described to play a role in prostate cancer gene expression³⁴ (Figure 5F, bottom). These data indicate that ERG's lineage-specific transcriptional activity is associated with binding to cell type-specific super-enhancers, and suggests cooperativity with distinct lineage-specific factors.

ERG controls the gene expression profile in HUVEC by regulating the enhancer and super-enhancer landscape.

We investigated whether ERG is required for H3K27 acetylation at endothelial enhancers by performing H3K27ac ChIP-seq analysis in HUVEC treated with control or ERG-siRNA (Online Figure VIA,B). In control HUVEC, 56347 H3K27ac-bound regions were identified, which significantly correlated with those reported in the ENCODE data²⁵ (Online Figure VIC). In ERG-deficient cells, H3K27ac was modulated globally, with a decrease in H3K27ac at 5277 regions (loss) and an increase in 1648 regions (gain) (Figure 6A,B). Changes in H3K27ac enrichment were validated by ChIP-qPCR across regions associated with selected endothelial ERG target genes (Figure 6C). Globally, in ERG-deficient cells changes in H3K27ac also correlated with the expression profile: expression of genes associated with loss of H3K27ac was significantly downregulated whilst expression of genes associated with gain of H3K27ac was significantly upregulated (Figure 6D and Online Figure VID). Genomic Regions Enrichment of Annotations Tool (GREAT) analysis of ERG-depleted HUVEC showed that loss of H3K27ac was associated with enrichment of Notch and VEGF receptor signaling, pathways positively controlled by ERG^{8,35} (Figure 6E). In contrast, gain of H3K27ac was associated with TGFβ-SMAD signaling, a pathway repressed by ERG¹¹ (Figure 6E).

To investigate the effect of ERG depletion on the recruitment of basal transcriptional machinery to enhancers, we carried out ChIP-seq for MED1 in control and ERG-deficient HUVEC. Loss or gain of MED1 occupancy in ERG-deficient cells coincided with a decrease or increase in H3K27ac, respectively (Figure 6F). These data indicate that ERG plays a role in modulating endothelial enhancers. MED1 is part of a large complex (Mediator) which interacts with super-enhancers¹⁹. We therefore investigated the role of ERG in the organization of endothelial super-enhancers. H3K27ac ChIP-seq analysis in control HUVEC identified 1015 super-enhancer clusters (Figure 7A, siCtl), in line with the HUVEC super-enhancers profile identified from ENCODE data (see Figure 2A). ERG depletion by siRNA caused changes in H3K27ac

levels leading to a redistribution of endothelial super-enhancers (Figure 7A; siERG). Comparison of H3K27ac super-enhancers in control versus ERG-depleted HUVEC identified a subset of 107 super-enhancers with decreased H3K27ac levels following loss of ERG. Amongst the ERG-regulated super-enhancers were those associated with key endothelial genes including *DLL4*, *NRARP* and *CLDN5* (Figure 7B). The majority of super-enhancers showed no significant changes following ERG siRNA, and only a few super-enhancers showed increased H3K27ac. MED1 occupancy was also reduced in the subset of ERG-regulated super-enhancers (*decreased SE*) compared to those unchanged (*constant SE*) (Figure 7C). Importantly, the ERG-dependent decrease in H3K27ac levels correlates with reduced expression of ERG target genes (Online Figure VIIA). Thus, ERG is functionally required to dynamically modulate H3K27ac levels in endothelial cells leading to redistribution of a subset of super-enhancers.

Cooperative TF binding and p300 recruitment in the regulation of HUVEC super-enhancers

ERG has been shown to bind to p300³⁵; thus we hypothesized that ERG-dependent changes in H3K27ac at super-enhancers might be linked to the recruitment of p300 by ERG. This was tested by ChIP-qPCR for p300 enrichment on selected loci associated with validated ERG targets (*CLDN5*, *DLL4*), where H3K27ac was decreased upon loss of ERG (*decreased SE*). These showed a significant decrease in p300 occupancy following ERG inhibition, suggesting that ERG is required to recruit p300 at these sites (Figure 7D). However, ERG inhibition did not consistently affect p300 recruitment at *constant SE* typified by *IL6* and *PXN* (Figure 7D). Online Figure VIIB illustrates the ERG-dependent decrease in H3K27ac and MED1 occupancy observed at the *CLDN5* and *DLL4* gene loci, compared to loci associated with *constant SE*, *IL6* and *PXN*. These findings suggest that ERG regulates a subset of super-enhancers partly through recruitment of the histone acetyltransferase p300.

We speculated that in the super-enhancers that remain constant following loss of ERG, other TFs might compensate for its absence. Previous studies have identified GATA and AP-1 (FOS/JUN) TF families as cooperating with ETS factors in regulating endothelial gene expression^{16, 36}; moreover, cJUN has been shown to bind p300³⁷. Analysis of ChIP-seq data from ENCODE for GATA2, cFOS, and cJUN in HUVEC²⁵ showed significant global overlap with ERG-bound sites (Online Figure VIIC). Higher occupancy of GATA2, cFOS and cJUN was present at *constant SE* compared to *decreased SE* (Figure 7E). This global distribution is reflected at representative loci for the two groups; *CLDN5*, *DLL4* (*decreased SE*) and *IL6*, *PXN* (*constant SE*) (Online Figure VIIB).

These data suggest a model (Figure 7F) in which the majority of SE are regulated by a cooperative TF network involving ERG, AP-1 and GATA2 that provide a strong transcriptional complex; thus loss of ERG can be compensated. However, in a subset of SE-associated lineage genes including *CLDN5* and *DLL4*, AP-1 and GATA2 are less abundant, and therefore there is low cooperativity and SE assembly and gene expression are strongly dependent on ERG.

Risk variants for cardiovascular and other diseases are enriched at ERG super-enhancers.

Several studies have recently shown that disease-associated SNPs identified through GWAS are preferentially enriched in the super-enhancer regions of disease-relevant cells, and can map to TF binding sites^{18, 38}. Endothelial dysfunction is implicated in many diseases. We examined the enrichment of disease-associated variants at ERG binding loci, ERG-bound enhancers and ERG super-enhancers, using SNPs reported in the NCBI dbGaP³⁹ and NHGRI-GWAS catalogs⁴⁰. We determined enrichment by using a null distribution of background population variants. Analysis at ERG binding loci identified association with SNPs for immune diseases and CVD (Online Figure VIIIA). At ERG-bound enhancers the significance of enrichment is greatly amplified, with a similar repertoire of disease traits (Online Figure VIIIB). Interestingly, analysis at ERG super-enhancers identified SNPs for CVD as the most highly associated disease trait ($P = 1.1 \times 10^{-14}$) (Figure 8A). ERG super-enhancers were also enriched in diseases for digestive

system ($P = 5.7 \times 10^{-5}$) and respiratory tract ($P = 6.1 \times 10^{-5}$) (Figure 8A). This enrichment was not identified at size- and chromosome-matched regions randomly shuffled (permuted) across permissive chromatin. Further interrogation of the CVD-associated SNPs revealed strong significant enrichment for traits such as abdominal aortic aneurysm ($P = 1.7 \times 10^{-13}$), coronary artery disease ($P = 2.6 \times 10^{-10}$), myocardial infarction ($P = 5.1 \times 10^{-9}$) and hypertension ($P = 1.1 \times 10^{-8}$), but not heart failure (Figure 8B), suggesting a closer association with diseases where EC contribute most to the pathogenesis.

These results identify a novel mechanisms through which disease associated non-coding SNPs may cause vascular dysfunction and increased disease risk, and suggest a possible functional link between ERG-dependent transcriptional regulation of endothelial gene expression and the predisposition to CVD.

DISCUSSION

In this study, we characterize the transcription factor ERG as a crucial regulator of enhancers and super-enhancers in HUVEC. Multiple studies have shown that ERG is essential to maintain endothelial homeostasis (reviewed in ²). Here we define the ERG-dependent endothelial epigenome and associate this with genetic variants linked to CVD and other diseases, suggesting novel potential strategies for biomarkers and target identification.

Our study identifies ERG as a positive regulator of a core set of putative endothelial super-enhancers in HUVEC. We describe an ERG-dependent subset of super-enhancers, associated with essential endothelial genes such as *DLL4*, *CLDN5* and *NRARP*. Crucially, ERG-dependent super-enhancers are significantly associated with ERG-activated genes. Analysis of *DLL4* and *CLDN5* loci shows that recruitment of p300 at these super-enhancer is controlled by ERG; direct interaction between ERG and p300 ³⁵ suggests a mechanism by which ERG recruits p300 to genomic loci for H3K27 acetylation. However, inhibition of ERG expression in HUVEC did not perturb activity in most super-enhancer regions. This is not surprising since super-enhancers are characterized by the presence of multiple TF binding sites and a high degree of enrichment of transcriptional co-activators, providing opportunities for cooperative binding and synergistic gene activation^{18,19}. Transcriptional networks consisting of members of the ETS (including ERG's closest homologue FLI1), AP-1 and GATA families have been shown to bind endothelial enhancers^{16,41}. Furthermore, ERG and AP-1 have been shown to functionally interact at composite DNA binding sites in non-EC⁴². As suggested from studies on AP-1 ⁴³, endothelial enhancer selection may be facilitated by cooperative ERG-AP-1 binding where ERG is acting as the endothelial-specific transcription factor. A recent study investigated the effect of combined knockdown of ERG and its closest homologue FLI1 on global H3K27ac levels in HUVEC and found a significant loss of H3K27ac on key endothelial genes⁴¹, supporting the notion that multiple TFs are directing cooperative transcriptional regulation. In our study, we found that the majority of endothelial super-enhancers are co-occupied by ERG, AP-1 members cFOS/cJUN, and GATA2. Interestingly, we found that the subset of super-enhancers not affected by ERG depletion is highly co-occupied by ERG, AP-1 and GATA2, whilst lower levels of all TF are present at the subset sensitive to ERG depletion. We propose that a cooperative TF network is able to compensate for ERG depletion at most super-enhancers; however, a specific subset of core super-enhancers is strictly dependent on ERG function, highlighting its key role in regulating endothelial gene expression.

ERG may modulate super-enhancer activity through mechanisms other than p300 recruitment. A potential mechanism may be through targeting the activity of the BAF (BRG1-associated factors) chromatin remodeling complex which disrupt histone-DNA interactions to control access to DNA⁴⁴. Recent studies have indicated a role for both AP-1 and ERG in binding to BAF subunits to establish accessible chromatin for enhancer selection and target gene regulation^{43,45}. The mechanism through which ERG modulates MED1 recruitment remains unclear. Mediator complex has been implicated in long-range chromatin

interactions functionally combining enhancers from many kilobases away. In fact, super-enhancers have been shown to form higher-order 3D chromatin structures which are likely to coordinate their activity in an orchestrated manner^{38,46}. In this study, we followed the current standard convention when annotating super-enhancer-associated regions with ERG-bound loci, namely by their linear distance along the epigenome²⁸. This methodology does not take into account the complex 3D chromatin structure. Further studies will be required to map super-enhancers using long-range chromatin interactions in the HUVEC genome.

Aberrant expression of lineage-specific TF in other tissues cause deregulated activation of a transcriptome profile detrimental to the cell¹⁵; this is indeed the case with ERG, which acts as an oncogene when overexpressed in cells such as the prostate epithelium¹³. We show that ERG-associated super-enhancer profiles are markedly different in HUVEC compared to VCaP cells. Thus, ERG does not co-opt an endothelial genomic profile in VCaP cells but controls fundamentally different pathways in these two cell types, partly through selective super-enhancer binding. These findings provide some insight into the molecular basis for ERG's homeostatic versus oncogenic functions. We postulate that different ERG-dependent gene expression between HUVEC and VCaP cells may be regulated in part by the activity of cell-specific pioneer factors which act to prime chromatin for accessibility at at lineage-specific sites⁴⁷.

Super-enhancer regions are commonly enriched in cell type-specific disease-related SNPs^{18, 38}. Hogan *et al.*¹⁶ identified disease trait-associated SNPs for CAD and hypertension within aortic endothelial enhancers. Our analysis of ERG-bound super-enhancers revealed enrichment for SNPs associated with diseases that have a vascular component, including predisposition to cardiovascular diseases, such as atherosclerosis and CAD. These data supports the notion that active maintenance of endothelial homeostasis through transcriptional programs is essential protection against a number of diseases, most of all CVD. Interestingly, recent GWAS meta-analysis revealed a novel risk locus for abdominal aortic aneurysm within the ERG gene itself⁴⁸. Further studies will determine the functional role of non-coding variants associated with ERG enhancers, and will provide crucial insight into the contribution of ERG, cooperative TF and co-factor binding in complex disease susceptibility.

In conclusion, this study provides novel evidence on the transcriptional and epigenetic mechanisms which controls lineage-specific gene expression in EC and identifies a possible functional link between regulation of ERG activity and human disease. These associations will provide valuable insights for investigating the role of ERG-dependent regulatory programs in maintaining endothelial homeostasis and protecting against vascular diseases.

ACKNOWLEDGEMENTS

We thank Dr Joan Ponsà (Imperial College London, UK) for helpful discussions.

SOURCES OF FUNDING

This work was funded by grants from the British Heart Foundation (RG/11/17/29256; RG/17/4/32662; FS/15/65/32036; PG/17/33/32990) and Cancer Research UK.

DISCLOSURES

None.

REFERENCES

1. Dejana E, Hirschi KK, Simons M. The molecular basis of endothelial cell plasticity. *Nat Commun.* 2017;8:14361
2. 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
3. Vlaeminck-Guillem V, Carrere S, Dewitte F, Stehelin D, Desbiens X, Duterque-Coquillaud M. The ets family member erg gene is expressed in mesodermal tissues and neural crests at fundamental steps during mouse embryogenesis. *Mech. Dev.* 2000;91:331-335
4. Ginsberg M, James D, Ding BS, Nolan D, Geng F, Butler JM, Schachterle W, Pulijaal VR, Mathew S, Chasen ST, Xiang J, Rosenwaks Z, Shido K, Elemento O, Rabbany SY, Rafii S. Efficient direct reprogramming of mature amniotic cells into endothelial cells by ets factors and tgfbeta suppression. *Cell.* 2012;151:559-575
5. Liu F, Patient R. Genome-wide analysis of the zebrafish ets family identifies three genes required for hemangioblast differentiation or angiogenesis. *Circ. Res.* 2008;103:1147-1154
6. Birdsey GM, Shah AV, Dufton N, Reynolds LE, Osuna Almagro L, Yang Y, Aspalter IM, Khan ST, Mason JC, Dejana E, Gottgens B, HodiVala-Dilke K, Gerhardt H, Adams RH, Randi AM. The endothelial transcription factor erg promotes vascular stability and growth through wnt/beta-catenin signaling. *Dev Cell.* 2015;32:82-96
7. Vijayaraj P, Le BA, Mitchell N, Kondo M, Juliao S, Wasserman M, Beeler D, Spokes K, Aird WC, Baldwin HS, Oettgen P. Erg is a crucial regulator of endocardial-mesenchymal transformation during cardiac valve morphogenesis. *Development.* 2012;139:3973-3985
8. Shah AV, Birdsey GM, Peghaire C, Pitulescu ME, Dufton N, Yang Y, Weinberg I, Osuna Almagro L, Payne L, Mason JC, Gerhardt H, Adams RH, Randi AM. The endothelial transcription factor erg mediates angiopoietin-1-dependent control of notch signalling and vascular stability. *Nat Commun.* 2017;8: 16002
9. Yuan L, Nikolova-Krsteovski V, Zhan Y, Kondo M, Bhasin M, Varghese L, Yano K, Carman CV, Aird WC, Oettgen P. Antiinflammatory effects of the ets factor erg in endothelial cells are mediated through transcriptional repression of the interleukin-8 gene. *Circ. Res.* 2009;104:1049-1057
10. Sperone A, Dryden NH, Birdsey GM, Madden L, Johns M, Evans PC, Mason JC, Haskard DO, Boyle JJ, Paleolog EM, Randi AM. The transcription factor erg inhibits vascular inflammation by repressing nf-kappab activation and proinflammatory gene expression in endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* 2011;31:142-150
11. Dufton NP, Peghaire CR, Osuna-Almagro L, Raimondi C, Kalna V, Chuahan A, Webb G, Yang Y, Birdsey GM, Lalor P, Mason JC, Adams DH, Randi AM. Dynamic regulation of canonical tgfbeta signalling by endothelial transcription factor erg protects from liver fibrogenesis. *Nat Commun.* 2017;8:895
12. Hagglof C, Hammarsten P, Stromvall K, Egevad L, Josefsson A, Stattin P, Granfors T, Bergh A. Tmprss2-erg expression predicts prostate cancer survival and associates with stromal biomarkers. *PLoS. One.* 2014;9:e86824
13. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. Recurrent fusion of tmprss2 and ets transcription factor genes in prostate cancer. *Science.* 2005;310:644-648
14. Adamo P, Ladomery MR. The oncogene ERG: A key factor in prostate cancer. *Oncogene.* 2016;35(4):403-14
15. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* 2015;16:144-154
16. Hogan NT, Whalen MB, Stolze LK, Hadeli NK, Lam MT, Springstead JR, Glass CK, Romanoski CE. Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. *Elife.* 2017;6

17. Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, Glass CK. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell*. 2013;51:310-325
18. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934-947
19. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307-319
20. Kaikkonen MU, Niskanen H, Romanoski CE, Kansanen E, Kivela AM, Laitalainen J, Heinz S, Benner C, Glass CK, Yla-Herttuala S. Control of vegf-a transcriptional programs by pausing and genomic compartmentalization. *Nucleic Acids Res*. 2014;42:12570-12584
21. Brown JD, Lin CY, Duan Q, Griffin G, Federation AJ, Paranal RM, Bair S, Newton G, Lichtman AH, Kung AL, Yang T, Wang H, Lusinskas FW, Croce KJ, Bradner JE, Plutzky J. Nf-kappab directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol Cell*. 2014;56:219-231
22. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyaev T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190-1195
23. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, Yan J, Talukder S, Turunen M, Taipale M, Stunnenberg HG, Ukkonen E, Hughes TR, Bulyk ML, Taipale J. Genome-wide analysis of ets-family DNA-binding in vitro and in vivo. *EMBO J*. 2010;29:2147-2160
24. Birdsey GM, Dryden NH, Shah AV, Hannah R, Hall MD, Haskard DO, Parsons M, Mason JC, Zvelebil M, Gottgens B, Ridley AJ, Randi AM. The transcription factor erg regulates expression of histone deacetylase 6 and multiple pathways involved in endothelial cell migration and angiogenesis. *Blood*. 2012;119:894-903
25. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74
26. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. *Cell*. 2008;132:311-322
27. Birdsey GM, Dryden NH, Amsellem V, Gebhardt F, Sahnun K, Haskard DO, Dejana E, Mason JC, Randi AM. Transcription factor erg regulates angiogenesis and endothelial apoptosis through v-cadherin. *Blood*. 2008;111:3498-3506
28. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013;153:320-334
29. Chng KR, Chang CW, Tan SK, Yang C, Hong SZ, Sng NY, Cheung E. A transcriptional repressor co-regulatory network governing androgen response in prostate cancers. *EMBO J*. 2012;31:2810-2823
30. Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y, Cheng H, Laxman B, Vellaichamy A, Shankar S, Li Y, Dhanasekaran SM, Morey R, Barrette T, Lonigro RJ, Tomlins SA, Varambally S, Qin ZS, Chinnaiyan AM. An integrated network of androgen receptor, polycomb, and tmprss2-erg gene fusions in prostate cancer progression. *Cancer Cell*. 2010;17:443-454
31. Wang S, Kollipara RK, Srivastava N, Li R, Ravindranathan P, Hernandez E, Freeman E, Humphries CG, Kapur P, Lotan Y, Fazli L, Gleave ME, Plymate SR, Raj GV, Hsieh JT, Kittler R. Ablation of the oncogenic transcription factor erg by deubiquitinase inhibition in prostate cancer. *Proc Natl Acad Sci U S A*. 2014;111:4251-4256

32. Asangani IA, Dommeti VL, Wang X, Malik R, Cieslik M, Yang R, Escara-Wilke J, Wilder-Romans K, Dhanireddy S, Engelke C, Iyer MK, Jing X, Wu YM, Cao X, Qin ZS, Wang S, Feng FY, Chinnaiyan AM. Therapeutic targeting of bet bromodomain proteins in castration-resistant prostate cancer. *Nature*. 2014;510:278-282
33. Kanki Y, Nakaki R, Shimamura T, Matsunaga T, Yamamizu K, Katayama S, Suehiro JI, Osawa T, Aburatani H, Kodama T, Wada Y, Yamashita JK, Minami T. Dynamically and epigenetically coordinated gata/ets/sox transcription factor expression is indispensable for endothelial cell differentiation. *Nucleic Acids Res*. 2017
34. Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, Cejas P, Vazquez F, Cook J, Shivdasani RA, Bowden M, Lis R, Hahn WC, Kantoff PW, Brown M, Loda M, Long HW, Freedman ML. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat Genet*. 2015;47:1346-1351
35. Fish JE, Gutierrez MC, Dang LT, Khyzha N, Chen Z, Veitch S, Cheng HS, Khor M, Antounians L, Njock MS, Boudreau E, Herman AM, Rhyner AM, Ruiz OE, Eisenhoffer GT, Medina-Rivera A, Wilson MD, Wythe JD. Dynamic regulation of vegf-inducible genes by an erk-erg-p300 transcriptional network. *Development*. 2017;144(13):2428-2444
36. De Val S, Black BL. Transcriptional control of endothelial cell development. *Dev Cell*. 2009;16:180-195
37. Kim BK, Im JY, Han G, Lee WJ, Won KJ, Chung KS, Lee K, Ban HS, Song K, Won M. P300 cooperates with c-jun and parp-1 at the p300 binding site to activate rhob transcription in nsc126188-mediated apoptosis. *Biochim Biophys Acta*. 2014;1839:364-373
38. Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Moran I, Gomez-Marin C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, Garcia-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Gomez-Skarmeta JL, Muller F, McCarthy MI, Ferrer J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet*. 2014;46:136-143
39. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The ncbi dbgap database of genotypes and phenotypes. *Nat Genet*. 2007;39:1181-1186
40. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, Burdett T, Hindorff L, Flicek P, Cunningham F, Parkinson H. The new nhgri-ebi catalog of published genome-wide association studies (gwas catalog). *Nucleic Acids Res*. 2017;45:D896-D901
41. Nagai N, Ohguchi H, Nakaki R, Matsumura Y, Kanki Y, Sakai J, Aburatani H, Minami T. Downregulation of erg and fli1 expression in endothelial cells triggers endothelial-to-mesenchymal transition. *PLoS Genet*. 2018;14:e1007826
42. Madison BJ, Clark KA, Bhachech N, Hollenhorst PC, Graves BJ, Currie SL. Electrostatic repulsion causes anticooperative DNA binding between tumor suppressor ets transcription factors and junfos at composite DNA sites. *J Biol Chem*. 2018;293:18624-18635
43. Vierbuchen T, Ling E, Cowley CJ, Couch CH, Wang X, Harmin DA, Roberts CWM, Greenberg ME. Ap-1 transcription factors and the baf complex mediate signal-dependent enhancer selection. *Mol Cell*. 2017;68:1067-1082 e1012
44. Becker PB, Horz W. Atp-dependent nucleosome remodeling. *Annu Rev Biochem*. 2002;71:247-273
45. Sandoval GJ, Pulice JL, Pakula H, Schenone M, Takeda DY, Pop M, Boulay G, Williamson KE, McBride MJ, Pan J, St Pierre R, Hartman E, Garraway LA, Carr SA, Rivera MN, Li Z, Ronco L, Hahn WC, Kadoch C. Binding of tmprss2-erg to baf chromatin remodeling complexes mediates prostate oncogenesis. *Mol Cell*. 2018;71:554-566 e557
46. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Rolando DMY, Javierre BM, Atla G, Farabella I, Morgan CC, García-Hurtado J, Beucher A, Morán I, Pasquali L, Ramos M, Appel EVR, Linneberg A, Gjesing AP, Witte DR, Pedersen O, Garup N, Ravassard P,

- Torrents D, Mercader JM, Piemonti L, Berney T, de Koning EJP, Kerr-Conte J, Pattou F, Fedko IO, Prokopenko I, Hansen T, Marti-Renom MA, Fraser P, Ferrer J. Human pancreatic islet 3d chromatin architecture provides insights into the genetics of type 2 diabetes. *bioRxiv*. 2018. DOI: <http://dx.doi.org/10.1101/400291>.
47. Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev*. 2016;37:76-81
48. Jones GT, Tromp G, Kuivaniemi H, *et al*. Meta-analysis of genome-wide association studies for abdominal aortic aneurysm identifies four new disease-specific risk loci. *Circ Res*. 2017;120:341-353.



Circulation Research

ONLINE FIRST

FIGURE LEGENDS

Figure 1. Characterization of ERG-bound enhancers in HUVEC. (A) Percentage genomic distribution of ERG ChIP-seq peaks. (B) Gene ontology analysis showing pathways associated with differentially regulated ERG-bound endothelial target genes (genes activated in blue, genes repressed in red). Significance shown as $-\log(p\text{-value})$. (C) ChIP-seq binding profiles for ERG, H3K4me1, H3K27ac, H3K4me3 and DNase I hypersensitivity in HUVEC; loci of ERG-activated *CDH5* (top) and ERG-repressed *ICAM1* (bottom). The x-axis represents the genomic position; the y-axis the ChIP-seq signal in reads per million per base pair (rpm/bp). ERG binding sites are shown as black bars. (D) Number of ERG binding sites associated with specific epigenomic features versus size-matched random regions. Active enhancers are defined by combined H3K4me1 and H3K27ac, poised enhancers by H3K4me1 and H3K27me3, and active promoters by H3K4me3 and H3K27ac. (E) Luciferase reporter plasmids (pGL4) containing the *CDH5* promoter with or without region E4 or a mutant enhancer (E4mut), were co-transfected into HUVEC along with an ERG cDNA expression plasmid (pcDNA-ERG) or empty vector control, pcDNA3.1. Values are represented as the fold change in relative luciferase activity over the empty vector alone. Values are mean \pm SEM, $n = 3$. A two-way ANOVA showed significance ($P < 0.001$) and a post hoc test using Tukey multiple comparisons test shows pairwise differences between specific groups, $***P < 0.001$.

Figure 2. Characterization of HUVEC super-enhancers identifies enrichment for ERG and ERG-target genes. (A) HUVEC enhancer regions identified by H3K27ac and H3K4me1 were ranked by enrichment of H3K27ac ChIP-seq signal in rpm/bp. Super-enhancer (SE) clusters are shown to the right of the gray dashed line. (See also Online Table V) (B) HUVEC SE regions indicated as red bars above density plots of ERG ChIP-seq signal alongside H3K27ac at distal enhancers of *VWF*, *ICAM2*, and *ERG* loci. (C) Transcriptome profiling of HUVEC shows significantly higher gene expression in SE than in typical enhancers. $***P < 0.0001$, Wilcoxon rank-sum test. (D) GSEA of the top 500 SE-associated genes compared with the ranked gene list from transcriptome profiling of ERG-deficient HUVEC. Normalised enrichment score (NES) = 2.13, $P < 0.001$.

Figure 3. ERG defines SE in HUVEC. (A) ERG binding at SE and typical enhancer regions. SE have significantly more ERG bound than typical enhancers; $***P < 0.0001$, Fisher's exact test. (B) Scatterplot of the correlation between ERG and H3K27ac occupancy (\log_2 transformed) in the 917 SE regions. Each point is the mean ChIP-seq signal across each SE region; $P < 0.001$. (C) SE ranked by ERG enrichment on identified enhancer regions. Known ERG-target genes proximal to SE are indicated. (See also Online Table VI) (D) GSEA of the top 500 genes associated with ERG-identified SE compared with the ranked gene list from 917 H3K27ac-enriched SE. NES = 1.54, $P < 0.001$. (E) Gene ontology analysis was carried out on the genes associated with SE identified by enrichment for H3K27ac or ERG. The top 7 biological processes ranked by significance ($-\log(p\text{-value})$) are depicted.

Figure 4. ChIP-seq analysis identifies a lineage-specific program for ERG: ERG expression and binding profile in prostate cancer cells. (A) Gene ontology pathway analysis of differentially regulated ERG-bound target genes in VCaP prostate cancer cells (activated genes in blue, repressed genes in red). Significance of pathway enrichment as $-\log(p\text{-value})$. (B) Venn diagram comparing ERG bound, differentially regulated genes in HUVEC and VCaP cells. Only 249 genes (in bold) are common to the two cell types. None of the fractions overlap significantly between the 2-way comparisons using a hypergeometric distribution test. (C) Heatmap of expression levels of the shared 249 bound and regulated putative ERG target genes following ERG inhibition in HUVEC and in VCaP cells. (D) Genomic view of percentage distribution of ERG peaks relative to TSS in regions bound by ERG that are shared between HUVEC and VCaP cells, HUVEC only or VCaP cells only. (E) Overlap of ERG binding sites at H3K4me3-enriched and RefSeq assigned TSS promoter regions (left), and H3K27ac/H3K4me1-enriched enhancers (right), in HUVEC and VCaP cells. Significance reported by a hypergeometric test p-value.

Figure 5. ERG associates with lineage-specific super-enhancers. (A) Identification of VCaP SE by H3K27ac ChIP-Seq signal in rpm/bp. (See also Online Table VII) (B) ERG binding at SE and typical enhancers in VCaP cells. SE have significantly more ERG binding than typical enhancers; *** $P < 0.0001$, Fisher's exact test. (C) Gene ontology analysis on the SE-associated genes in HUVEC (red) and VCaP cells (blue). The top 5 biological processes in both cells are ranked by significance ($-\log(p\text{-value})$). (D) ChIP-seq binding profiles for ERG and H3K27ac occupancy in HUVEC (left) and VCaP cells (right). H3K27ac-defined SE regions are depicted as blue bars below tracks. (E) Aggregate plots of active histone modifications H3K27ac and H3K4me1 histone modifications from HUVEC and VCaP cells in HUVEC SE (top panels) and VCaP SE (bottom panels). Plots centered on the SE center. The average size of VCaP SE are smaller than those identified in HUVEC. (F) Motif analysis at ERG binding sites ($\pm 200\text{bp}$) in HUVEC and VCaP. Top 5 most enriched transcription factor families are shown with the most highly occurring member of each family represented. Motifs for the binding of critical lineage transcription factors coincide with the ERG motif in both HUVEC and VCaP cells.

Figure 6. ERG contributes to enhancer activation in HUVEC. (A) Volcano plot showing \log_2 fold change (FC) vs $-\log(p\text{-value})$ of differential H3K27ac enrichment at siCtl H3K27ac regions in response to ERG knockdown. Loss and gain enhancer regions are selected by $-1 \geq \log_2\text{FC} \geq 1$; $-\log(p\text{-value}) \geq 4$. (B) Heatmap of H3K27ac enrichment over input in all loss and gain regions. Signal ± 5 kb from the centre of H3K27ac siCtl or siERG regions as tag density/bp. (C) ChIP-qPCR of H3K27ac at enhancers of *CLDN5*, *DLLA* and *VWF* with a negative control gene desert, in siCtl vs siERG treated HUVEC. Graph represents fold change over IgG, $n=3$. * $P < 0.05$, paired two-tailed t-test. (D) Boxplot representing $\log_2\text{FC}$ from transcriptome profiling data following ERG knockdown in H3K27ac regions changed in response to siERG (as defined in A). $P < 0.0001$, Kruskal-Wallis test and post-hoc test using Wilcoxon rank-sum test **** $P < 0.0001$; * $P < 0.05$. (E) Pathway analysis of genes associated with loss and gain of H3K27ac. (F) Boxplot showing the $\log_2\text{FC}$ of MED1 occupancy in siERG-treated HUVEC in H3K27ac regions changed in response to siERG. $P < 0.0001$, Kruskal-Wallis test and post-hoc test using Wilcoxon rank-sum test, **** $P < 0.0001$.

Figure 7. ERG regulates distinct endothelial super-enhancers. (A) SE identification on siCtl H3K27ac using H3K27ac enrichment from siCtl and siERG-treated HUVEC. Selected genes relevant to endothelial function are indicated in red. (See also Online Table VIII) (B) Volcano plot depicting the $\log_2\text{FC}$ vs $-\log(p\text{-value})$ of differential super-enhancers at 1015 H3K27ac SE identified in siCtl treated cells. Significantly up or downregulated super-enhancers are selected according to $-0.58 \geq \log_2\text{FC} \geq 0.58$ ($-1.5 \geq \text{FC} \geq 1.5$); $-\log(p\text{-value}) \geq 4$. (C) Boxplot representing the $\log_2\text{FC}$ of MED1 occupancy in siERG-treated HUVEC in *decreased SE* and *constant SE*. **** $P < 0.0001$, Wilcoxon rank-sum test. (D) ChIP-qPCR of p300 enrichment in siCtl or siERG-treated HUVEC at selected SE constituent enhancers (E) associated with *decreased SE* (*CLDN5*, *DLLA*) or *constant SE* (*IL6*, *PXN*). Data are represented as fold change over IgG, $n=4$. * $P < 0.05$, paired two-tailed t-test. (E) Genomic occupancy of ERG and collaborative TFs: GATA2, cFOS and cJUN at *decreased SE* compared to *constant SE*, measured by the ChIP-seq signal in HUVEC. **** $P < 0.0001$; *** $P < 0.001$; * $P < 0.05$, Wilcoxon rank-sum test. (F) Model of ERG-dependent SE assembly in EC. Cooperative SE activation is associated with a strong transcriptional complex (ERG, AP-1, GATA2) at *constant SE* (left). In a subset of SE, activation of SE is strongly dependent on ERG due to less abundance of transcription factor network partners with reduced cooperativity.

Figure 8. ERG super-enhancers prioritize cardiovascular disease variants. (A) Overlap of GWAS SNPs associated with disease trait classes within ERG defined SE (orange) with chromosome and size matched random SE (gray). Significance of enrichment was calculated by binomial distribution test with red dashed line indicating $P = 0.05$. Some points overlay. (B) Test for enrichment of selected CVD trait-associated SNPs within ERG defined SE (orange) with chromosome and size matched random SE (gray), as performed in A. Random region were restricted to open chromatin by excluding placement in repressed chromatin states in HUVEC.

NOVELTY AND SIGNIFICANCE

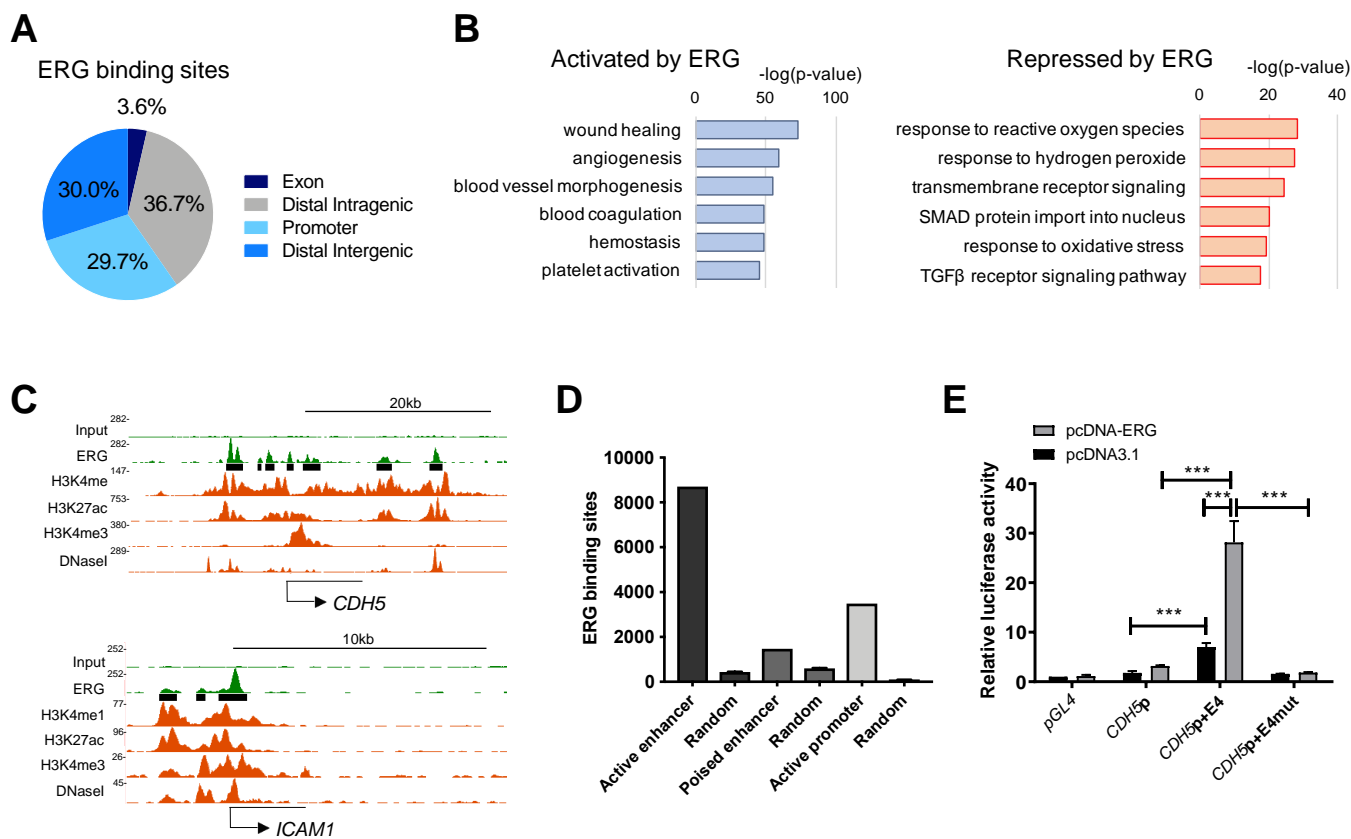
What Is Known?

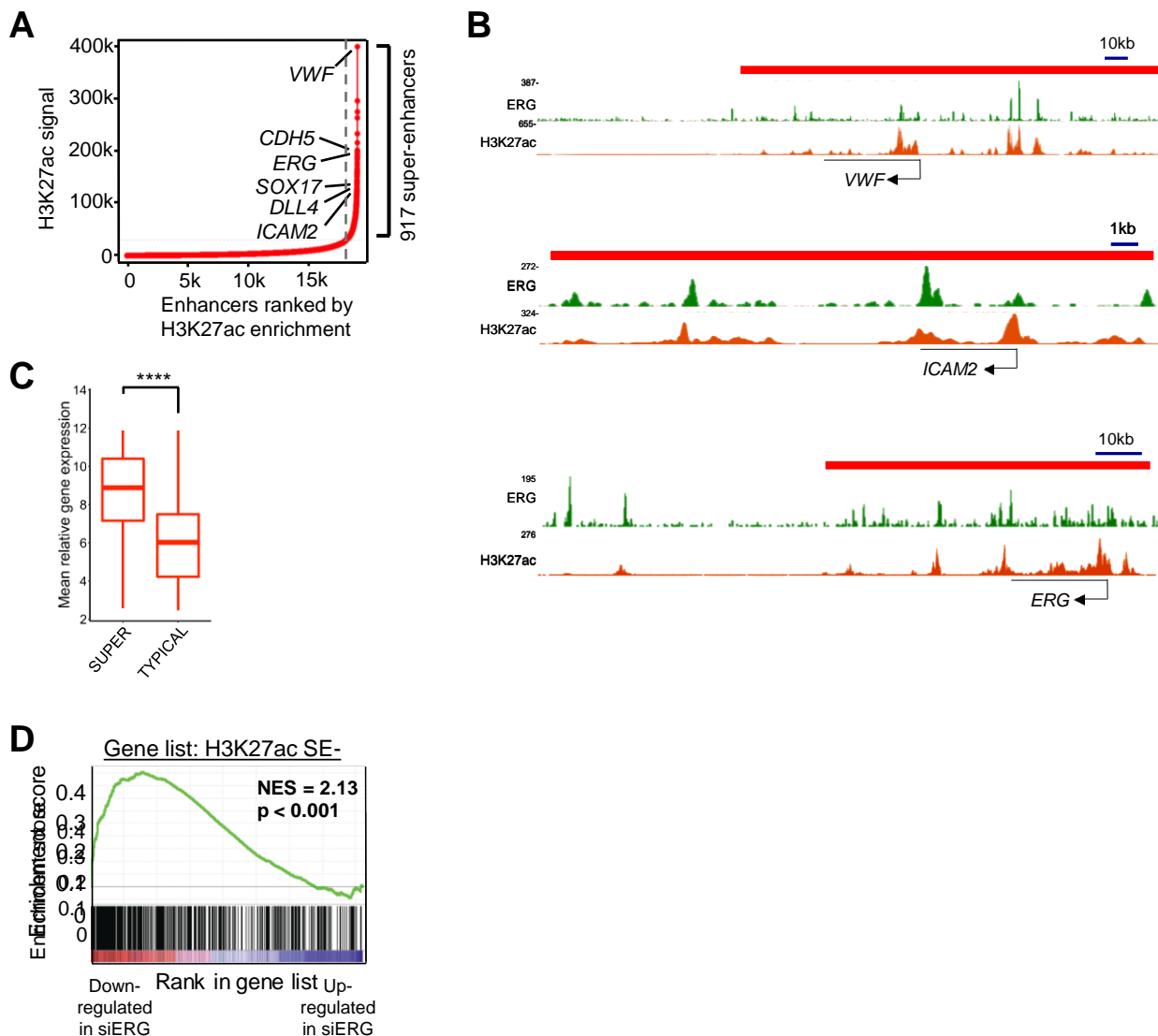
- The ETS transcription factor ERG is essential for endothelial homeostasis, whereas its aberrant over-expression in cancer is oncogenic.
- Lineage-specific transcription factors bind to chromatin regulatory elements, called super-enhancers, which are enriched in single nucleotide polymorphisms (SNPs) associated with specific diseases, including CVD.
- Epigenomic mechanisms are emerging as key players in cardiovascular disease (CVD); little is known about the epigenetic regulation of endothelial function and its links to CVD.

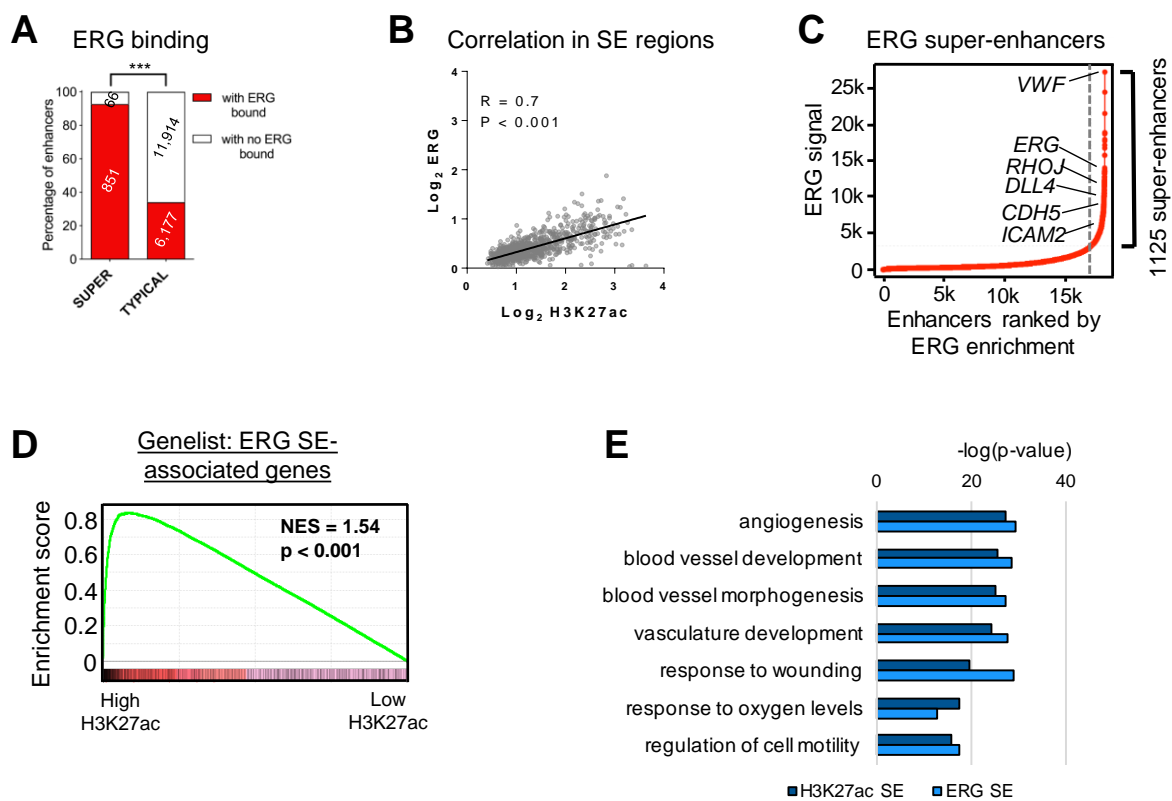
What New Information Does This Article Contribute?

- The transcription factor ERG drives endothelial lineage genes via super-enhancers.
- Comparison of human umbilical vein endothelial cells (HUVEC) and prostate cancer.
- TMPRSS2:ERG fusion-positive VCaP cells reveals distinctive lineage-specific transcriptome and super-enhancer profiles.
- ERG regulates a core set of endothelial super-enhancers that have reduced transcription factor cooperativity.
- CVD-associated SNPs are enriched at ERG super-enhancers.

There is an emerging link between the transcriptional and epigenetic regulation of endothelial gene expression and CVD. Most disease-associated SNPs lie within noncoding regions of the genome and perturb transcription factor recognition sequences. A key question is whether disruption of transcription factor pathways which promote endothelial cell homeostasis modulates disease risk. This is suggested by the presence of SNPs in endothelial cell enhancers linked to CVD. In this study, we focus on the transcription factor ERG which is required for endothelial lineage specification, vascular development and angiogenesis, and plays an essential role in maintaining vascular homeostasis. Profiling global ERG DNA binding reveals that ERG binds to and regulates endothelial super-enhancers. This is lineage-specific since oncogenic ERG in prostate cancer VCaP cells binds different super-enhancers compared to HUVEC. ERG binding at endothelial super-enhancers is associated with CVD-associated SNPs, suggesting that perturbation of ERG DNA-binding motifs in super-enhancers may modulate disease risk. In summary, we identify a novel mechanism through which the ERG transcription factor promotes endothelial cell homeostasis via regulation of super-enhancers. Binding of ERG to super-enhancers may have functional consequences for CVD risk. These findings may open new avenues of research focussing on the epigenomic mechanisms underlying CVD pathologies.







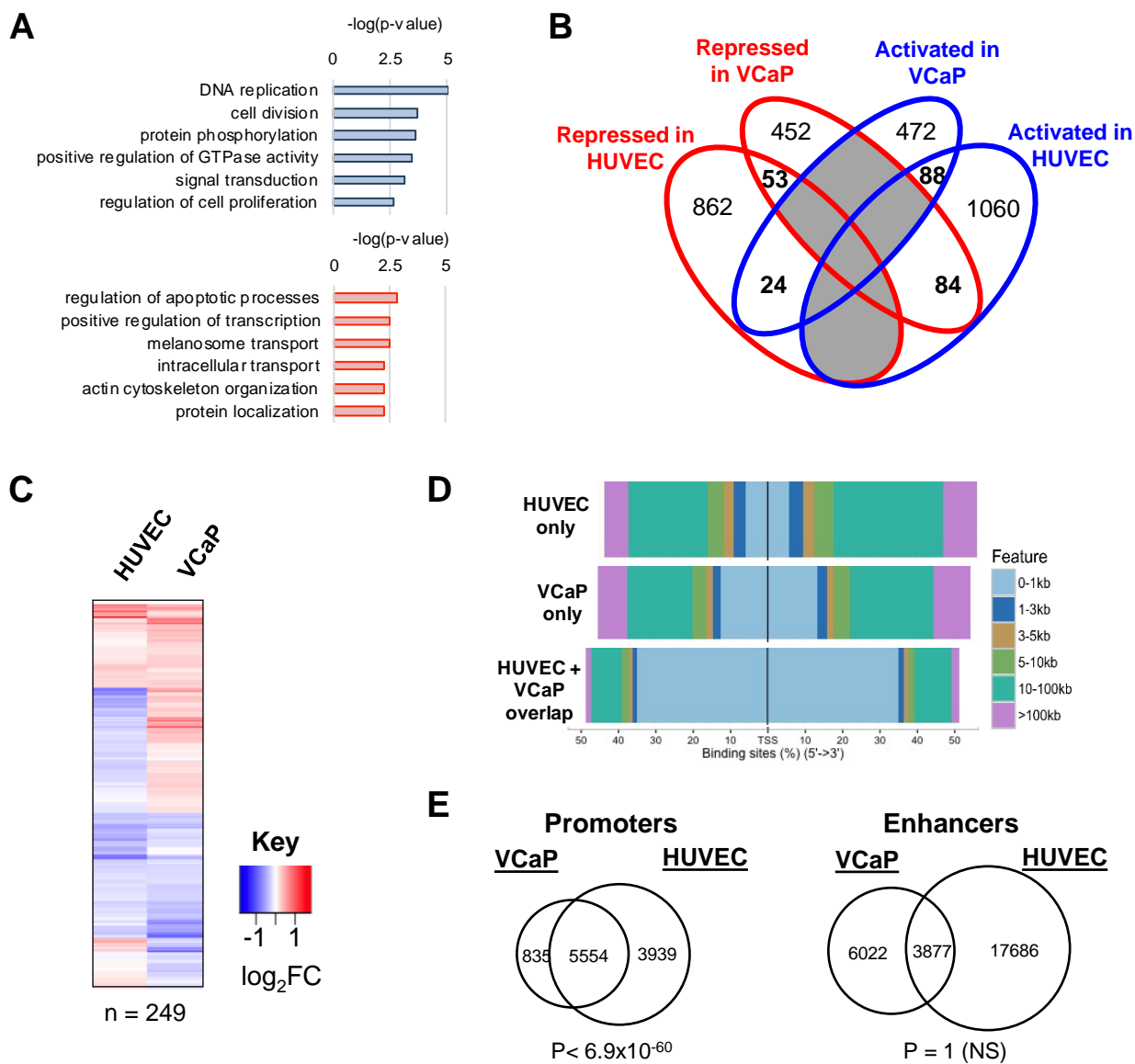
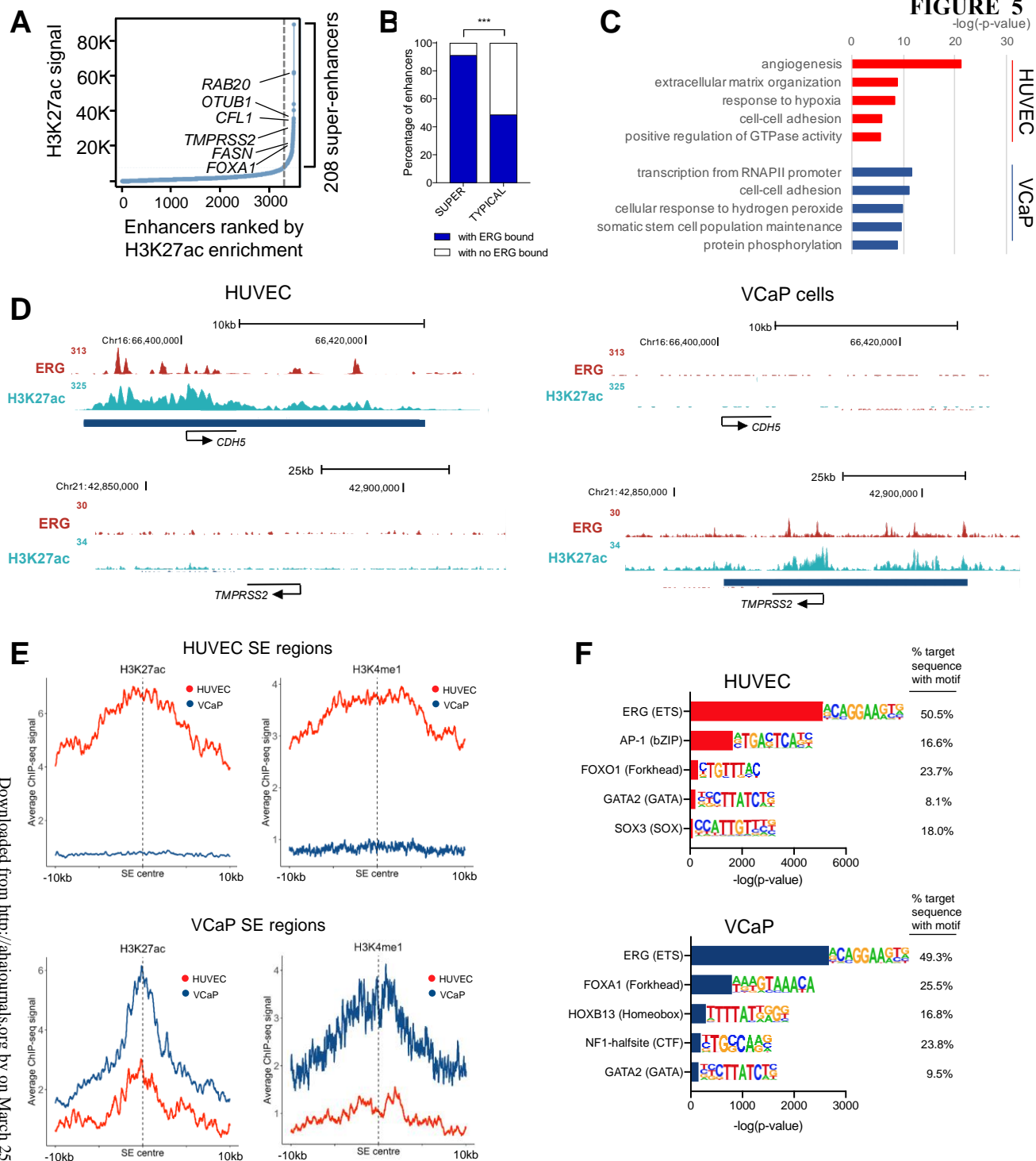


FIGURE 5



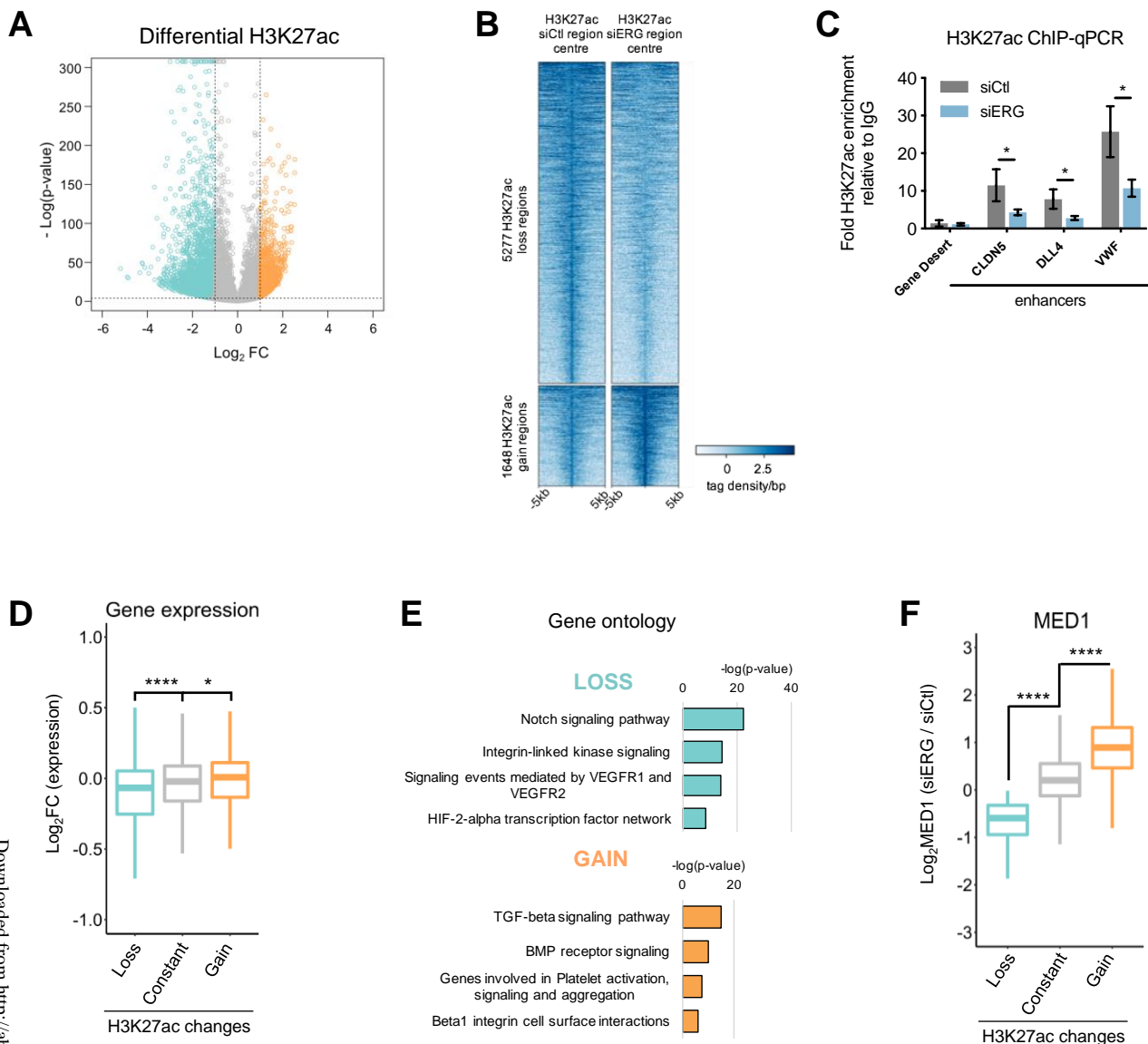
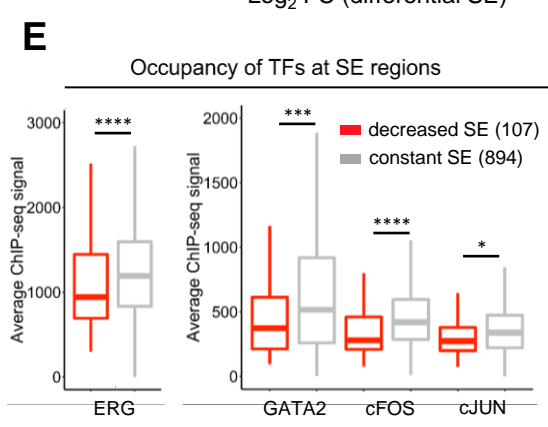
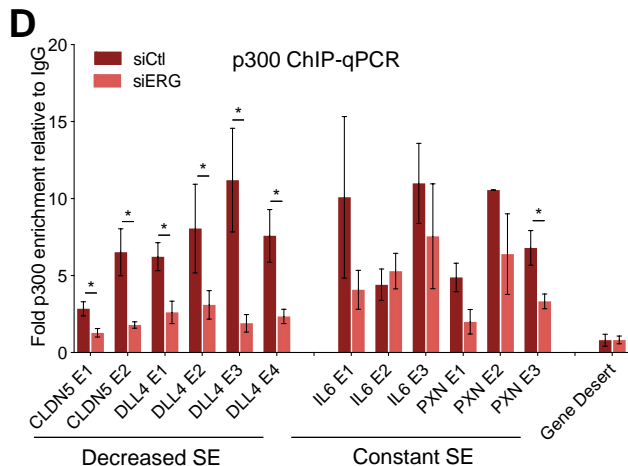
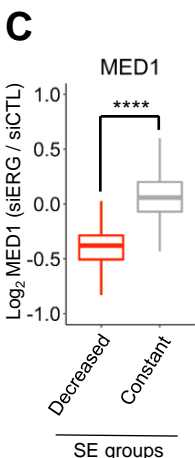
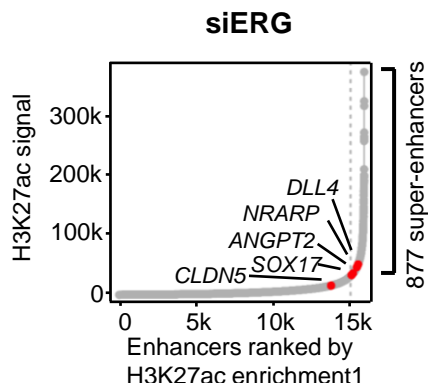
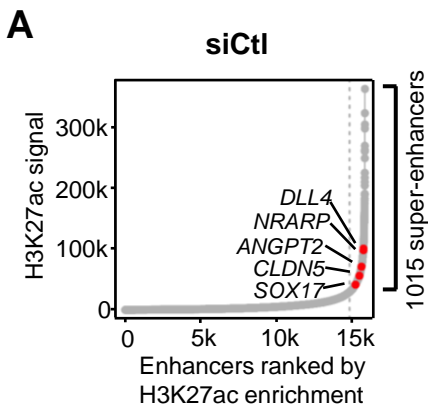
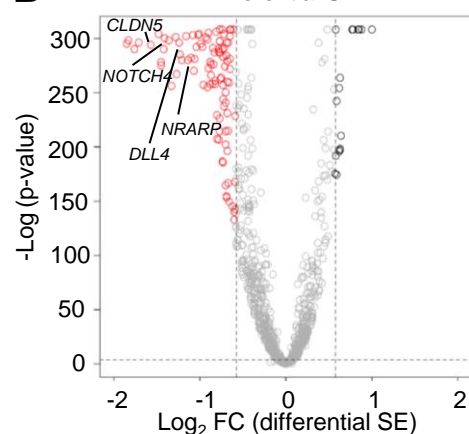
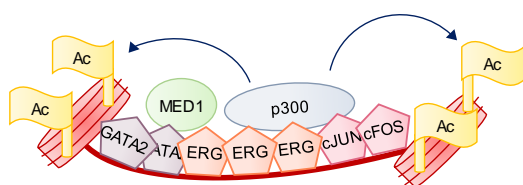


FIGURE 7

Differential SE

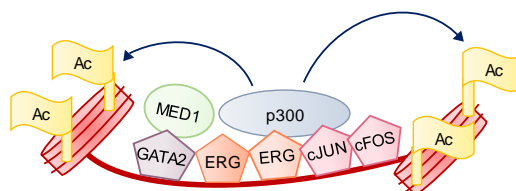


F Cooperative super-enhancer activation (*IL6*, *PXN*)



- Stronger transcriptional complex
- Cooperative binding between TFs

ERG-dependent super-enhancer activation (*DLL4*, *CLDN5*)



- Weaker transcriptional complex
- Low cooperativity between TF partners

