Imperial College London
National Lung and Heart Institute
Vascular Sciences Section

Transcriptional and epigenetic regulation of lineage identity in endothelial cells by the transcription factor ERG via super-enhancers

---

Viktoria Kalna

November 2019

A thesis submitted to Imperial College London for the degree of

Doctor of Philosophy

DECLARATION

I declare that the work presented in this thesis is that of my own, and that any work of others has been appropriately acknowledged and cited in the text.

Viktoria Kalna

ABSTRACT

Transcriptional programs establish and maintain cell identities. Regulation of gene expression is mediated by sequence-specific transcription factors (TFs) and *cis*-regulatory elements present in the genome. More recently, cell identity has been associated with lineage-defining super-enhancers: comprising dense TF platforms. Endothelial cells are key players of vascular integrity. The ETS TF ERG is constitutively expressed in endothelial cells and essential for endothelial lineage specification, vascular homeostasis and angiogenesis. However, the genomic programs that are regulated by ERG in endothelial cells are poorly understood. In this thesis, I show that ERG densely occupies super-enhancers in human umbilical vein endothelial cells (HUVEC), and its high occupancy can identify super-enhancers. I find that variants associated with cardiovascular disease are enriched in ERG-defined super-enhancers providing insight to the endothelial contribution to complex disease. Depletion of ERG causes profound modulation of the active enhancer mark H3K27ac genome-wide and in recruitment of the transcriptional co-activator Mediator complex. Loss of ERG leads to a decrease in 107 endothelial super-enhancers that have reduced co-occupancy of TFs GATA2 and AP-1. This indicates that ERG plays an essential role as a positive regulator of a core set of endothelial super-enhancers. Interestingly, aberrant ERG overexpression in prostate cancer via androgen-responsive TMPRSS2:ERG fusion proteins is oncogenic. Comparison between HUVEC and prostate cancer TMPRSS2:ERG fusion-positive VCaP cells revealed distinct lineage-specific transcriptome and super-enhancer profiles. In endothelial cells, I show that ERG is required at promoters and enhancers yet assembles distinct TF complexes at these two regions. ERG also colocalises with structural chromatin regulator CTCF in HUVEC, implying a role for ERG in coordinating chromatin structural organisation. Finally, I adopt CRISPR-Cas9 gene editing technology to genetically dissect the super-enhancer of adhesion molecular VE-cadherin. The mechanistic exploration of ERG and endothelial super-enhancers provide insight into the regulation of the endothelial-specific gene expression program.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my principal supervisor Professor Anna Randi for the opportunity to do this PhD and explore the computational and experimental avenues of vascular biology. Thank you for your continued enthusiasm in this work and allowing me to be involved in many other projects. Your dedication to scientific integrity has taught me the rigour required to be a successful research scientist. I am further grateful for your support, guidance and helpful discussion over the last four years.

I am also grateful to my second supervisor, Dr Graeme Birdsey, for his supervision and knowledge of all things 'ERG'. I am thankful to have benefited from your guidance in the lab. An extended thank you goes to your patience during the publication of our paper.

I'm grateful to the British Heart Foundation for my funding without which this work wouldn't exist. It also wouldn't exist without the National Heart and Lung Institute which have provided research infrastructure in the Vascular Sciences department.

A special thank you goes to all members of the group, past and present. It has been a pleasure to work with such talented and friendly company. I especially want to thank Claire Peghaire for her guidance and contribution to my project, and helping me out in all aspects of the lab. Your expertise is second to none and it has been extremely rewarding for me to work with you. I further thank Neil 'Mr Grumpy' Dufton for all your guidance, blue-sky thinking, humour and encouragement over the years; Gaye Saginc for the critical assessment of all my omics data analysis, enthusiasm and crazy fun nature; Claudio Raimondi for knowledge and help in the lab; Astrid Stroobants for your great friendship, fun and support in my final years; Josefin Edqvist for the many fun discussions in and out the lab; Koralia Paschalaki for all your positive encouragement; Koval Smith and Luke Payne for guiding me in the unknown in my first years and Youwen Yang for

beginning this exciting project.

I further give appreciation to the collaboration with Professor Jorge Ferrer and his team. I thank Jorge, Inês Cebola and Joan Ponsa-Cobas for their help and guidance in the many areas of super-enhancers, CRISPR gene editing and genomic variants.

I extend my thanks to my colleagues on the 5th floor of ICTEM, Dr Joseph Boyle and Dr Mike Johns for technical expertise and Samata Pandey for your friendship, support and stress release.

The greatest thank you goes to my own Joseph for your endless support and understanding. You have reassured and encouraged me throughout and kept me on track in the tough times. To my family, my Dad for his positive outlook and my brother for his support. Last and most importantly, to my Mother, my greatest encouragement, I was inspired by you to pursue a PhD in biomedical research and I am continually guided by your light. Sadly you could not see me finish my PhD, but Mum, I did it.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Ang-1 Angiopoietin-1

AP-1 Activating protein-1

ATAC-seq Assay for transposase accessible chromatin

bp base pair

BRD4 Bromodomain-containing protein 4

CAD Coronary artery disease

CAGE Cap analysis of gene expression

CBP CREB-binding protein

ChIP-seq Chromatin immunoprecipitation followed by high-throughput sequencing

CRE *Cis*-regulatory element

CRISPR clustered regularly interspaced short palindromic repeats

CTCF CCCTC-binding factor

Ctl Control

DLL4 Delta like canonical notch ligand 4

DNA Deoxyribonucleic acid

DNase I deoxyribonuclease I

E Enhancer

EC Endothelial cells

ECFC Endothelial colony forming cells

ENCODE Encyclopedia of DNA elements

EndMT Endothelial-to-mesenchymal transition

ERG ETS related gene

eRNA enhancer RNA

ESC Embryonic stem cells

ETS E-twenty-six

FC Fold change

FDR False discovery rate

FLI1 Friend of leukaemia integration-site 1

FPKM Fragments per kilobase per million mapped reads

GATA2 GATA-binding factor 2

gDNA genomic DNA

GO Gene ontology

GREAT Genomic Regions Enrichment of Annotations Tool

gRNA guide RNA

GRO-seq Global run-on sequencing

GSEA Gene set enrichment analysis

GTEx Genotype-Tissue Expression

GWAS Genome-wide association studies

H3K27ac Histone 3 lysine 27 acetylation

H3K27me3 Histone 3 lysine 27 trimethylation

H3K4me1 Histone 3 lysine 4 monomethylation

H3K4me3 Histone 3 lysine 4 trimethylation

HAEC Human aortic endothelial cells

HAT Histone acetyltransferase

HDAC Histone deacetyltransferase

Hi-C High throughput chromosome conformation capture

HSCs Haematopoietic stem cells

HUVEC Human umbilical vein endothelial cells

kb kilobase pair

KRAB Krüppel-associated box

LD Linkage disequilibrium

MAF Minor allele frequency

MAPK Mitogen- activated protein kinase

MED1 Mediator complex subunit 1

MI Myocardial infarction

NOTCH4 Notch Receptor 4

PCR Polymerase chain reaction

RNA Ribonucleic acid

RNA-seq RNA-sequencing

ROSE Rank ordering of super-enhancers

rpm reads per million

RT-qPCR Real-time quantitative polymerase chain reaction

sc single-cell

SNP Single nucleotide polymorphism

SPOP Speckle-type POZ protein

TAD Topologically associated domain

TF Transcription factor

TMPRSS2 Transmembrane protease, serine-2

TSS Transcription start site

USP9X ubiquitin-specific peptidase 9, X-linked

VCaP Vertebral-cancer of the prostate

VEGF Vascular endothelial growth factor

VEGFR Vascular endothelial growth factor receptor

1 INTRODUCTION

1.1 Regulation of gene expression

The majority of the cells in the human body share the same genome. Yet, many different cell types exist and are phenotypically and functionally distinct. The gene expression programs that are established during differentiation are maintained in differentiated cells and these define the cell identity. These gene expression programs are controlled by changes at the level of the chromatin. This epigenetic information is harboured in the likes of histone modifications and chromatin structural conformation. Assays to elucidate these chromatin features specify the position of cis-regulatory elements which regulate the gene expression program of a given cell type.

1.1.1 Cis-regulatory elements (CREs)

The genomic DNA is partitioned into functional units for the spatial and temporal control of gene expression. Gene expression initiates from the core promoter, also regarded as the transcription start site (TSS), which engages RNA polymerase II (Pol II) for transcription initiation. At this level, transcription is generally low and requires enhancer elements distant to the TSS to facilitate maximal transcriptional output (Visel et al. 2009).

Cis-regulatory elements (CREs), predominantly promoters and enhancers, provide binding modules for one or more trans-acting transcription factors (TFs). These functional sites reside in nucleosome-depleted regions hypersensitive to deoxyribonuclease I (DNase I) cleavage. The precise mapping of these CREs in different cell types has provided a systematic strategy to infer the regulatory logic that coordinates gene regulation. In this section I describe the CREs central to lineage-specificity, the chromatin architecture and the trans-acting factors functioning in concert.

1.1.2 Histone modifications and chromatin state classification

The opening of chromatin is controlled by the modifications at histone tails which affect the relative spacing of nucleosomes through electrostatic interactions with the DNA. Histone acetylation neutralises the positively charged histone core of the nucleosome reducing DNA compaction and remodelling the local chromatin environment to expose TF motifs for binding (Zentner et al. 2011). Histone acetylation is a dynamic process regulated by histone acetyltransferases that covalently attach acetyl groups to lysine side chains and histone deacetylases that cause histone deacetylation. On the contrary, histone methylation does not change the electric charge of histone proteins preserving the histone-DNA structure. The effects of histone methylation on gene regulation are more diverse; methylation of H3K27 and H3K9 causes transcriptional repression and in contrast methylation of H3K4 and H3K36 is linked to transcriptional activation (Zentner et al. 2011).

Distal enhancer regions act to coordinate with promoters to regulate gene expression. They are found to be enriched at monomethylated and dimethylated H3K4 (H3K4me1 and H3K4me2) and conversely lack the trimethylated H3K4 (H3K4me3) found instead to be associated with promoter regions (Heintzman et al. 2007). To distinguish active genomic regions from inactive regions histone modification acetylated H3K27 (H3K27ac) is a powerful predictor of activation at enhancers and promoters. In relation, histone acetyltransferse p300, which deposits the acetyl modification onto H3K27, is also distributed to active regions and in particular functional enhancers. P300 is highly related to the CREB-binding protein (CBP) and both are shown to be involved in multiple cell types and is a functionally diverse coactivator which interacts with over 400 TFs and regulatory proteins (Bedford et al. 2010).

Despite these convenient enhancer and promoter specifications, all three H3K4 (me1/ 2/ 3) histone modifications have been detected at enhancers along with histone variant H3.3 associated with active transcription and H2A.Z known for limiting DNA accessi-

bility (Barski et al. 2007). This highlights the promiscuity of histone modifications at different enhancers in the genome.

Regardless of the reported complexity, the changes in chromatin state at enhancers changes dynamically with the functional state observed most frequently at enhancers being either active or poised (Battle et al. 2019). Histone H3K27ac/H3K4me1 co-occupancy distinguishes active enhancers whilst poised enhancers are elements with H3K4me1 only (Creyghton et al. 2010). Additionally, enhancers marked with H3K4me1 and H3K27me3 are actively repressed and these enhancers are known to mediate transcriptional repression at genes whose expression is inappropriate for the cell type (Rada-Iglesias et al. 2011). Importantly, dysregulation of these epigenetic modifications have been implicated in the development and maintenance of disease states (Heintzman et al. 2007).

1.1.3 Regulation by enhancers

The first enhancer element deciphered was a 72 bp sequence of the viral SV40 genome that enhanced transcription of a gene reporter 100-fold in HeLa cells (Banerji et al. 1981). Enhancers are accessible elements as a result of chromatin and nucleosomal remodelling and are present upstream or downstream of their target genes and also in intronic regions (Heintzman et al. 2007). Multiple enhancers can control the same target promoter and conversely one enhancer can interact with more than one promoter. Mechanistically, enhancers are capable of crossing over the most proximal gene to regulate a more distant one (Andersson et al. 2014; Sanyal et al. 2012). In addition to position and distance independence, enhancers also act independent of orientation. Enhancers possess multiple binding sites for transcription factors (TFs) which recruit the likes of p300, Mediator complex and chromatin remodelling complexes and these embellishments allow them to create promoter-enhancer interactions and loop out intervening chromatin (Nolis et al. 2009).

The docking of TFs and subsequent transcriptional machinery at enhancers is undisputed; however, three key models have been proposed for the functionality of enhancers: (1) the enhancesome model, (2) the billboard model and (3) TF collective model.

The enhancesome model specifies that the TFs are recruited to enhancers by specific and predefined TF motifs that promote highly cooperative DNA binding (Spitz et al. 2012). In this model transcriptional output is achieved only when all the TF are integrated (Long et al. 2016; Spitz et al. 2012).

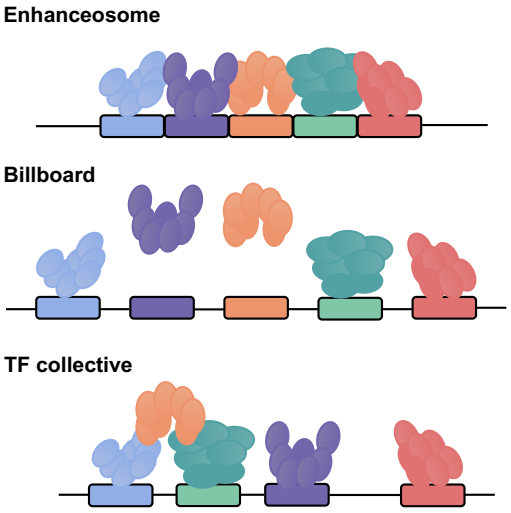
The billboard model does not strictly specify TF binding position or organisation but considers the presence of particular TFs. The model emphasizes the need for direct TF-DNA binding but only requires some of the TFs to be present to be active.

Finally, the TF collective model exemplifies that there is no restriction on motif presence or position with the possibility of TFs binding the DNA directly or indirectly through protein-protein interactions (Long et al. 2016; Spitz et al. 2012).

Overall, it is difficult to categorise all enhancers into one of these models and in fact most enhancers will function as a combination of the rigid motif organisation of the enhancesome model, the more flexible motif organisation of the billboard model and the indirect TF binding of the TF collective model (Long et al. 2016). The models are collectively summarised in Figure 1.1.

Alternatively, more recent work investigating the organisation of TF motifs within enhancer elements alternatively suggested a different model whereby TFs play a distinct role in the function of a given enhancer determined by their given position in that enhancer element (Grossman et al. 2018).

Enhancers have sparked particular interest as they tend to largely control lineage specific gene expression programs (Heintzman et al. 2007; Thurman et al. 2012; Visel et al. 2009). Thousands of enhancers have been identified to be active in any given cell type (ENCODE Project Consortium 2012; Thurman et al. 2012) and these establish a specific



Motifs	Binding assembly	Practicality of model
• Sites present for every factor • Rigid motif positioning	• Cooperative and direct binding • Factor positions are fixed	• Not common due to the need for high levels of sequence conservation
• Flexible motif sequence • Rigid motif positioning	• Cooperative and direct binding • Additive binding	• Common as not all factors need to be bound for activation
• Flexible motif sequence • Flexible motif sequence	• Cooperative binding • Direct DNA binding not necessary	• Common as evolutionary conservation is lowest of all models (feature of enhancers)

Figure 1.1: The three key models of enhancer activity.

chromatin signature for specifically modulating target gene expression without affecting other pleiotropic genes (Long et al. 2016). Additionally, enhancers show a rapid rate in genomic evolution as opposed to promoters which are more highly conserved regulatory regions (Villar et al. 2015). Indeed, analysis across the genomes of 20 mammalian species, it was discovered that approximately half of all active enhancers (20,000 to 25,000) have evolved in a lineage- or species-specific manner to drive phenotypic adaptations (Villar et al. 2015).

1.1.4 Super-enhancers

Gene regulation by enhancers is often not constrained to a single enhancer regulating a single gene. Often a large cluster of multiple enhancers, termed super-enhancers, are implicated in gene expression control to attain high levels of transcription in a cell-type specific manner (Hnisz et al. 2013; Lovén et al. 2013; Whyte et al. 2013). Super-enhancers represent a small fraction of the total enhancers in the genome but have several features that distinguish them from isolated typical enhancers. These include dense occupancy

of master regulators such as master TF and chromatin activators Mediator complex and bromodomain-containing protein 4 (BRD4). Mediator complex is a co-activator that relays signals from enhancer-bound TFs to RNA Pol II at core promoters by coordinate looping (Yin et al. 2014). BRD4 is a member of the BET family proteins which recognise acetylated histones, associate with Mediator complex and interact with RNA Pol II for transcriptional regulation (Hnisz et al. 2013).

Additionally, super-enhancers show elevated H3K27ac levels providing a domain for synergistic and higher-order chromatin assembly (Hnisz et al. 2013; Lovén et al. 2013; Whyte et al. 2013). This high-density assembly has been proposed to be achieved by phase-separation of the transcriptional co-activators in the nucleus that form condensates through weak protein-protein interactions (Sabari et al. 2018). The plethora of factors concentrated at super-enhancers has allowed their identification firstly in embryonic stem cells (ESCs) (Hnisz et al. 2013; Whyte et al. 2013) and many other cell-types (Brown et al. 2014; Huang et al. 2016; Pinz et al. 2016; Vahedi et al. 2015). Identification uses an algorithm termed Rank Ordering of Super-Enhancers (ROSE) (Lovén et al. 2013; Whyte et al. 2013) which requires chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) binding signatures of active enhancer marks such as H3K27ac, Mediator complex subunit 1 (MED1) and lineage-defining TFs (Whyte et al. 2013). Enhancer elements within 12.5 kb are concatenated and then ranked based on ChIP-seq signal which are classified as super-enhancers beyond an identified inflection point (Lovén et al. 2013; Whyte et al. 2013).

The emergence of the super-enhancer concept coincided with the identification of so-called “stretch enhancers” (Parker et al. 2013) described as extend enhancers which span more than 3 kb. Super and stretch enhancers have often been considered as discrepant for their association with lineage identity genes however super-enhancers have been characterised as being more cell-type specific and transcriptionally active (Khan et al. 2018).

Studies dissecting the structure of super-enhancers *in vitro* have reported that individual constituent enhancers are both essential and dispensable. One study determined that a subset of super-enhancers have a hierarchical structure whereby one constituent enhancer acts as the hub enhancer that attributes most to gene expression and chromatin organisation (Huang et al. 2018). Shin et al. (2016) performed mutational analysis to decipher the contribution of three enhancers of the mammary gland super-enhancer gene *Wap* using small deletions or mutations *in vivo*. Disruption of all three enhancers completely abrogated *Wap* expression and a hierarchical structure was demonstrated with the most distal enhancer regulating more than 90% of *Wap* expression (Shin et al. 2016). A schematic to illustrate the composition of super-enhancer structure is shown in Figure 1.2.

However, some studies have not supported the super-enhancer paradigm as exceptional regulatory elements of gene expression. Assessment of a super-enhancer associated with α-globin in erythroid cells showed that constituent enhancers of the cluster act independently and additively to gene expression and similarly to chromatin structure in the form of chromatin accessibility and chromatin interactions (Hay et al. 2016). In another study, isolated enhancers of embryonic stem cells were found to control robust transcriptional regulation and enhancers in a super-enhancer cluster were identified to act partially in a redundant manner to fine tune gene expression (Moorthy et al. 2017).

Thus, it is likely that constituent enhancers function in a manner dependent on cell type and environmental condition in super-enhancer composition.

1.1.5 DNA sequence variations at enhancer regions and disease risk

DNA sequence variations which are disease-linked are most frequently located at enhancers in the genome. Single nucleotide polymorphisms (SNPs) are the most prevalent variation in the genome and their position at enhancers has made it a challenge to link them to their target gene or genes. Genome wide association studies (GWAS) have

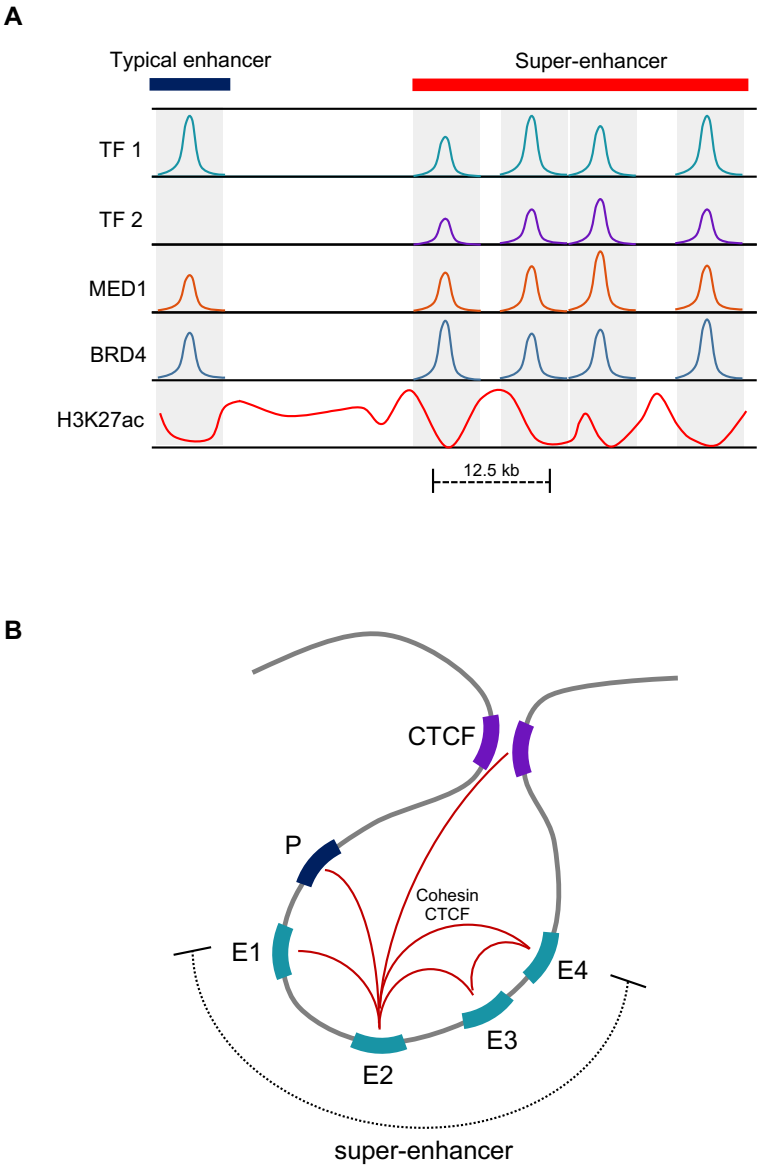


Figure 1.2: Super-enhancer structure and composition. (A) Linear structure of a super-enhancer and typical enhancer with the respective markers for identification which include cell-type specific TFs, co-activators (such as MED1 and BRD4) and active histone modifications (predominantly H3K27ac). Multiple, closely spaced (within 12.5 kb) enhancer sites are concatenated together to and ranked by accumulative ChIP-seq signal to define super-enhancers of several kilobases. Typical enhancers are isolated from the cluster of enhancers and are typically less than 1 kb in size. (B) Schematic of the composition of a super-enhancer in three-dimension whereby each constituent enhancer has the propensity to interact to its neighbouring enhancers and the gene promoter via CTCF/cohesin (red lines). The diagram specifically represents a super-enhancer that is organised in a hierarchical structure with a hub enhancer, E2, which is highly enriched in CTCF/cohesin binding and arranges the other non-hub enhancers to form an integrated structure.

classified a number of disease-associated SNPs including those in linkage disequilibrium (LD). More than 150 loci have been identified to be associated with cardiovascular diseases including coronary artery disease (CAD) and myocardial infarction (MI). Many studies have painstakingly characterised enhancer-occupying SNPs by identifying the target gene and the functional causal mechanisms. A non-coding SNP associated with five vascular diseases was linked to an enhancer region of the aortic tissue and found to be regulating endothelin-1 expression in EC (Gupta et al. 2017). Another recent report identified a SNP associated with CAD and ischaemic stroke to reside in an endothelial enhancer of the *PLPP3* gene. Functional validation by CRISPR interference experiments of the SNP locus resulted in a decrease in *PLPP3* expression in HAEC (Krause et al. 2018).

Importantly, super-enhancers have been found to be enriched for disease or trait-associated SNPs linked to the cell type in which they were defined. Hnisz et al. (2013) associated non-coding SNPs linked with the likes of Alzheimer's disease, type I diabetes and systemic lupus erythematosus to super-enhancers of brain tissue, pancreatic β cells and B cells respectively.

1.1.6 3D organisation of the genome

Higher-order chromatin structure is vital to gene regulatory mechanisms. Long-range interactions from distant enhancers contact promoters by looping out intervening DNA. The genome is organised into domains with high chromatin interactions and regions with fewer, dispersed interactions collectively referred to as topologically associated domains (TADs). These TADs are largely invariant between different cell types (Dixon et al. 2012). Given the TAD structure, genes are commonly co-regulated in the same TAD with the possibility that enhancers in one TAD may sample all promoters in that TAD. However, they do not seem to represent functional domains that have segregated genes that share similar functions. One exception to this is the presence of an active and

repressed domain for regulation of the *Hoxa* cluster (Kim et al. 2011).

These domains of *Hoxa* and all other TADs are separated by CCCTC-binding factor (CTCF) binding that further docks cohesin complex for the formation of long-range chromatin loops. Depletion of CTCF does not vastly affect the TAD organisation but does show an increase in interdomain interactions (Zuin et al. 2014), suggesting that CTCF has a functional role as an insulator.

1.2 Transcriptional regulation by trans-acting factors

The genetic information of enhancers is read by a large range of trans-acting protein factors which include DNA-binding sequence-specific TFs and recruited transcriptional co-regulators. In lineage-determination it is reported that only a small subset of TF, known to regulate each other, bind together in the genome (Boyer et al. 2005; Chen et al. 2008). The classical example is the conversion of fibroblasts into induced pluripotent stem cells (iPSCs) using lineage reprogramming via expression of the "Yamanaka" factors Oct4, Sox2, c-Myc and Klf4 (Okita et al. 2007; Takahashi et al. 2006; Wernig et al. 2007).

There exists around 1,600 TFs and it is known that the same TF can regulate different genes in different cell types. Although, TF assembly can be cell type-specific yet experimentally elucidating these complexes is challenging. TFs are classified as activators or repressors of gene expression and can commonly conduct both. Interestingly, a high-throughput study using 15,000 putative regulatory elements distinguished TF binding sequences with an activating role, repressor functions or both in the same cell type (Ernst et al. 2016).

In relation, transcriptional co-regulators implement chromatin modifications by acetylating and deacetylating histones without direct DNA binding. They further facilitate communication between the TFs and RNA Pol II for target complex assembly and gene regulation. Here I describe the function of TFs (commonly cell-type specific) and co-

regulators (expressed more ubiquitously) on the chromatin landscape and how these coordinate together to regulate transcription.

1.2.1 Transcription factor DNA binding

The establishment of terminally differentiated cells is initiated from cell-type specific enhancers that are predominantly inaccessible in early development. The action of nucleosome-binding TFs, termed pioneer TFs, is to precisely remodel these nucleosome occupied sites by nucleosome eviction or destabilisation to increase the chromatin accessibility and subsequently promote the deposition of enhancer-associated histone marks (Heintzman et al. 2007; Zaret et al. 2011). However, other TFs do not have these properties yet they and transcriptional co-regulators can subsequently bind to the newly opened chromatin regions to direct cell-type specific transcriptional outputs for determining cell fate (Iwafuchi-Doi et al. 2014).

The structural constraints dictated by nucleosomes affect TF-DNA binding. Biochemical structural assays to understand the relationship between TFs and their binding to nucleosomal DNA have shown different positional preferences. Zhu et al. (2018) confirmed that most TFs are unable to bind to nucleosomes and that nucleosomes also affect the orientation of TF binding. Finally, it was also identified that some TFs can stabilise nucleosomes and these regions correlate with gene repression (Zhu et al. 2018). For TFs that bind to open chromatin there appears to be a positional distribution of their motifs that is closely related to distinct TF functions including their binding stability, cell-type specificity and interaction with other TFs (Grossman et al. 2018). Accordingly, it is known that the TF binding strength is correlated with levels of DNA accessibility and active histone modification H3K27ac (Grossman et al. 2017).

1.2.2 Combinatorial transcription factor regulation

The binding of multiple TFs that function in a combinatorial manner is imperative for cell-specific transcriptional control (Yamamoto 1985). As mentioned earlier, a set of TFs are required as opposed to only one TF for lineage reprogramming (Takahashi et al. 2006).

It has been reported that only a few TFs are required for the identity of any given cell-type and these function both as ubiquitous housekeeping TFs and specific cell-lineage associated TFs (Wei et al. 2017). Interestingly, it has been suggested that besides the action of pioneer TFs to open inaccessible chromatin, the combination of specific TFs can also remodel a subset of nucleosome-rich sites. The "Yamanaka" factors Oct4, Sox2 and Klf4 which selectively bind enhancers were identified to co-occupy sites of pronounced chromatin accessibility during stages of fibroblast to ESC reprogramming (Chronis et al. 2017).

The selection and priming of cell-type-specific enhancers is mediated by TFs in both a collaborative and hierarchical fashion. Firstly, enhancer selection is collaborative since an enhancer will frequently not be bound by a TF if its collaborating TF motif is mutated (Heinz et al. 2013) and it is hierarchical as enhancers already bound by TFs are more likely to be bound by new TFs (Heinz et al. 2013; Kaikkonen et al. 2013).

It is evident that TFs have different modes of combinatorial interactions on enhancers. The reported effects are that some TF pairs can (1) enhance each other's activity (superadditive), (2) function independently of each other (additive) and (3) mutually inhibit each other's activity (subadditive). A model of these three states is illustrated in Figure 1.3. The classification of these types is dependent on the spatial and structural features of a given TF. Interestingly, super-enhancers appear to be hotspots for cooperative TF binding (Siersbæk et al. 2014).

One underlying mechanism that promotes TF-TF interactions on DNA are the ex-

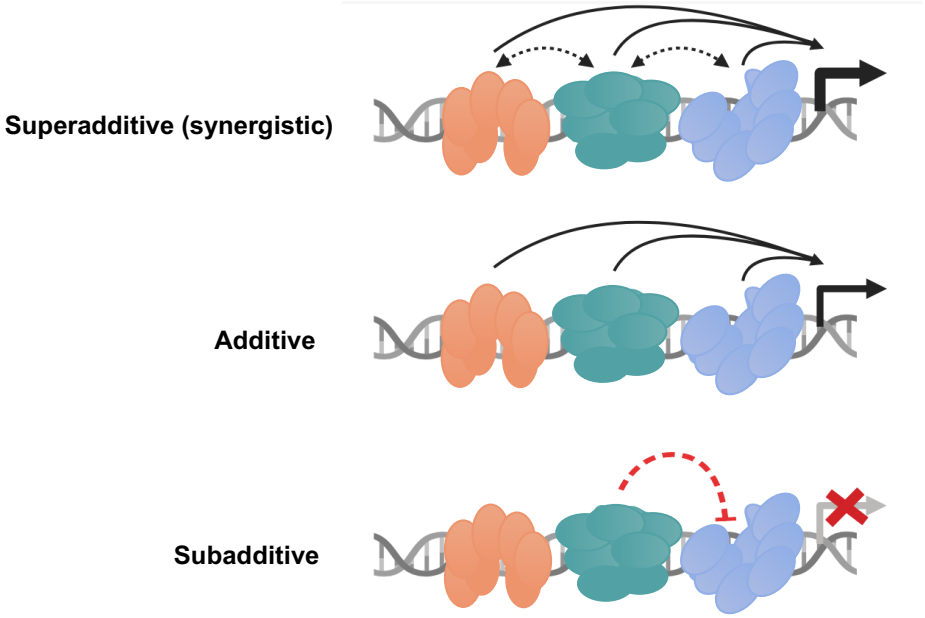


Figure 1.3: Modes of combinatorial TF interactions on enhancers. Superadditive transcriptional output occurs when co-occupancy of TFs is synergistic whereby they aid one another's activity. Alternatively, TFs can act additively in an independent manner and finally they can be subadditive by one or several TFs suppressing the active function of another, resulting in no transcriptional output.

istence of composite TF motifs in the genome. These are often enriched in enhancers of a specific cell-type to promote heterodimer protein binding for enhancer activation (Grossman et al. 2017).

1.2.3 Co-regulation factors

In line with the combinatorial binding between TFs, proteins associated with co-activating or co-repressing functions are further recruited to TF-bound DNA sites for facilitating transcriptional outputs. The majority fall into the categories of chromatin modifiers, chromatin remodellers and scaffolding proteins.

The chromatin modifiers act to first write post-translational modifications on histone tails (eg. p300/CBP histone acetyltransferase), chromatin erasers reverse these modifications (eg. histone deacetylases (HDACs)) then chromatin readers bind to modified proteins (eg. BRD4 and the BAF complex components, also referred to as mSWI/SNF (mammalian switching defective/sucrose non-fermenting)) (Gillette et al. 2015).

Chromatin remodellers are ATP-dependent enzymes that catalytically disrupt nucleosomes and alter their spacing. Alternatively other evidence has revealed that the complex BAF, in its role as a remodeller, evicts the antagonistic Polycomb repressive complexes (PRCs) to form open chromatin (Kadoch et al. 2017).

Finally, scaffolding proteins act to demarcate CREs for assembly of diverse regulatory complexes. One example is KAP1 (KRAB (Krüppel-associated-box)-associated protein 1) that assembles a repressive transcriptional complex by KRAB binding to KAP1 and functioning as a scaffolding protein to further recruit histone methyltransferases and deacetylases (Iyengar et al. 2011). A schematic to summarise these particular chromatin modifiers is shown in Figure 1.4.

The other large, multi-subunit complex is Mediator which functions as a key co-activator. As described, Mediator relays signals between RNA Pol II at promoters

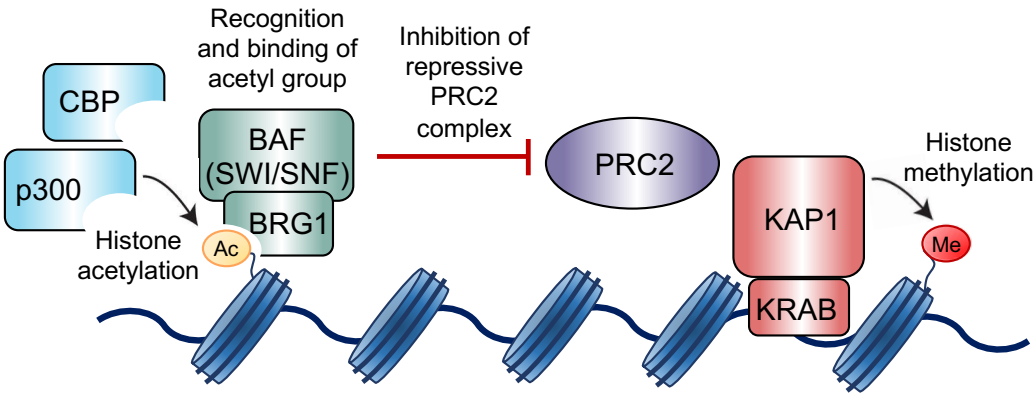


Figure 1.4: Transcriptional control mediated by co-factor activity. Chromatin modifier histone acetyltransferase p300 deposits an acetyl group on the tail of a histone protein including H3 Lys-27 (H3K27). The chromatin reader BAF complex (SWI/SNF) has sunbunit BRG1 (Brahma-related gene 1) that recognises and binds to the acetylated residue. The BAF complex is known to suppress the function of the repressive complex PRC2 that functionally establishes heterochromatin via methyltransferase activity. Epigenetic formation of heterochromatin formation is induced by the DNA-binding KRAB repressor domain which upon binding to DNA recruits KAP1 that acts as a scaffolding protein to engage other methyltransferases and deacetylases for chromatin heterochromatic silencing.

and TFs at enhancers but is further involved in almost all stages of gene transcription. Studies have shown it to be involved in transcriptional elongation, mRNA processing, DNA loop formation and most recently super-enhancer formation (Huang et al. 2012; Kagey et al. 2010; Takahashi et al. 2011; Whyte et al. 2013). In humans, Mediator is formed of 26 subunits and many have been shown to interact with a plethora of TFs, other co-regulators and long-non-coding RNA (lncRNA) but not directly to DNA itself (Yin et al. 2014).

1.2.4 Methods to study cis-regulatory elements in transcriptional regulation

Several techniques have been established to explicitly elucidate the function of CREs and global lower-order and higher-order chromatin structure. Through continuous advances these methods are achieving higher precision, specificity and resolution. Integration of multiple high-throughput methodologies have allowed the study of chromatin

structure and function for insightful genome-wide characterisation.

DNase I hypersensitive sites sequencing (DNase-seq) and transposase-accessible chromatin using sequencing (ATAC-seq)

To detect the open chromatin sites on a genome-wide scale, techniques that leverage exposed DNA sites have been employed. Both DNase I hypersensitive sites sequencing (DNase-seq) (Song et al. 2010) and transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro et al. 2013) use enzyme-based methods to target accessible DNA followed by high-throughput sequencing. These assays map the range of CREs in the genome, define nucleosome positioning and confer potential TF binding sites. Both techniques infer chromatin accessibility to a high degree but ATAC-seq has become popular for its feasibility with low cell numbers obtained from *in vivo* and patient samples (Buenrostro et al. 2013). Interestingly, it has been reported that the strongest hypersensitive sites and ATAC-seq signals are distributed almost exclusively to enhancer sites as opposed to promoter regions (Grossman et al. 2018).

Chromatin immunoprecipitation with deep sequencing (ChIP-seq)

The relationship between the genetic sequence and its occupying proteins has been widely investigated by chromatin immunoprecipitation followed by sequencing (ChIP-seq). ChIP can also be analysed in a locus-specific manner by performing a quantitative polymerase chain reaction (qPCR) on the enriched pool of target bound chromatin. Figure 1.5 provides a schematic overview of the ChIP-seq and ChIP-qPCR method.

ChIP-seq has most commonly been used for modified histones to deduce chromatin states and TF binding sites to map functional chromatin. The Encyclopedia of DNA Elements (ENCODE) project have thus far mapped more than 10,000 histone modifications and transcriptional regulators in human tissues and cells provide a rich resource for comparative and integration studies (ENCODE Project Consortium 2012).

Limitations of ChIP-seq include the fact that only one protein can be investigated

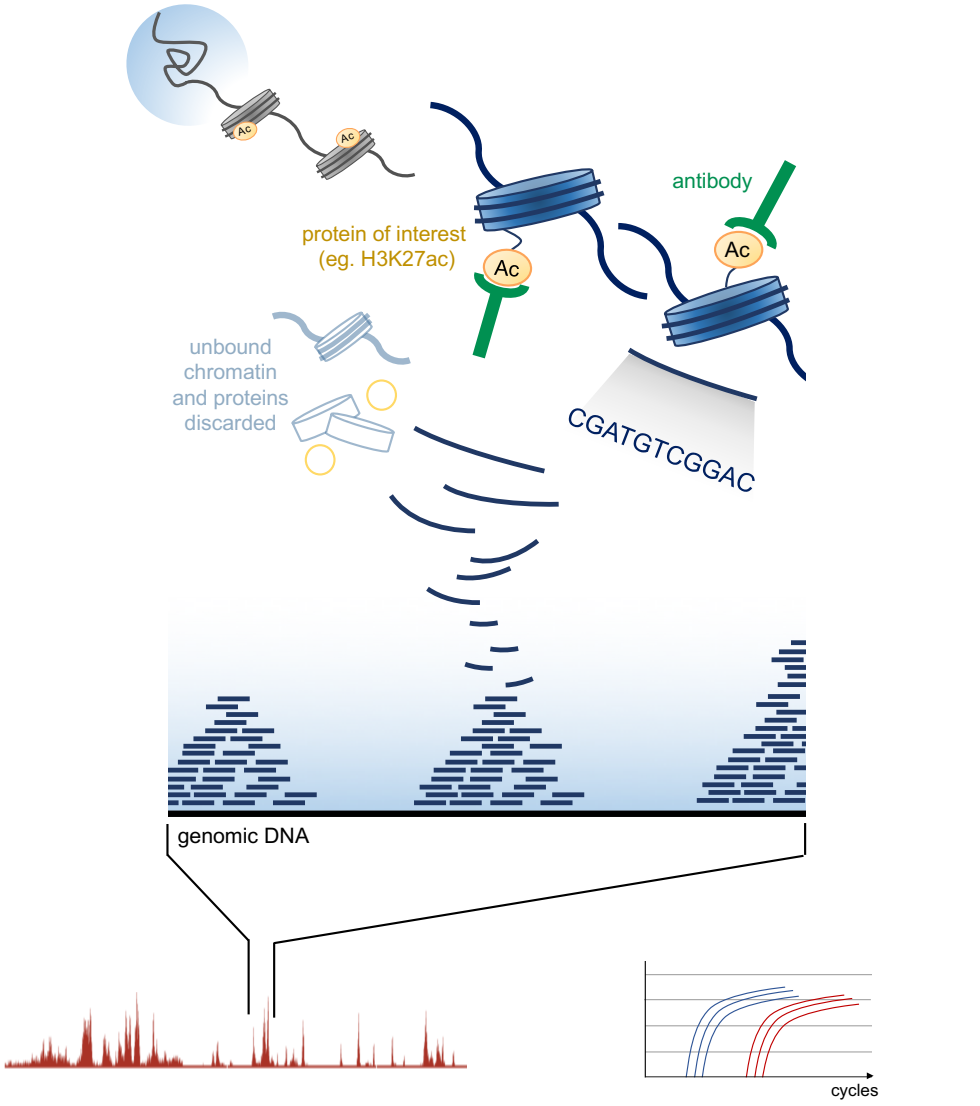


Figure 1.5: ChIP-seq is a powerful method to identify enriched chromatin binding of proteins. To assess the binding loci of proteins of interest in the chromatin a multi-step protocol is used. Firstly, the chromatin is cross-linked to establish DNA-protein complexes. These are sheared to fragments and a specific antibody is used to immunoprecipitate the specific protein, usually histone modifications and TFs. The cross-links are reversed and the originally bound DNA is quantified genome-wide through (1) high-throughput sequencing which is then mapped back to the genome or locus-specifically by (2) polymerase chain reaction.

at a time and common caveats in sample preparation relating to antibody efficacies, quantity of input sample and material loss during processing can confound high-quality data generation (Park 2009).

In combination with RNA-sequencing (RNA-seq) data and microarray transcriptome profiling novel insights into the structure and function of regulatory networks can be established within and across various cell types.

Assays to quantify transcriptional regulatory elements

DNase-seq, ATAC-seq and ChIP-seq are high-throughput genomic assays that are independent of regulatory activity. To overcome this, techniques have been developed to study nascent RNA which includes protein-coding mRNA, long non-coding RNA, microRNA and enhancer RNA (eRNA), all synthesised by RNA Pol II. Global and high-sensitivity technologies have emerged to quantify nascent RNA production. Start-seq is a single nucleotide RNA method for selection of short 5' ends of capped RNA and is particularly useful for identification of eRNA, hallmarks of active enhancers (Henriques et al. 2018). Another method that assays enhancer activity is the self-transcribing active regulatory region sequencing (STARR-seq) (Muerdter et al. 2015). This functional assay determines enhancer strength based on activity levels of candidate enhancer sequences. Finally, run-on techniques, the most common of which is global run-on sequencing (GRO-seq) (Lopes et al. 2017) marks nascent RNAs with labelled nucleotides and maps the docking sites of RNA Pol II in the genome. It has provided a technique to robustly cover the enhancer and promoter-associated RNA transcripts allowing all varieties of nascent RNA to be detected in the genome.

Chromatin conformation capture

Experimental approaches to identify enhancer contacts *in situ* has relied on genome-wide chromosome conformation capture (Hi-C) that determines the TAD structure. However, confidence in chromatin contacts are limited when using Hi-C due to the

restriction fragment length generated by the assay; the lowest achieved resolution being 5 kb (Rao et al. 2014). At this resolution promoter and enhancer contacts can not be accurately distinguished. To circumvent this shortcoming, methods to enrich the contacts of interest have been designed whereby promoter interactions are enriched by using promoter-capture HiC (Mifsud et al. 2015) and enhancer interactions are enriched by the H3K27ac HiChIP method (Mumbach et al. 2017).

CRISPR-Cas9

The other highly efficient tool to have emerged to dissect the function of putative cis-regulatory elements is the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 system (Doudna et al. 2014). Using CRISPR-Cas9, manipulations to the genome are achieved endogenously by DNA-targeting using a single guide RNA (gRNA) and nickase activity of Cas9 enzyme to introduce double-stranded DNA breaks that are repaired by non-homologous end joining (NHEJ) (Doudna et al. 2014). Alternatively, enzymatically deactivated Cas9 (dCas9) fused to desired modulators for activation or repression have been used to modify chromatin structure (Gilbert et al. 2014).

Furthermore, the technique has been multiplexed by the use of CRISPR screens to study CREs. These employ a library of gRNAs and outcomes are assessed at the functional or phenotypic level. A CRISPR-Cas9 dropout screen was able to identify novel enhancers regulating oncogenic p53 and estrogen receptor alpha (ERα) in cancer cells by scoring enhancer elements based on senescence induction (Korkmaz et al. 2016). Another study performed a mutagenesis screen with a pool of around 18,000 gRNAs to elucidate the functional effects of enhancers which confer resistance to a BRAF inhibitor vemurafenib in a melanoma cell line. Functional enhancer elements associated with cancer drug resistance were distinguished and causally linked to gene expression loss and a disease-linked phenotype (Sanjana et al. 2016).

Although CRISPR has become the predominantly attractive technique for studying

enhancer function a range of difficulties exist when assessing the effects of CRISPR-modulated enhancers due to enhancer redundancy (where another enhancer may adopt the function of the target one), ambiguity of the correct gene target and insufficient disruption of an enhancer region.

1.3 Vascular endothelial lineage specification, maintenance and function

At its core the vascular endothelium maintains a general homeostatic state. The organised vascular network is established during embryonic development in a process termed vasculogenesis. Following the formation of a vascular network new sprouting blood vessels can arise from existing ones during angiogenesis which is required for vascular remodelling. Here the activated endothelium responds to inflammatory cues to produce angiocrine growth factor signals to re-establish homeostatic conditions (Rafii et al. 2016).

The endothelium is pivotal to the maintenance of healthy vasculature and is attained by the regulation of a plethora of functional properties; permeability, regulation of vascular tone, paracrine signalling and junctional stability among others. Here I describe these multitude of functions, the means by which endothelial cells (EC) are derived and the recent studies on the molecular mechanisms of their function.

1.3.1 The vascular system

The vascular system is a fundamental organ in all vertebrates. It supplies oxygen and nutrients to surrounding cells and eliminates the byproducts of metabolism through a network of arteries, capillaries and veins. Only the vascular system along with the immune system consists of cells that are in contact with almost every other cell in

the body. The vascular ECs, which line the inner blood vessels, attain tissue-specific functionality. The ECs acquire specialised identities presumably from signals of the regional organ environment to support the specific functions of the organ (Adams et al. 2007; Carmeliet 2003). This is portrayed by the presence of endocardial cells in the heart, distinct sinusoidal ECs of the liver and bone and the ECs of the Schlemm's canal of the eye (Augustin et al. 2017).

Dysfunction of the vascular system entails dysfunction at the level of large vessels, the microvasculature and the underlying endothelium. Of these, endothelial dysfunction is known to contribute most and acts as the perpetrator of many inflammatory, infectious and immune disorders, as well as metastasis in cancer (Carmeliet 2003).

1.3.2 Control of endothelial cell differentiation

During the early stages of mammalian embryonic development the structure referred to as the the primitive streak develops in the blastula through Wnt/β-catenin, activin/nodal and bone morphogenetic protein (BMP) signalling (Patsch et al. 2015). Mesodermal cells then differentiate into angioblasts which migrate into the blood islands of the yolk sac where they coalesce and create the primitive network that then differentiate into ECs in response to vascular endothelial growth factor (VEGF)/ VEGF receptor (VEGFR) signalling (vasculogenesis). Modulation in VEGF concentrations control the arteriovenous axis; high VEGF levels drive arterial commitment and low VEGF levels promote venous specification (Sriram et al. 2015). The schematic to illustrate the processes of vasculogenesis, angiogenesis and vascular wall formation are depicted in Figure 1.6.

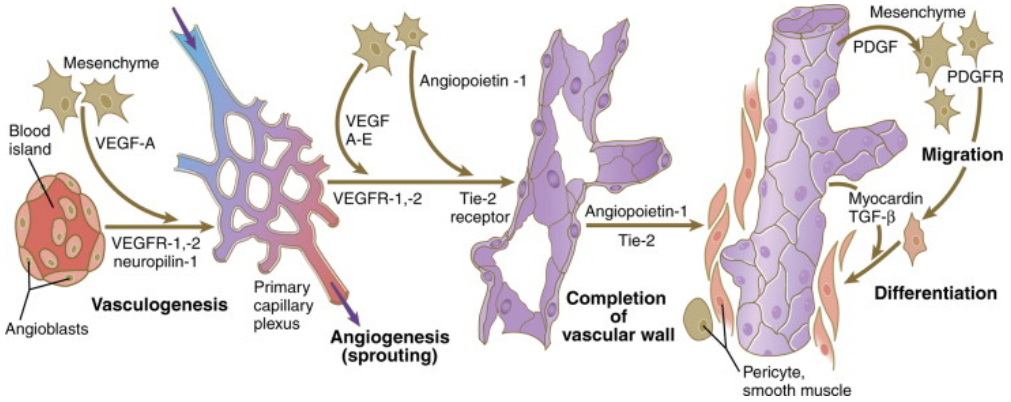


Figure 1.6: The development of the vascular system. Angioblasts migrate into the blood islands and express VEGFR2 which when stimulated by VEGF-A forms the primitive vascular network during vasculogenesis. In response to growth factor stimuli the ECs of the plexus can form vascular sprouts in the early stages of angiogenesis. The vascular wall is then stabilised by the recruitment of cells from the mesenchyme; pericytes and smooth muscle cells. Image reproduced with permission from the rights holder, Elsevier from Reference Module in Biomedical Sciences, B. Carlson, 2014.

During differentiation, endothelial progenitors are in close lineage association with haematopoietic progenitors as they share a number of the same TFs and signalling molecules. It has previously been speculated that a common precursor exists to derive both cell types termed the hemangioblast (Eichmann et al. 1997; Vogeli et al. 2006). However, endothelial progenitor cells have been detected prior to haematopoietic cells in the mouse embryo (Furuta et al. 2006) with the overall association between the two cell types remaining widely debatable.

1.3.3 Heterogeneity of endothelial cells

Distinct and versatile functions of tissue-specific endothelial cells is critical to their many functions. It is still unclear whether endothelial heterogeneity arises during development in their environmental niche or is acquired post differentiation. A study has investigated the EC heterogeneity of four major organs (heart, lung, liver and kidney) from human fetal tissue and showed profound differences in angiogenic potential,

metabolic regulation and vessel barrier functions (Marcu et al. 2018). Another interesting tissue-specific property of ECs are their association with mural cells (vascular smooth muscle cells and pericytes). For example, in the heart mural cell have been found to originate from a common progenitor population during the developmental stage. This shared progenitor is not evident in other tissues as ECs and mural cells segregate early in the differentiation stage (Chen et al. 2016).

Another property mediating the heterogeneity of EC is their exposure to varying biophysical forces potentiated by blood flow of different vessel types. Via their many receptors ECs sense mechanotransduction signals that activate signalling pathways leading to changes in the vascular phenotype and function (Chistiakov et al. 2017). Disturbed flow patterns lead to clinical disease with the most characterised being the spatially predictable instances of atherosclerosis at branch points and curved regions of arteries. The arterial architecture of these regions is particularly permissive to levels of high shear stress promoting pro-inflammation, pro-coagulation and reduced NO production required for vasodilation (Davies et al. 2013).

At the molecular level studies have suggested that a combination of multiple factors are responsible for tissue-specific EC types in mice. These include different TFs, discussed later, and angiocrine factors such as higher expression of CD37 in the bone marrow, liver and kidney and CD177 (*Kit*) in liver (Nolan et al. 2013).

1.3.4 Pathways regulating homeostasis in endothelial cells

The fundamental homeostatic state of the endothelium is achieved by physiological mechanisms to prevent endothelial dysregulation. There are multiple functions of the endothelium which define the homeostatic state.

Post differentiation, the majority of endothelial cells are considered quiescent as they do not replicate frequently; exceptions being ECs residing in the corpus luteum and

uterus which undergo extensive proliferative cycles (Aird 2007). The molecular mechanisms underlying endothelial quiescence have been investigated in the lung endothelium of mice where changes in the DNA methylation state were demonstrated, with methylation increases in intronic enhancer regions and attenuation in intragenic regions that specifically direct endothelial quiescence (Schlereth et al. 2018).

Vascular tone is also a process that requires homeostatic control. The communication between ECs and underlying smooth muscle cells mediates release of vasoactive factors from the endothelium. These include vasodilators (including nitric oxide (NO), prostacyclin (PGI_2) and endothelial-derived hyperpolarizing factor (EDHF)) and vasoconstrictors (including endothelin-1 (ET-1), angiotensin II (AII) and vasopressin (AVP)) to regulate vessel tone and reactivity by signalling through a multitude of pathways. These mechanisms are not uniform in the vascular system. Yet, the vascular beds that utilise these mechanisms allow physiological blood pressure to be tightly regulated.

Further physiological balance governed by the endothelium is the maintenance of blood in a fluid state. This is regulated by limiting the manifestation of coagulation, thrombosis and platelet activation. The role of the ECs in haemostasis are to respond to local stresses by expressing anti-platelet and anti-coagulation factors that prevent platelet aggregation. One key component of this pathway is the large multimeric glycoprotein von Willebrand factor (VWF). VWF is constitutively expressed by ECs and is functionally important in haemostasis as it binds a number of molecules and critically adheres to platelets (André et al. 2000).

Another physiological process of blood vessels is the highly effective distribution of blood to the many parts of the body. In an integrated manner, they regulate permeability to transfer molecules between tissue and blood ranging from water and ions to whole cells such as lymphocytes during an inflammatory response (Matheny et al. 2000). Permeability is dependent on the type and physiology of a tissues. Some tissues are highly permeable such as the high endothelial venules and other tissues are completely imper-

meable such as the brain vasculature (Hirase et al. 1997; Matheny et al. 2000). These differences are determined by the composition of the intercellular junctions. The two key intercellular junctions are adherens junctions and tight junctions with the prior proving critical to the organisation of the latter (Taddei et al. 2008). The major protein component of adherens junctions in the endothelium is vascular endothelial cadherin (VE-cadherin) that is ubiquitously expressed in all ECs and that functionally binds to several intracellular proteins including α-catenin and β-catenin (Vestweber 2008).

Finally, a key active endothelial process is the growth of new blood vessels from existing ones as described by angiogenesis. The most common form is sprouting angiogenesis in which a endothelial tip cell guides a capillary sprout through concentration differences of an angiogenic stimulus. VEGF and its receptor are the master regulators of angiogenesis (Gerhardt et al. 2003). The high oxygen demand within tissues largely governs angiogenesis where hypoxic tissue secretes VEGF, a transcriptional target of hypoxia-inducible factors (HIFα and HIFβ), and together with its receptor VEGFR facilitates EC migration. Additionally, the HIF pathway also regulates a host of other pro-angiogenic factors including platelet-derived growth factor (PDGF) and its receptor PDGFR and ligands angiopoietin-1 (Ang-1) and angiopoietin-2 (Ang-2) and their shared receptor Tie2 (a receptor tyrosine kinase) (Krock et al. 2011).

The angiopoietin ligands have more general functions in EC survival, maturation and general control of vascular homeostasis. Ang-1 specifically inhibits EC apoptosis (Kwak et al. 1999; Papapetropoulos et al. 2000), stimulates EC migration (Carlson et al. 2001) and promotes vascular tube formation (Chen et al. 2004). At the cell signalling level, Ang-1 potently activates mitogen-activated protein kinase (MAPK) and phosphoinositide 3-kinase (PI3K) downstream to trigger cell mitosis (Kanda et al. 2005).

1.3.5 Experimental models to study endothelial cells

With the aforementioned heterogeneity of different ECs, the choice of EC models for molecular interrogation are consequential for the experimental setup. Human umbilical vein ECs (HUVECs) have been the widely adopted and characterised EC model yet they do not specifically represent an universal model for all molecular and functional studies. Other macrovascular ECs used are human aortic EC (HAEC) for their higher relevance to atherosclerotic progression. For detailed organ-specific work, microvascular cells have become a valuable resource. Studies have employed microvascular cells derived from mice (Nolan et al. 2013; Sabbagh et al. 2018) and human (Heidemann et al. 2003; Lang et al. 2003; Marcu et al. 2018) to investigate cellular, molecular and transcriptional EC diversity.

Additionally, endothelial colony forming cells (ECFC) are progenitor EC that are derived from various blood sources and show clonal potential. They are a highly attractive model for regenerative medicine with potential in vascular repair, gene therapy and tissue bioengineering (Paschalaki et al. 2018). Promising studies *in vivo* have reported successful outcomes following ischaemic events. Endothelial progenitor cells were shown to decrease infarct size and increase vessel density in a pig model (Dubois et al. 2010) and elsewhere shown to ameliorate the effects of acute kidney injury following renal ischaemia in rats (Collett et al. 2017).

1.3.6 Endothelial transcription factor families

The TF families that are highly specific for ECs include, but are not limited to, ETS, FOX, GATA and SOX. The ETS factors have been found to regulate most, if not all, recognised endothelial genes although they are evidently not endothelial-specific (Meadows et al. 2011). During development ets variant 2 (ETV2) is the most important, expressed transiently in mice before being rapidly degraded and undetectable in the adult

(Lee et al. 2008). Other ETS factors are maintained throughout adulthood including ERG, ETS1 and FLI1 (friend leukaemia integration-site 1). Interestingly, ablation of these ETS factors result in milder vascular phenotypes than would be expected, likely due to compensation for the loss by other ETS factors (Meadows et al. 2011).

The other large TF family that mediates many EC functions is the FOX family. Notably, the promoters of several endothelial genes possess FOX:ETS binding motifs (De Val et al. 2008). Gene expression profiles report the regulation of angiogenesis and vascular remodelling-associated genes by FOXO1 and FOXO3 (Potente et al. 2005). Members of the C subgroup of forkhead TFs, FOXC1 and FOXC2, have been found to regulate arterio-venous specification (Papanicolaou et al. 2008) with malformations exhibited in double *Foxc1/Foxc2* knockout mice (Kume et al. 2001).

GATA factors are preferentially known for their role in haematopoietic cells. Yet, key findings have implicated a role for GATA2 in vessel development (Lim et al. 2012) and maintenance of endothelial gene expression, including that of endothelial genes *ENDOMUCIN* and *VEGFR2* (Kanki et al. 2011).

The SOX TFs found to be related to vascular functions are SOX7, SOX17 and SOX18 which are expressed early in embryonic development. Sox17 and Sox18 seem to be redundant in cardiovascular development with double null embryos showing more severe defects than Sox17-single null embryos (Sakamoto et al. 2007). In other work, inactivation of Sox17 was accompanied with the absence of arterial differentiation and vascular remodelling (Corada et al. 2013).

1.3.7 Transcriptional and epigenetic regulation of endothelial cells

The derivation of EC for therapeutic approaches has prompted the studies into the generation of ECs. The most direct conversion was from adult skin fibroblasts and human embryonic stem cells (ESCs) into ECs using only ETV2 (Elcheva et al. 2014; Morita et al.

2015). The cells derived characteristically expressed key endothelial markers, CD31 and VE-cadherin, and vascular properties such as formation of capillary structures (Morita et al. 2015) or formation of vascular tubes (Elcheva et al. 2014). Interestingly, the skin fibroblast derived ECs were proposed as a venous EC subset. Another study also highlighted the indispensable expression of *ETV2* for conversion of human ESCs into EC. In this study ETV2 was transiently expressed along with constitutive expression of other ETS factors ERG and FLI1 and the inhibition of transforming growth factor β (TGFβ) (Ginsberg et al. 2012). Another model was initiated from mouse ESCs cultured in a differentiation media (including fetal calf serum) for 4 days before sorting for cells expressing VEGFR2 (Flk-1). These cells were then induced with VEGF to derive ECs and these cells showed a rapid onset of *Etv2* gene expression followed by the expression of EC master regulators *Gata2*, *Fli1*, *Sox7* and *Sox18* (Kanki et al. 2017). The models from these studies are summarised in Figure 1.7.

The molecular signatures of differentiated EC have been defined by profiling different EC models. In HAEC, it was reported that the core TF complex at endothelial-specific enhancers consists of AP-1, ETS and GATA TFs (Hogan et al. 2017). Furthermore, an *in vivo* model with a biotin-labelled bio peptide linked to the p300 protein (termed $Ep300^{fb}$ bioChIP-seq) has been used to accurately determine endothelial-specific enhancers (Zhou et al. 2017). In this study, motif determination within these enhancer regions identified ETS and GATA motifs along with the ETS:FOX heterodimer to be most highly enriched highlighting the importance of ETS factors in the endothelial gene expression program. In organ-specific ECs, the ETS factor family members Erg, Ets1 and Fli1 were all expressed across brain, liver, lung and kidney EC from mice. For other TFs, tissue-specific expression patterns were observed with enrichment of Zic3 and Lef1 in brain, Gata4 and Maf in liver, Foxf1 and Gata3 in the lung and Irx3 and Hoxa7 in kidney (Sabbagh et al. 2018).

Recent studies have explored chromatin architecture of the endothelial landscape. Local chromatin accessibility determined by ATAC-seq in ECs shows further tissue speci-

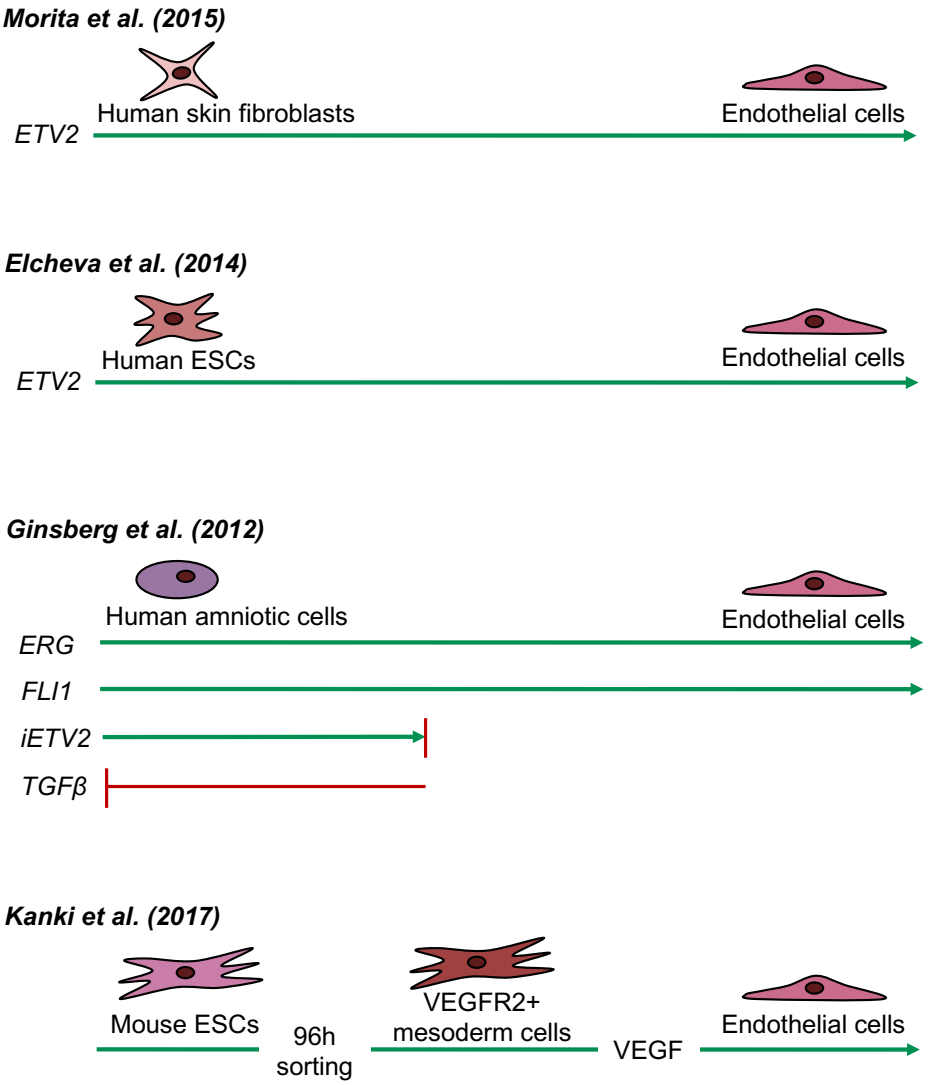


Figure 1.7: Models of EC induction. Morita et al. (2015) and Elcheva et al. (2014) directly converted human adult skin fibroblasts and human ESCs respectively to ECs with only the expression of pioneer TF ETV2. Ginsberg et al. (2012) converted human amniotic cells into EC by constituent expression of TFs ERG and FLI1 and transient expression of an inducible ETV2 (iETV2) in conjunction with the repression of TGFβ signalling. Kanki et al. (2017) did not overexpress TFs to induce EC but rather allowed the differentiation of mouse ESCs for 96h, followed by sorting for VEGFR2 positive mesoderm cells which were converted to EC upon the stimulation with VEGF.

ficity in mice as only 14% of promoter-distal ATAC-seq peaks are shared between brain, liver, lung and kidney ECs with the brain endothelium being most divergent versus the other three peripheral tissues (Sabbagh et al. 2018). Large structural conformation of the endothelial chromatin as determined by TAD composition has reportedly been largely unaffected in response to various stimuli including hypoxia assayed by tethered conformation capture (TCC) (Niskanen et al. 2018) and TNFα investigated using Hi-C (Lalonde et al. 2019).

1.3.8 Super-enhancer control in endothelial cell identity

To investigate the control of EC identity studies have identified super-enhancers in different EC types and under different physiological stimuli. One study identified super-enhancers in basal HAEC and reported their association with TFs of the ETS and AP-1 families (Hogan et al. 2017). A study using HUVEC showed that compartments of the genome which are actively transcribed in response to VEGF-A stimulation are those harbouring super-enhancer identity (Kaikkonen et al. 2014). Furthermore, regions in which chromatin interactions within a TAD where enriched in HUVEC versus endothelial progenitor cells were also regions identified as HUVEC-specific super-enhancers (Niskanen et al. 2018). A study by Brown et al. demonstrated the redistribution of super-enhancers in HUVEC in response to TNF-α stimulation by using the occupancy of BRD4 as a proxy for super-enhancer sites. The activation of NF-κB establishes the recruitment of BRD4 for the induction of a rapid inflammatory response (Brown et al. 2014). These studies demonstrate the importance of super-enhancers as regions in the endothelial genome critical to regulation, chromatin structure organisation and dynamic signal responses.

1.4 ETS related gene (ERG)

The E-twenty-six (ETS) TF family members are one of the most predominant sets of TFs governing EC biology. The ETS proteins are highly conserved among each other sharing the ETS DNA-binding domain that has specificity for the sequence recognition motif 5'-GGA(A/T)-3'. In humans there are 28 different ETS members and half of them are ubiquitously expressed in all cell types currently tested (Hollenhorst et al. 2011). Others are more cell type restricted including the ETS related gene (ERG) which is the highest expressed ETS factor in mature, quiescent EC. In prostate cancer, ERG has also been implicated as an oncogene as a result of chromosomal rearrangements with a key deregulation. In this section I will describe the current knowledge on ERG structure and function in EC and other cell types focussing on some of the most recent genomic studies.

1.4.1 Expression and regulation of ERG

ERG is expressed from the chromosome 21q22 locus which can expression over 30 ERG isoform variants (Zammarchi et al. 2013). ERG expression has largely been restricted to ECs and otherwise it has only been detected in haematopoietic stem cells (HSCs) and some haematopoietic progenitors namely T- and B-lymphocytes and megakaryocytic cells (Mohamed et al. 2010). Its expression is related to the derivation of EC and haematopoietic cells from the mesoderm during development. ERG is well conserved across species with studies across xenopus, zebrafish, mouse and human. In mouse ECs converted from ESC via VEGF *ERG* expression was shown to peak late after VEGF stimulation (Kanki et al. 2017). Another study deleted ERG in murine ESC and reported a significant reduction in their ability to differentiate in EC (Nikolova-Krstevski et al. 2009). In differentiated mouse and human ECs ERG expression remains high in the healthy vasculature (Vlaeminck-Guillem et al. 2000; Yuan et al. 2009). ERG's closest homolog is Friend leukemia integration 1 (FLI-1) sharing 65% homology overall and 98%

homology at the DNA binding domain. This would suggest that the two TF can occupy similar loci in the genome. *ERG* and *FLI-1* have been shown to genetically interact in haematopoiesis (Kruse et al. 2009) and their relationship in ECs has been explored by simultaneous knockdown of *ERG* and *FLI-1* in human pulmonary arterial EC and shown to have a synergistic or additive effect on gene expression suggesting the two TFs can co-regulate gene expression (Looney et al. 2017).

The regulation of ERG levels in ECs have been investigated at the mRNA and protein levels. One mechanism by which ERG levels are maintained are via the ubiquitin-proteosome pathway. ERG levels are shown to be rapidly degraded when the enzyme ubiquitin-specific peptidase 9, X-linked (USP9X), known to stabilise ERG is inhibited (Wang et al. 2014). Additionally, an E3 ubiquitin ligase, speckle-type POZ protein (SPOP) has been shown to actively promote ubiquitination and proteosome degradation of ERG protein (An et al. 2015). Both these degradation targeting pathways have yet to be directly confirmed in ECs. ERG protein levels are also known to be dramatically reduced in response to pro-inflammatory stimuli such as TNF-α (McLaughlin et al. 1999), lipopolysaccharide (LPS) and IL-1β (Yuan et al. 2009). This has further been implicated in tissues from human disease; human coronary atherosclerotic plaques show a loss of ERG expression in the overlying endothelium (Sperone et al. 2011). ERG levels were also attenuated in liver tissue derived from patients with chronic end-stage liver fibrosis which correlates with endothelial-to-mesenchymal transition (EndMT) (Dufton et al. 2017). Representative images from these studies are reproduced in Figure 1.8.

The activity of ERG is known to be controlled by signal transduction cascades that regulate its transcriptional activity. A study has demonstrated that VEGF/Mitogen-activated protein kinase (MAPK) signalling mediates the activity of ERG occupancy at regulatory elements of the delta like canonical notch ligand 4 (*DLL4*) and Notch receptor 4 (*NOTCH4*) genes in arterial EC (Whyte et al. 2013). Further studies have revealed that the VEGF/MAPK signalling pathway leads to the phosphorylation of ERG, predominantly by MAPK1, at serine residues 96, 215 and 276 and demonstrated that these

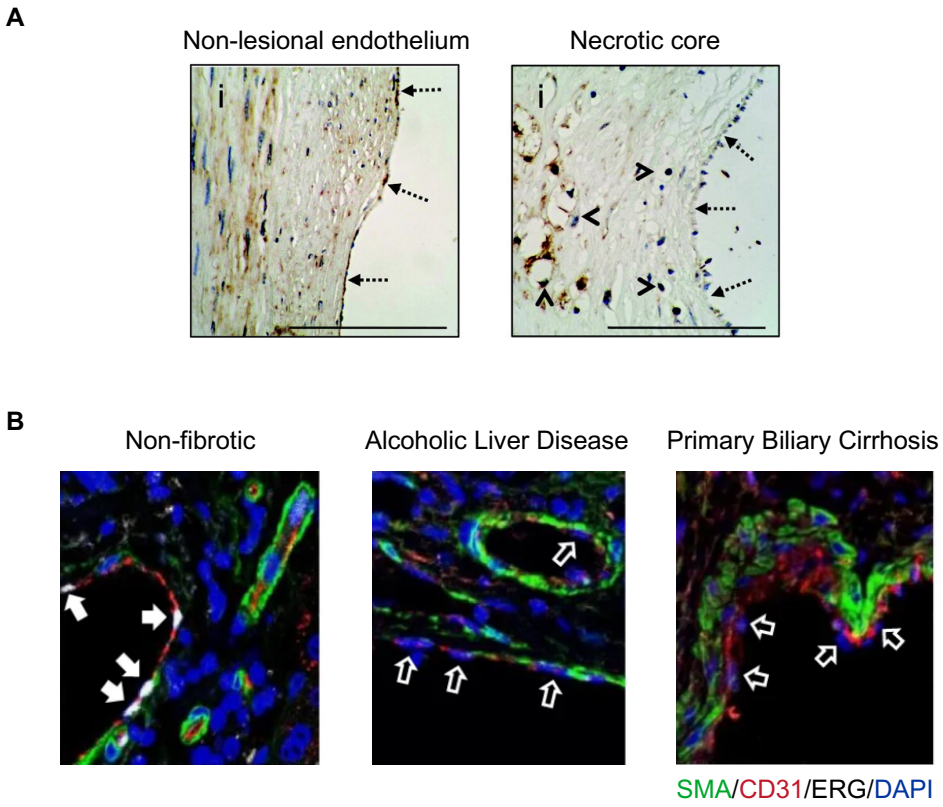


Figure 1.8: ERG expression is absent in human disease. (A) ERG expression is present at non-lesional endothelium, the unaffected section of the vessel, in coronary artery endothelium. ERG expression is lost in the affected region of the endothelium above the necrotic core. ERG staining is indicated by the dotted arrows. Scale bar represents 100 μm at 40x objective. Images from Sperone et al. (2011). (B) Immunofluorescence images of human liver sections from patient donors with non-fibrotic, end-stage alcoholic liver disease and primary biliary cirrhosis. Staining is for smooth muscle actin (SMA; green), CD31 (red), ERG (white) and DAPI (blue). Full white arrows in the non-fibrotic section indicates co-staining for ERG and CD31 whereas empty white arrows in chronic end-stage sections show CD31 staining but no ERG. Scale bar represents 50 μm. Images from Dufton et al. (2017).

post-translational modifications increase ERG binding, with DNA and other TFs, and therefore also enhance transactivation potential (Selvaraj et al. 2015). Interestingly, structural studies have shown that ERG can also autoregulate its own DNA-binding by an autoinhibitory mechanism through moderate allosteric changes at regions N- and C-terminally adjacent to its ETS DNA-binding domain (Regan et al. 2013).

1.4.2 ERG protein structure

The ERG protein structure belongs to the superfamily of TFs characterised by a basic helix-loop-helix domain. Beyond the ETS domain shared by all ETS factors for DNA binding, ERG and ten other ETS factors also share the pointed (PNT) domain implicated in protein-protein interactions. It remains unknown whether interactions between ETS TFs occur via the PNT domain directly. Homodimerization of the ERG protein was shown to involve the ETS domain and a region consisting of the first 200 residue from the N-terminal domain which includes the PNT domain (Carrère et al. 1998). However, isolated PNT domains remain monomeric in solution indicating that the ETS domain is indispensable for protein-protein interactions (Mackereth et al. 2004). Functionally ERG dimers are unable to bind DNA (Carrère et al. 1998) and further studies would determine whether dimerization is a inhibitory mechanism *in vivo*. I discuss interactions between ERG and other ETS factors further below.

1.4.3 Molecular functions of ERG

ERG is a TF that can act to both transcriptionally activate and repress its target genes. In ECs it drives the genes required for the endothelial lineage and for homeostasis (Birdsey et al. 2008), meanwhile repressing the genes that are pro-inflammatory such as ICAM-1 and interleukin-8 (IL-8) (Sperone et al. 2011; Yuan et al. 2009). Specifically, ERG has been reported to be involved in the regulation of endothelial junctional adhesion proteins VE-cadherin (CDH5) (Birdsey et al. 2008), claudin-5 (CLDN5) (Yuan et al.

2012) and ICAM-2 (McLaughlin et al. 1999). Furthermore, ERG has been shown to regulate cytoskeletal proteins such as histone deacetylase 6 (HDAC6) (Birdsey et al. 2012) and the GTPase Ras Homolog Family Member J (RHOJ) (Yuan et al. 2011). Functionally, ERG inhibition has been shown to reduce HUVEC migration rate with attenuation in lamellipodia formation (Birdsey et al. 2012).

Critically, during development and in the adult tissue Erg is essential in mice. Constitutive EC-specific deletion of Erg in mouse embryos ($Erg^{cEC\text{-}KO}$) results in blood vessel disorganisation, reduction in number of blood vessels, haemorrhages that ultimately lead to embryonic lethality at E10.5-11.5 (Birdsey et al. 2015). In a mouse model with global Erg deletion the observed lack of blood vessel formation was a result of improper EC migration (Vijayaraj et al. 2012), simulating the findings *in vitro*.

ERG has been implicated in a number of signalling pathways including the Wnt/β-catenin, Notch and TGFβ pathways (Birdsey et al. 2015; Dufton et al. 2017; Shah et al. 2017). Through Wnt/β-catenin signalling ERG regulates angiogenesis and vascular stability both *in vitro* and *in vivo* (Birdsey et al. 2015). ERG was further found to coordinate vascular maturation and stability through the regulation of Notch ligands *DLL4* and *Jag1* by potentiating Ang-1-dependent signals (Shah et al. 2017). This study mechanistically demonstrated that Ang-1 phosphorylates ERG and enhances the occupancy of ERG at the enhancers of *DLL4*. Key ERG genetic targets and functionalities are illustrated in Figure 1.9.

Interestingly, ERG has a pivotal role to play in HSC maintenance. It has been found to be required for HSC self-renewal (Ng et al. 2011) and further studies using a heterozygous ERG deletion mouse model showed that in bone marrow transplants the partially ERG-depleted bone marrow could not support haematopoiesis in a healthy host (Taoudi et al. 2011).

A recent finding highlighted a novel role for ERG with regulatory mechanisms at the post-transcriptional level. ERG was found to be recruited to mRNA via RNA-binding

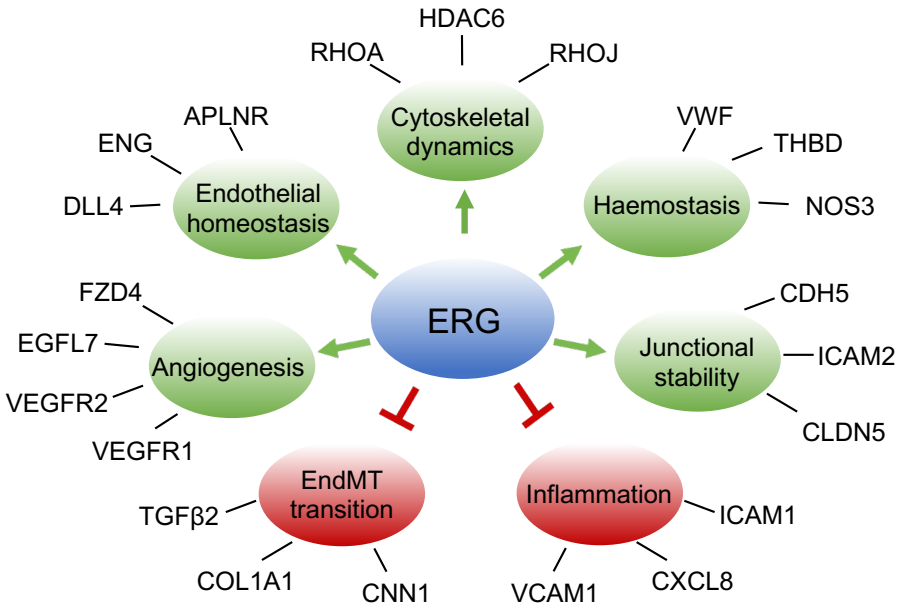


Figure 1.9: ERG is a central modulator of the endothelial lineage. ERG acts to both drive (green) endothelial functions angiogenesis, homeostasis, cytoskeletal dynamics, haemostasis and junctional stability by activating the expression of genes labelled for each function. ERG also represses (red) functions including endothelial-to-mesenchymal (EndMT) transition and inflammation by directly repressing the targets listed with the functions. The regulation by ERG of the listed genes were reported as follows: APLNR (Lathen et al. (2014)), CDH5 (Birdsey et al. (2008)), CLDN5 (Yuan et al. (2012)), CNN1, COL1A1, TGFβ2 (Dufton et al. (2017)), CXCL8 (Yuan et al. (2009)), DLL4 (Whyte et al. (2013)), EGFL7 (Le Bras et al. (2010)), ENG (Pimanda et al. (2006)), FZD4 (Birdsey et al. (2015)), HDAC6 (Birdsey et al. (2012)), ICAM1, VCAM1 (Sperone et al. (2011)), ICAM2 (McLaughlin et al. (1999)), RHOA, VWF (McLaughlin et al. (2001)), RHOJ (Yuan et al. (2011)), THBD (Peghaire et al. (2019)), VEGFR1 (Wakiya et al. (1996)), VEGFR2 (Meadows et al. (2009)).

protein RBPMS and promote mRNA degradation by binding to CNOT2, a component of the CCR4-NOT deadenylation complex responsible for 3'-5' degradation. Globally these ERG-targeting mRNAs were predominantly Aurora kinases which require to be decayed at S phase of the cell cycle for entry to mitosis (Rambout et al. 2016).

1.4.4 Oncogenic role of ERG in prostate cancer

In striking contrast to its homeostatic role in EC, ERG is aberrantly expressed in multiple cancer types. Through a range of gene fusions ERG plays an oncogenic role in Ewing's sarcoma, acute myeloid leukemia (AML) and prostate cancer (Sorensen et al. 1994; Thoms et al. 2011; Tomlins et al. 2005). Its most prominent role is in prostate cancer where a chromosomal rearrangement occurs between the 5' promoter region of *TMPRSS2*, transmembrane serine, protease 2, and the 3' region of *ERG* resulting in full-length or N-terminally truncated ERG protein (Tomlins et al. 2005). See Figure 1.10 for the *TMPRSS2-ERG* mRNA and encoded protein structure. The *TMPRSS2-ERG* fusion is clinically prevalent in approximately 50% of prostate cancers (Tomlins et al. 2005). Harbouring the fusion gene is linked to enhanced invasiveness, increased metastasis and poorer prognostic outcomes (Yoshimoto et al. 2008) which are similar characteristics of ERG chromosomal rearrangements in AML (Pigazzi et al. 2012). The high expression of *ERG* from the fusion gene is a result of the fact that *TMPRSS2* is an androgen-regulated gene and prostate cancer cells are regulated by androgen signalling (Tomlins et al. 2005). In fact it has been reported that androgen signalling induces chromosomal movements which bring the loci of *TMPRSS2* and *ERG* in closer proximity thus increasing the probability of a gene fusion event (Mani et al. 2009). Further work has demonstrated that the *TMPRSS2-ERG* fusion product also binds to the native ERG locus and drives the expression of wild-type ERG in prostate cancer tissue (Mani et al. 2011).

At the transcriptional level oncogenic expression of ERG in prostate cancer is regulated by a number of distinct molecular mechanisms. Recently, it was identified that the

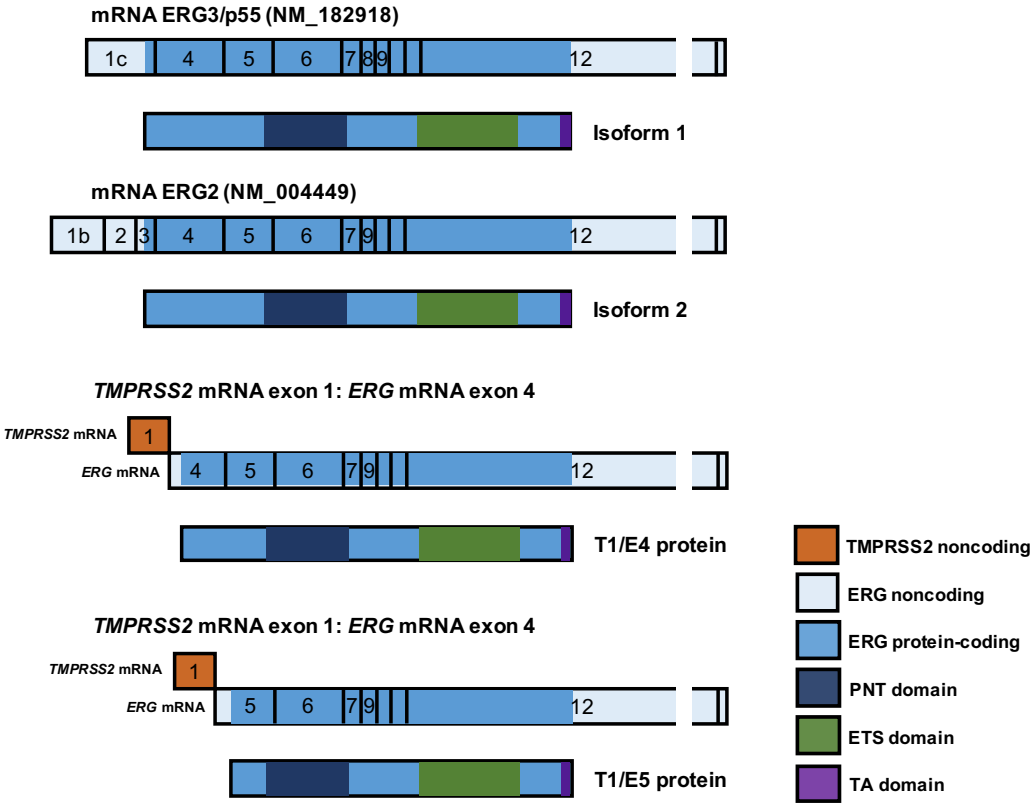


Figure 1.10: ERG transcript variants and corresponding protein isoforms in ECs and prostate cancer. Transcript variants show all coding and non-coding exons for the most commonly expressed ERG3/p55 variant which produces isoform 1. ERG2 encodes the protein isoform most representative of the protein encoded by the TMPRSS2-ERG fusion in prostate cancer. The TMPRSS2-ERG fusion most frequently occurs by the translocation of *TMPRSS2* exon 1 to either *ERG* exon 4 or 5 whereby the expressed protein has all the functional domains as the wild-type EC ERG; PNT, ETS and TA (transactivation) domains. Crucially, the N-terminus of the ERG encoded by the TMPRSS2-ERG fusion is lost. Coding and non-coding sequences and domains of the ERG protein are indicated in the key.

ubiquitin ligase SPOP marks the transcriptional co-activator BRD4, a critical oncoprotein, for destruction in prostate cancer however in tumours that confer SPOP mutations BRD4 levels are elevated and consequently act with TFs androgen receptor (AR) and ERG to facilitate prostate tumourigenesis (Dai et al. 2017). Furthermore, ERG co-opts AR by recruiting it for co-binding to novel sites across the chromatin (Chen et al. 2013; Yu et al. 2010). In a study that utilised the ERG-expressing prostate cancer cell line VCaP (vertebral cancer of the prostate) long-range chromatin interactions were profiled and found that ERG and AR cooperate to establish a global chromatin interactome in prostate cancer cells which acts to transcriptionally regulate AR target genes (Zhang et al. 2019). In line with this, a previous report demonstrated that the chromatin conformation in prostate epithelial cells was dramatically rearranged upon the over-expression of ERG and the regions with differential chromatin interactions were enriched for ERG binding sites (Rickman et al. 2012). Finally, a recent report showed that the super-enhancer associated with endogenous *TMPRSS2* expression spreads into the *ERG* locus and facilities its high expression (Kron et al. 2017).

1.4.5 Functional interactions between ERG and co-regulators

In relation to the interaction between ERG and AR in prostate cancer, ERG has been demonstrated to interact directly and indirectly with other trans-acting factors. Its interactions with other ETS TFs have been determined by *in silico* studies where physical interactions with members FLI-1, ETS-2, ER81 and PU-1 were detected whereas, interestingly, ETS-1 was detected to have little to no interaction with ERG (Carrère et al. 1998). A more recent report observed the impact of depleting ERG and FLI-1 in combination in HUVEC and indicated a significant loss of active histone mark H3K27ac (Nagai et al. 2018).

Further studies have detected interaction between ERG and other TFs. The ubiquitously expressed members of the AP-1 family, FOS and JUN, have both been shown to

bind to the ERG protein including to form a ternary complex of ERG/FOS/JUN in the presence of DNA at their respective DNA binding consensus motifs (Carrère et al. 1998). This synergistic relationship between ERG and JUN has been further characterised in live cells using fluorescence resonance energy transfer (FRET) analysis (Camuzeaux et al. 2005). Interestingly, amino acid residues in ERG that were responsible for the binding interaction between ERG and AP-1 have been elucidated. Mutations at arginine Arg^{367} abolished both the interaction with JUN and DNA binding and mutations at tyrosine residue Tyr^{371} abrogated the synergistic interaction with the FOS/JUN heterodimer (Verger et al. 2001). More recently, an intricate study used genetic approaches which included leveraging the natural genetic variations in mice to establish a central role for AP-1 as a collaborative TF that co-selects enhancers in collaboration with cell-type specific TFs (Vierbuchen et al. 2017).

Interactions have also been identified between ERG and other TFs that are essential to endothelial development and identity. ERG was found to associate with KLF2 in a physical complex that synergistically activates the expression of VEGF receptor 2 (VEGFR-2) (Meadows et al. 2009). Binding has also been detected between ERG and FOXO1 in prostate cancer VCaP cells and by use of recombinant protein constructs the interaction was identified to occur between ERG amino acid residues 40 to 200 (Yang et al. 2017). Moreover, a direct interaction was also identified between ERG and RUNX1, a TF known to regulate HSC differentiation (Wilson et al. 2010). It is also plausible that ERG can associate with other TFs indirectly as part of the larger transcriptional complex. In HSCs a heptad of TFs has been described which include ERG, FLI-1 and GATA2 by the integration of multiple ChIP-seq datasets (Wilson et al. 2010).

Binding interactions have also been explored between ERG and co-transcriptional regulators. For transcriptional activation ERG has been shown to cooperate with histone acetyltransferase p300 (Jayaraman et al. 1999). More recently this interaction was confirmed with ERG shown to recruit and physically interact with p300 in response to VEGF stimulation (Fish et al. 2017). Additionally, ERG forms a binding

interaction with transcriptional co-activator and chromatin remodelling complex BAF (SWI/SNF) (Sandoval et al. 2018). ERG is demonstrated to target BAF complexes to occupy chromatin for transcriptional activation by specifically interacting with ARID1A (BAF250A), SMARCA4 (BRG1) and SMARCC1 (BAF155) in prostate cancer VCaP cells (Sandoval et al. 2018). Further studies in prostate cancer cells have shown a physical interaction between ERG and BRD4 (Blee et al. 2016). Through this interaction, a mechanism by which ERG is brought into proximity with acetylated histone lysine residues is depicted as these are recognised by bromodomains of BRD4.

The role of ERG in gene repression is critically important to the maintenance of endothelial homeostasis. ERG has been found to block the binding of TF nuclear factor (NF)-κB to prevent the rapid activation of a pro-inflammatory gene cascade (Sperone et al. 2011). Moreover, ERG has been proposed to recruit the polycomb repressive complex 2 (PRC2) to deposit methyl groups to establish histone modification H3K27me3 to the gene targets inhibited by ERG. Specifically, a direct interaction was confirmed between ERG and the enzymatic methyltransferase domain of PRC2; enhancer of zeste 2 (EZH2) in prostate cancer cells (Kedage et al. 2017).

1.4.6 ERG in vascular diseases

With functions central to vascular endothelial homeostasis it is unsurprising that levels of ERG are important to physiological vascular functions. ERG in mouse tissue has been found to be most highly expressed in the heart and the lung (Yuan et al. 2009). Studies performed by our group in an inducible EC-specific ERG knockout mouse model ($ERG^{iEC\text{-}KO}$) highlight the functions of ERG in vascular physiology as compared to wild-type control mice with $ERG^{iEC\text{-}KO}$ mice defective angiogenesis in the retina postnatally with reduced vascular stability (Birdsey et al. 2015). In lung pathology, ERG has been found to be downregulated in the lungs of patients with pulmonary arterial hypertension (PAH) (Looney et al. 2017).

In other work on patients a novel SNP was identified in the ERG locus associated with abdominal aortic aneurysm (AAA) through a meta-analysis of GWAS data. This SNP narrowly achieved genome-wide statistical significance in the meta-analysis and was shown to be a significant expression quantitative trait loci (eQTL) in blood monocytes only, but not in more relevant tissues such as the aortic adventitia and aortic media (Jones et al. 2017). These findings prompt the loss of ERG as a biomarker of idiopathic diseases.

1.5 Aims of this thesis

It is evident from the existing studies that the transcriptional mechanisms to potentiate endothelial-specific gene expression are only beginning to be uncovered. ERG is a cell-type-restricted TF that functions collaboratively with other transcriptional machinery to establish lineage-specific transcriptional regulation. In this thesis I present the work which investigates the regulatory functions of ERG at enhancer and super-enhancer elements to maintain homeostasis in endothelial cells. The basis of each chapter in this thesis is as follows.

In Chapter 2, I integrate ChIP-seq and transcriptome profiling data from HUVEC to characterise the chromatin and transcriptional signatures in basal ECs regulated by ERG. I identify super-enhancers in HUVEC using histone modification H3K27ac, transcriptional co-activator MED1 and ERG itself. This chapter further presents the functional role ERG plays on preserving endothelial super-enhancers by its effects on H3K27ac and transcriptional co-activator Mediator. It further describes the impact of physiological stimuli Ang-1 on the super-enhancer landscape and, importantly, the link between ERG, super-enhancers and disease-associated genetic variants revealing a prevalent association with cardiovascular diseases.

Chapter 3 explores the comparison of ERG in HUVEC and prostate cancer VCaP

cells at the chromatin and gene regulation level with a focus on the lineage-specificity of super-enhancers. This chapter investigates the lineage-determining potential of ERG and postulates that molecular mechanisms in distinct cell types facilitates selectivity of cell type-specific enhancers and super-enhancers via ERG activity.

In Chapter 4, I explore the TF network that works in concert with ERG to gain functional insight into the interplay between the key TFs in ECs. Furthermore, I determine the functional basis by which ERG-dependent super-enhancers are controlled in endothelial cells which are delineated by co-regulator p300 activity and TF abundance at the chromatin. I partition the chromatin landscape of EC to define functional states associated with TF complex assembly, gene expression and chromatin interaction dynamics.

In Chapter 5, I use genome editing to deconstruct the endothelial super-enhancer linked to VE-cadherin expression. I use both CRISPR-Cas9 for deletion and CRISPR interference for heterochromatin formation in HUVEC and ECFC to target constituent enhancer sites bound by ERG. I extend the method to generate clones and attempt to assess the contribution of each enhancer constituent to the large super-enhancer and its direct effect on VE-cadherin gene expression levels.

2 ERG DRIVES A LINEAGE-SPECIFIC PROGRAM IN ENDOTHELIAL CELLS VIA SUPER-ENHANCERS

2.1 Introduction

Non-coding *cis*-regulatory elements, termed enhancers, play an important role in gene regulation. A subset of highly active enhancers that form large clusters have been described and termed super-enhancers (Hnisz et al. 2013; Lovén et al. 2013; Whyte et al. 2013). Super-enhancers are densely occupied by mediator complex components, transcription factors (TFs) and by histone modification H3K27ac. Functionally, super-enhancers have been reported to coordinate cell identity by driving high expression of genes required for cell type specificity. Hnisz et al. (2013) identified super-enhancers in 86 human cell types including those in HUVEC using H3K27ac. Further studies in HUVEC to determine the effect of proinflammatory stimuli on the distribution of super-enhancers demonstrated that tumour necrosis factor (TNF)-α treatment caused basal super-enhancers to be decommissioned and new inflammatory super-enhancers to be established (Brown et al. 2014). It remains unclear how super-enhancers are formed and maintained integrally in order to maintain high expression of functionally critical genes.

The transcriptional regulation of genes important in endothelial cell identity is directed by key endothelial TFs. Previous work in the Randi group and others have described the prominent role of ERG in regulating multiple pathways in the vasculature, including vascular homeostasis and survival (Birdsey et al. 2012; Birdsey et al. 2008). The mechanism by which ERG controls a lineage-defining transcriptional profile remains unknown. We hypothesised that ERG could be fundamental to maintaining endothelial super-enhancers in mature HUVEC through which it regulates gene expression. In this chapter I show that ERG binds and drives endothelial target genes via super-enhancers and that ERG can be used as a surrogate marker for super-enhancer profiling in en-

dothelial cells. I reveal that ERG-deficient HUVEC demonstrate a loss of enhancers and a subset of super-enhancers. I show functional implications linked with disease by highlighting that ERG super-enhancers are enriched for SNPs associated, most prevalently, with cardiovascular diseases.

2.2 Methods and Data

Culture of HUVEC

Pooled human umbilical vein endothelial cells (HUVEC) were purchased from Lonza (Wokingham, United Kingdom) and cultured in EGM-2 media (Lonza). HUVEC were plated on 1% gelatin.

Chromatin immunoprecipitation with sequencing (ChIP-seq) sample preparation

ChIP-seq for ERG in HUVEC was carried out by by Dr. Youwen Yang (postdoc in Randi's group). ChIP assays were conducted using chromatin from pooled HUVEC (10^7 cells) were performed using 2 µg of rabbit polyclonal antibody against ERG (sc-354X, Santa Cruz Biotechnology), as described (Wilson et al. 2010), with minor modifications by introducing an additional washing step with high salt buffer (0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 500 mM NaCl). Library generation was performed using TruSeq ChIP Sample Prep Kit (Illumina). Briefly, 10 ng DNA ends were polished, 5'-phosphorylated, and 3'-dATP added before genomic adapters were ligated onto the samples, following manufacturer's instructions (Illumina). A library was also generated for an input sample which was an aliquot of sonicated cell lysate with no subsequent immunoprecipitation step. The DNA libraries were amplified by PCR and sequenced using Illumina HiSeq 2000.

HUVEC for ChIP-seq following siRNA treatments were prepared together with Dr.

Claire Peghaire (postdoc in Randi's group) for H3K27ac and MED1 ChIP-seq. Human ERG was inhibited using 20 nM siRNA against ERG exon 6 (Qiagen; 5' - CAGATCCTACGCTATGGAGTA - 3') or 20 nM AllStars Negative Control siRNA (Qiagen) using AtuFECT01 lipid (1 µg/ml, Silence Therapeutics, Berlin, Germany). After 48 h of transfection HUVEC (10^7 cells) were crosslinked for 10 min with formaldehyde and snap-frozen. Subsequently the cell pellet samples were shipped to Active Motif for processing with a custom protocol which includes an immunoprecipitation using 4 µg anti-H3K27ac (cat no. 39133, Active Motif) and 5 µg anti-MED1 (A300-793A, Bethyl Laboratories) antibodies and high-throughput sequencing using Illumina NextSeq 500.

HUVEC for ERG ChIP-seq with Ang-1 treatment were prepared by Claire Peghaire. HUVEC were serum-starved in M199 media with 1 % BSA for 1h prior to stimulation with 250 ng/ml human modified Ang-1 (Regeneron Pharmaceuticals, Inc.). HUVEC were then collected as above and shipped to Active Motif for ERG ChIP using 8 µg anti-ERG (sc-354X, Santa Cruz Biotechnology).

ChIP-seq data mapping and peak calling

Raw fastq files for ERG, ENCODE / siCtl / siERG H3K27ac, siCtl / siERG MED1, H3K4me1 and all respective input datasets in HUVEC were aligned to the hg19 human reference genome using Bowtie 2 with default parameters (Langmead et al. 2012). Resulting SAM files were converted to BAM files using SAMtools (Li et al. 2009). Duplicate reads were marked by Picard from the Broad Institute and removed using SAMtools, along with mapping quality scores less than 5. I called peaks using MACS version 2.1.2 (Zhang et al. 2008) with input datasets as controls. To detect transcription factor binding sites I used a cutoff detection p-value of 1e-5 and for histone modifications I used a p-value of 10e-9.

ChIP-seq gene annotation and data visualisation

I assigned significant peaks to genes using Bioconductor package ChIPseeker in R

(Yu et al. 2015). ChIPseeker uses the proximity of the peak genomic location to the nearest gene. BigWig and BED files were visualised in the UCSC Genome Browser (https://genome.ucsc.edu).

Functional enrichment, functional annotation and motif analysis

GSEA software (v2.2.0) from the Broad Institute at MIT (Subramanian et al. 2005) was used to identify how different gene sets distribute in gene lists ranked by either microarray gene expression fold change values or by H3K27ac ChIP-seq enrichment on super-enhancers.

Biological functions were assigned to regulatory regions using Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al. 2010) with enriched gene ontology (GO) Biological Processes or using the collection from the Molecular Signatures Database (MSigDB) from Broad Institute (Liberzon et al. 2011) reported at a significance cutoff FDR < 0.001.

To plot correlations and heatmaps from ChIP-seq data I first normalised to each experimental input dataset and then I used the bamCompare tool in deepTools2.0 (Ramírez et al. 2016). I report and plot the data as $\log_2$ of the ratio between experiment and input. Resulting bigWig files were used to obtain a multiBigwigSummary in deepTools2.0 and these were used to plot correlations in R using the ggplot2 package. Scores in genomic regions of interest were calculated using computeMatrix to generate the heatmaps.

To find the most enriched motifs in a given ChIP-seq peak set the 'findMotifsGenome.pl' script from HOMER was performed (Heinz et al. 2010). For transcription factor binding sites the genomic interval $\pm$ 200bp from the centre of the binding site was used. The script adopts a set of sequences from the hg19 reference genome build as a background control that is matched in size and GC content. The significance of enriched motifs reported is reflected in the number and size of genomic regions assessed. To identify transcription factor motifs across the human reference genome hg19 for ERG,

cJUN and GATA2 I applied the 'scanMotifGenomeWide.pl' script from HOMER (Heinz et al. 2010).

Analysis of microarray data

Human transcriptome profiling of HUVEC following knockdown of ERG using microarray were obtained from Birdsey et al. (2012). Analysis of differential expression from the Affymetrix microarray datasets were performed in R version 3.2.4 using the Limma package (Ritchie et al. 2015). Briefly, the data was first normalised to account for technical variability and then a linear model is fit to the normalised gene expression data. A contrast matrix is formed to allow for all pairwise comparisons and an empirical Bayes framework is implemented to distinguish the larger variances in the genes that are truly differentially expressed. The significantly differentially expressed genes were assigned if the Benjamini and Hochberg adjusted p-value < 0.10. This allowed for a representational proportion of potentially meaningful gene expression changes to be validated.

Defining super-enhancer

Active enhancer regions were assigned by co-occurrence of H3K27ac and H3K4me1 ChIP-Seq enrichment in HUVEC using the ROSE algorithm to define super-enhancers (Lovén et al. 2013; Whyte et al. 2013). Active enhancers separated by less than 12.5 kb and located further than 2 kb from annotated promoters were concatenated. The ChIP-seq occupancy of H3K27ac in reads per million per bp (rpm/bp) were used to classify super-enhancers from typical enhancers by ranking stitched enhancers based on an increasing H3K27ac ChIP-seq signal. The inflection point provided a cut-off to define super-enhancers. An analogous procedure was used to define super-enhancer regions by enrichment of ERG at enhancers. Active enhancers were still assigned by H3K27ac/H3K4me1 co-occupancy for ERG super-enhancers. Super-enhancer associated genes were assigned as described earlier using ChIPseeker (Yu et al. 2015).

Differential ChIP-seq enrichment analysis

ChIP-seq alignment files (bam files) were converted to tag directories using the makeTagDirectory command in HOMER. Called H3K27ac peaks assigned using MACS (cutoff detection p-value of 1e-9) in the control siRNA (siCtl) sample were compared to ERG siRNA samples by getDifferentialPeaks to define loss (less density tags) or gain (more density tags) H3K27ac regions. Differential peaks were reported with $\log_2$ fold changes and a cumulative Poisson p-value < 0.0001. Similarly, this command was used to find differential super-enhancers in the control siRNA super-enhancer set relative to siERG. Volcano plots were produced with $\log_2$ fold changes and p-value cutoffs in R.

ChIP-qPCR

ChIP experiments were performed using the ChIP-IT express kit (Active Motif). HUVEC transfected with control or ERG siRNA were crosslinked with 1 % formaldehyde. Chromatin was sheared for four cycles of 30s on and 30s off using a Bioruptor ICD-200 ultrasound sonicator (Diagenode). Generated fragments of 200 - 1,000 bp were chromatin immunoprecipitated using 2 µg anti-H3K27ac (cat no. 39133, Active Motif) antibody. The respective negative controls were rabbit IgG (PP64, Chemicon, Millipore). RT-qPCR were performed using PerfeCTa SYBR Green Fastmix (Quanta Biosciences) on a Bio-Rad CFX96 system for selected genomic regions on immunoprecipitated DNA using primers listed in Table 2.2. Statistical significance was determined using a two-tailed paired sample t-test in R with a statistical significance cut-off p-value < 0.05. Data were plotted using Prism 8.0 (Graph Pad).

RT-qPCR primers used in this chapter are listed in Table 2.1.

Variant enrichment analysis

To perform GWAS SNP enrichment analysis, the R Bioconductor package traseR was used (Chen et al. 2016). SNP trait associations were obtained from dbGaP (Mailman et al. 2007) and NHGRI GWAS Catalog (MacArthur et al. 2017) combined. We also con-

Name	Orientation	Oligonucleotide sequences
Gene desert	Forward	TGAATAAGCCAATGAAACAATGACA
	Reverse	TGAAACATAGTATGGGTGGCAACT
VWF enhancer	Forward	AGGGGATTGGCCTCCTTTTA
	Reverse	CCATTTCCTTTCATTGTTTCC
CLDN5 enhancer (E1)	Forward	CCGGAAGCCAACTTGGAGTTT
	Reverse	GTGCAGAAGAATGCCCGGAA
CLDN5 E2	Forward	TCCTGCATCCCTGACCACTG
	Reverse	CTGGATGCTGCTCACATCGT
DLL4 E1	Forward	GCAGGTTGAGGGTGAATGGT
	Reverse	TGCCCAAGCACCAGAACTTT
DLL4 E2	Forward	CCCCAGGACCTATCCCAAGT
	Reverse	CACCATTTAGCAGAGCCGGA
DLL4 enhancer (E3)	Forward	GTTTCCTGCGGGTTATTTTT
	Reverse	CTTTCCAAAGGAGCGGAAT
DLL4 E4	Forward	CATGTGGGGGACAGGTAGGA
	Reverse	GCTCCCCATCTAGTGCATCA
IL6 E1	Forward	TGACTGAGCAAACCCATTTTCC
	Reverse	TCCTTATGTGGGAAGGTATGGC
IL6 E2	Forward	AGATTCCTCCACATTTGCCCA
	Reverse	GGCAACTCCAAGCCAGAACA
IL6 E3	Forward	GGTCACGCCACAACTGGAAT
	Reverse	CCATTCCTCACACCCACTGTT
PXN E1	Forward	GCCTCTCACCCCTGCTAATC
	Reverse	TTTGTTCGGGTCTCTGTGGG
PXN E2	Forward	GCATCACGTAGCAACAGAGC
	Reverse	GGTGTTGCTGACACATTCCG
PXN E3	Forward	GGTGGAGTAAAGCGTGAGCA
	Reverse	TGGGTGTAATGCCTGCCTTC

Table 2.1: Primer sequences used in this chapter for ChIP-qPCR.

sidered SNPs in LD from the 1000 Genomes Project ($r^2 > 0.8$, and located within 100kb of the lead SNP) giving a total of 78,247 unique trait-associated SNPs. Background SNPs were obtained from Utah residents with ancestry from Northern and Western Europe (CEU) population, available from the 1000 Genomes Project (Consortium 2015). These were used for hypothesis testing to assess the enrichment of trait-associated SNPs in ERG super-enhancers, ERG binding loci and ERG-bound enhancers in HUVEC. The

enrichment test follows the assumptions of the null hypothesis where the number of observed disease-associated SNPs out of background SNPs in ERG genomic intervals follows a binomial distribution with probability equal to the proportion of all genome-wide disease-associated SNPs out of background SNPs. Significance of enrichment was tested by binomial distribution. Analysis was performed for the enrichment of a select 10 trait class categories and all 573 available traits. A statistical p-value threshold $<$ 0.05 was used as a cutoff for significance. A random test set was created to assess the null hypothesis that random genomic intervals were significantly enriched for disease-associated SNPs. I generated chromosome and size matched genomic regions relative to ERG super-enhancers, ERG binding loci and ERG-bound enhancers. I obtained these random regions using shuffleBED from BEDTools (Quinlan et al. 2010). These regions were excluded from repressed (high H3K27me3) chromatin states in the HUVEC genome to provide a more rigorous test set.

Statistics

All statistical analyses were performed in R. Statistics associated with a particular analysis are mentioned above for each subsection. Fisher's exact test is performed across contingency tables when sample sizes were small (less than 1000) and a Chi-squared test for larger sample sizes (more than 1000). For data represented in a boxplot the sample distribution was tested for normality using a Shapiro-Wilk test. All data were found to be significantly not normally distributed and therefore a nonparametric test was used to test for significance between different groups. A Wilcoxon sum-ranks test was performed when comparing two groups. A Kruskal-Wallis test was used when comparing multiple groups (more than two) followed by post-hoc testing using the Wilcoxon sum-ranks test. Hypergeometric distribution testing to compare the overlap of Venn diagram regions was also performed in R.

Data

Data used in this chapter is listed in Table 2.2.

Experiment	Factor	Cell type	Condition	GEO ID	Citation / Distributor
ChIP-seq	ERG	HUVEC	untreated	GSM3557980	Kalna et al. 2019
DNaseI hyper-sensitivity		HUVEC	untreated	GSM816646	Thurman et al. 2012
ChIP-seq	cJUN	HUVEC	untreated	GSM935278	Linnemann et al. 2011
ChIP-seq	GATA2	HUVEC	untreated	GSM935347	Linnemann et al. 2011
ChIP-seq	H3K27ac	HUVEC	untreated	GSM733691	ENCODE/Broad Institute
ChIP-seq	H3K4me1	HUVEC	untreated	GSM733690	ENCODE/Broad Institute
ChIP-seq	H3K27me3	HUVEC	untreated	GSM733688	ENCODE/Broad Institute
ChIP-seq	H3K27ac	HUVEC	siCtl	GSM3557982	Kalna et al. 2019
ChIP-seq	H3K27ac	HUVEC	siERG	GSM3557983	Kalna et al. 2019
ChIP-seq	MED1	HUVEC	siCtl	GSM3557984	Kalna et al. 2019
ChIP-seq	MED1	HUVEC	siERG	GSM3557985	Kalna et al. 2019
ChIP-seq	ERG	HUVEC	PBS control	N/A	Unpublished
ChIP-seq	ERG	HUVEC	Ang1	N/A	Unpublished
Microarray		HUVEC	siCtl v siERG	GSE32984	Birdsey et al. 2012

Table 2.2: Data generated and used in this chapter.

2.3 Results

2.3.1 ERG binding and gene regulation in the HUVEC genome

ChIP-seq for ERG in HUVEC was performed by Dr Youwen Yang. Using this data I performed peak calling to identify 40,821 genomic ERG binding sites and by assigning peaks to genes I found that these sites were associated with 14,786 genes. The distribution of ERG binding sites across the genomic landscape was assessed by location. I found ERG bound promoters (± 2 kb from the TSS) accounting for 29.7 % of the total sites, yet many more distal regions were occupied (ie. bound by ERG) in both introns, 36.7 %, and intergenic sites, 30 % (Figure 2.1A). The classification of chromatin states was improved by using combinations of histone modifications from the HUVEC genome and showed that ERG was predominantly binding active enhancers, specifically 60 % of all ERG binding loci (Figure 2.1B).

By integrating the ERG ChIP-seq with transcriptome analysis of ERG-deficient HUVEC from a previously published microarray performed in the group by Dr Graeme Birdsey (Birdsey et al. 2008), I could determine the genes directly bound and regulated by ERG. Of genes activated by ERG, 1232 / 1454 were directly bound and 939 / 1180 of those repressed were directly bound (Figure 2.1C). There majority of sites, 12,615, were bound but were not directly regulating the assigned gene. The absence of regulation suggests functions of ERG other than direct gene regulation. This could include facilitating chromatin looping, co-activator assembly and pre-assembly for stimuli response. Gene ontology analysis of genes directly activated by ERG revealed functions previously known to be controlled by ERG such as “angiogenesis” and “wound healing” (Birdsey et al. 2012; Birdsey et al. 2015); similarly for directly repressed genes functions such as “SMAD protein import into the nucleus” and “TGFβ signalling pathway” are closely linked as SMAD proteins are effector signalling molecules that regulate gene transcription especially for TGFβ activation, a pathway tightly regulated by ERG (Dufton et al. 2017) (Figure 2.1D).

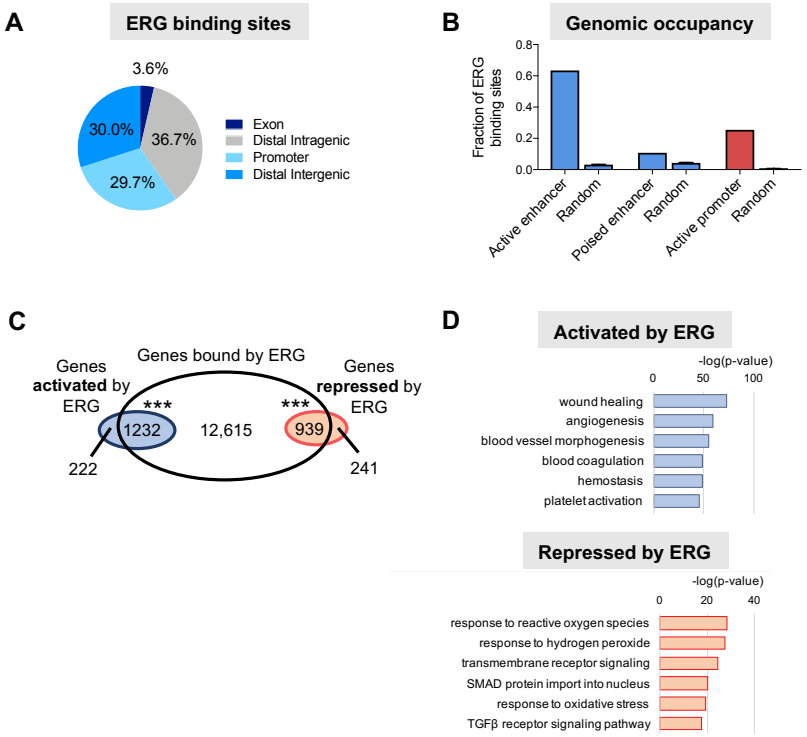


Figure 2.1: Genomic binding landscape and gene regulation by ERG in HUVEC. (A) Percentage distribution of ERG binding loci. Promoters are defined as ± 2 kb from the TSS, distal intragenic are sites located in introns and distal intergenic are sites located between genes. (B) Fraction of ERG binding sites at different chromatin states versus size-matched random regions. Active enhancers are defined by co-occupancy of H3K4me1 and H3K27ac; poised enhancers by H3K4me1 and H3K27me3 and active promoters by H3K4me3 and H3K27ac. (C) Overlap of ERG binding sites and genes regulated by ERG. The proportion of overlap of bound and both activated and repressed genes is significant, ***p-value < 0.001, Fisher's exact test. (D) Gene ontology analysis showing pathways associated with differentially regulated ERG-bound genes (activated genes in blue, repressed genes in red). Significance shown as -log(p-value).

2.3.2 Composition of ERG motif is specific to regulatory elements

The ETS binding motif is most commonly represented by the GGAA core sequence but the specific ERG binding motif is extended with flanking regions as identified from ChIP-seq in non-endothelial cells: (C/a/g)(A/C)GGAA(G/A) (Yu et al. 2010). By performing de novo motif analysis, I found that the ERG motif was most highly enriched at ERG binding sites (Figure 2.2A), as expected. Using the consensus ERG motif, I analysed the instances that ERG bound its transcription factor binding site (TFBS). I found the occurrence to be 2.5% (52,077 / 2,066,004 sites) which accounts for a small fraction of all potential ERG TFBSs. However, this is in line with other TFs in HUVEC: cJUN bound 2.2% of the total putative cJUN TFBSs (15,692 / 724,031) and GATA2 1.9% of the total putative GATA2 TFBSs (19,458 / 1,158,025). This small proportion of bound sites could partially be explained by some part of the chromatin in the genome being structurally inaccessible of TFs. If considering only the open regions of active chromatin in HUVEC, by co-occurence of DNaseI hypersensitivity and H3K27ac, 34.3% (38303 / 111596) of ERG TFBS are bound by ERG.

It has been reported that ERG binds the cytosine 5' of the core ETS motif GGAA (CGGAA) with significantly higher affinity than the 5' adenosine, AGGAA (Madison et al. 2018). These biochemical studies provided speculation on whether ERG was specifically utilising adenosine or cytosine in a particular chromatin or functional context. I assessed ERG binding sites by first quantifying the occurrence of AGGAA and CGGAA in different chromatin states defined by ChromHMM in HUVEC; active promoters, poised promoters, strong enhancers and weak enhancers. I found active promoter regions had approximately the same incidence of AGGAA as CGGAA when normalised for region length in kilobases. This was in stark contrast to enhancer regions where AGGAA was predominant and the CGGAA motif was less occurrent (Figure 2.2B). We also tested the hypothesis that the distribution of AGGAA and CGGAA directs the function of ERG as an activator or repressor of gene expression. However, I found that ERG binding sites associated with genes that are activated or repressed by ERG had

the same distribution of AGGAA and CGGAA sequences at ERG binding sites (Figure 2.2C), suggesting that gene regulation by ERG is independent of sequence composition, at least by the single nucleotide flanking the consensus ETS motif.

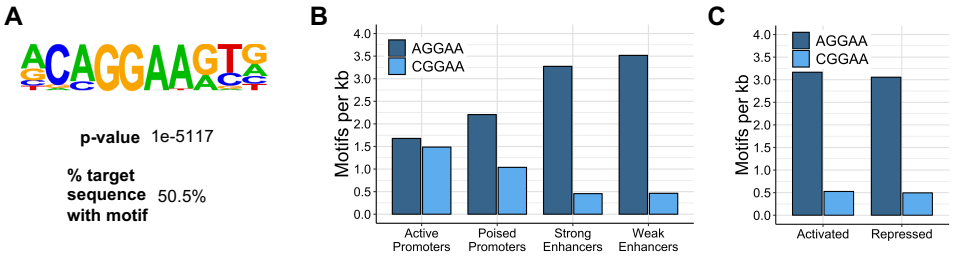


Figure 2.2: ERG motif occupancy in HUVEC. (A) Motif analysis +/-200bp from the centre of ERG binding sites. The core consensus sequence of ERG was the most enriched motif, p-value = 1e-5117 and occupied 50.5 % of ERG sites. (B) Occupancy of the AGGAA and CGGAA consensus sequence at ERG binding loci in specific chromatin regions. The CGGAA sequence is more common at promoters than enhancers were there is a much greater presence of the AGGAA motif. (C) Occupancy of AGGAA and CGGAA in genomic regions associated with genes activated or repressed by ERG. The AGGAA core sequence is predominantly occupied versus CGGAA.

2.3.3 Super-enhancers in endothelial cells

Super-enhancers have previously been characterised in HUVEC using H3K27ac (Hnisz et al. 2013) or BRD4 (Brown et al. 2014) as the ranking parameters. I used the ROSE algorithm as published (Hnisz et al. 2013; Lovén et al. 2013), and selected enhancer regions using a combination of H3K4me1 and H3K27ac at sites with DNaseI hypersensitivity signal in HUVEC. Identified enhancers were concatenated if within a distance of 12.5 kb, eliminating regions ± 2 kb from the TSS. Finally, these stitched enhancers were ranked according to the H3K27ac signal intensity. A schematic describing the protocol for identifying super-enhancers is shown in Figure 2.3A. I identified super-enhancer regions in HUVEC using H3K27ac ChIP-seq generated by the ENCODE/Broad Institute collaboration (ENCODE Project Consortium 2012). I found 917 super-enhancers associated with 822 genes, including key endothelial genes activated by ERG such as *VWF*,

CDH5 and *SOX17* (Figure 2.3B). These corroborated those reported by Hnisz et al. (2013); precisely a total of 86.3 % of super-enhancer regions reported here overlap with those published.

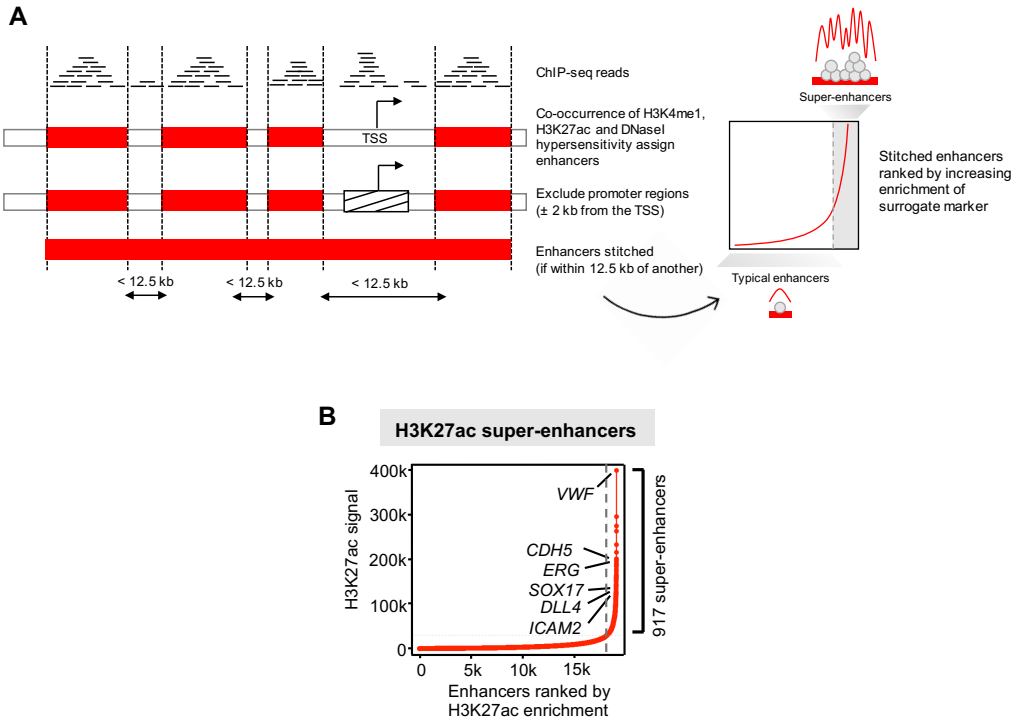


Figure 2.3: Identification of super-enhancers in endothelial cells. (A) Method used to identify super-enhancers. TSS; transcription start site. (B) Enhancer regions specified by H3K4me1 and H3K27ac co-occupancy and ranking of super-enhancers defined by H3K27ac enrichment. Super-enhancers and associated genes are indicated right of the grey dashed line.

By further exploring the characteristics of the identified super-enhancers, I integrated the expression profiles from ERG-deficient microarray analysis in HUVEC (Birdsey et al. 2012). By using the mean expression from the three control replicate samples I found that super-enhancer-associated genes have a significantly higher average gene expression in HUVEC compared to typical enhancer-associated genes (Figure 2.4A). It was postulated whether ERG gene regulation correlated with super-enhancer genes. Indeed, gene set enrichment analysis (GSEA) demonstrates a significant enrichment of ERG driven genes with the top 500 super-enhancer genes (Figure 2.4B). In addition, I found that ERG

bound the constituent enhancers of endothelial super-enhancers (177 binding sites / Mb), significantly more than typical enhancers (110 binding sites / Mb) (Figure 2.4C). The ERG motif was also significantly more occupied by ERG at super-enhancers than typical enhancers (35 % versus 25 %) (Figure 2.4D). However, interestingly the ERG motif occurred less frequently at super-enhancers than typical enhancers (0.63 versus 0.90 motifs / kb).

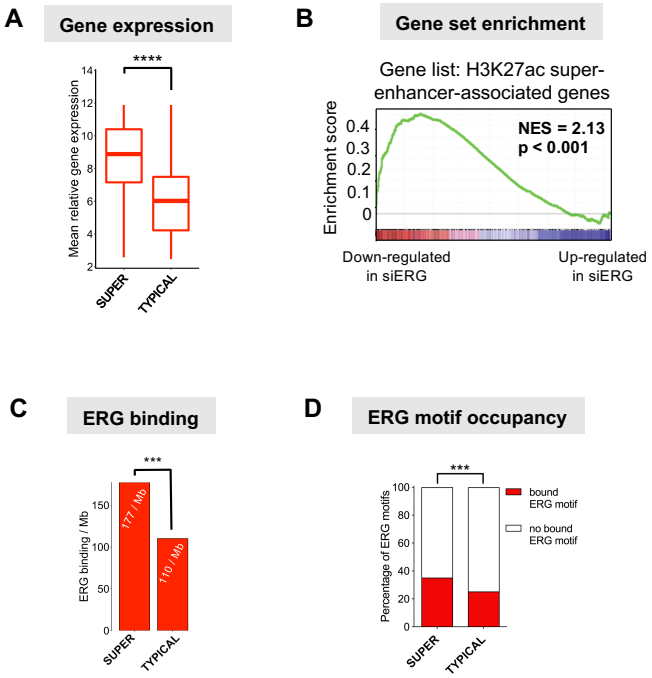


Figure 2.4: Characteristation of super-enhancers in endothelial cells. (A) Microarray transcriptome profiling in HUVEC shows relatively higher average gene expression in genes associated with super-enhancers than typical enhancers. (B) GSEA illustrates that the top 500 super-enhancer-associated genes are enriched in genes activated by ERG (ie. downregulated in ERG-deficient HUVEC). Normalised enrichment score (NES) = 2.13, p-value < 0.001. (C) ERG binding is significantly higher at constituent enhancers of super-enhancers than typical enhancers; ***p-value < 0.001, Pearson's Chi-squared test. (D) The occupancy of the ERG motif by ERG is significantly greater in super-enhancer than typical enhancer regions; ***p-value < 0.001, Pearson's Chi-squared test.

Next, given that enhancers and super-enhancers have previously been distinguished by transcriptional co-activators such as Mediator subunit 1 (MED1) and p300 (Shin et al. 2016; Vahedi et al. 2015; Zhang et al. 2013), Dr. Claire Peghaire (postdoc in Randi's group) and I prepared HUVEC for ChIP-seq of Mediator subunit 1 (MED1). The chromatin clean-up, immunoprecipitation and sequencing was performed by company Active Motif. I used ChIP-seq signals from H3K27ac and MED1 to determine the genome-wide correlation and demonstrate that the enrichment of these signals significantly aligns (p-value < 0.0001) (Figure 2.5A). Such a significant correlation implies that MED1 is a likely candidate for super-enhancer identification in HUVEC. I assigned constituent enhancer regions using H3K4me1 and H3K27ac and used the signal of MED1 to assign 893 super-enhancers (Figure 2.5B). When compared to super-enhancer regions defined by H3K27ac, I observed a 49 % overlap, a subset of super-enhancers that may potentially be more important since they are defined by multiple super-enhancer marks.

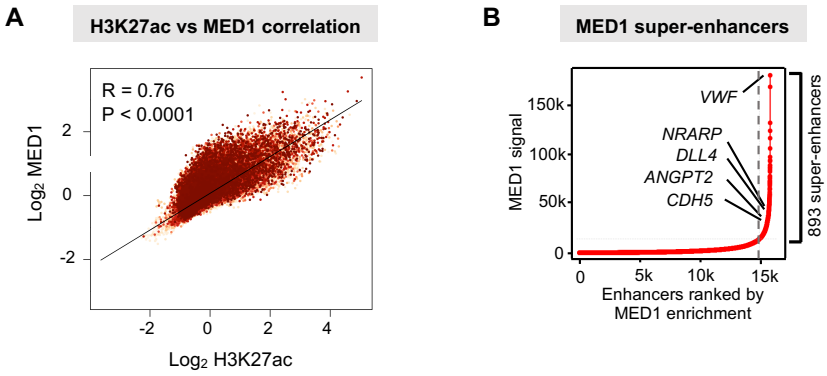


Figure 2.5: Super-enhancers defined by MED1 occupancy. (A) Scatterplot of the positive correlation between H3K27ac and MED1 occupancy ($\log_2$ transformed). Pearson correlation coefficient = 0.76, p-value < 0.0001. Each point represents the average ChIP-seq signal in a 50 bp window of the genome. (B) Super-enhancers defined by MED1 signal across stitched enhancer regions. MED1 super-enhancer-associated genes are similar to those identified by H3K27ac enrichment.

2.3.4 ERG defined super-enhancers in endothelial cells

To assess the relationship between ERG and super-enhancers I used the ChIP-seq signal for H3K27ac and ERG to unveil a significant positive correlation between the two at super-enhancers defined by H3K27ac (Figure 2.6A). I therefore used ERG as the ranking parameter and defined 1125 ERG super-enhancers (Figure 2.6B). By performing a GSEA with all super-enhancer-associated genes identified by ERG I found a strong positive correlation with H3K27ac-super-enhancer genes (Figure 2.6C). Gene ontology analysis performed in Genomic Regions Enrichment of Annotations Tool (GREAT) revealed mechanisms such as angiogenesis, blood vessel morphogenesis and response to wounding for both H3K27ac and ERG super-enhancers (Figure 2.6D). These results confirm the endothelial functions and characteristics of the super-enhancers defined by both H3K27ac and ERG itself.

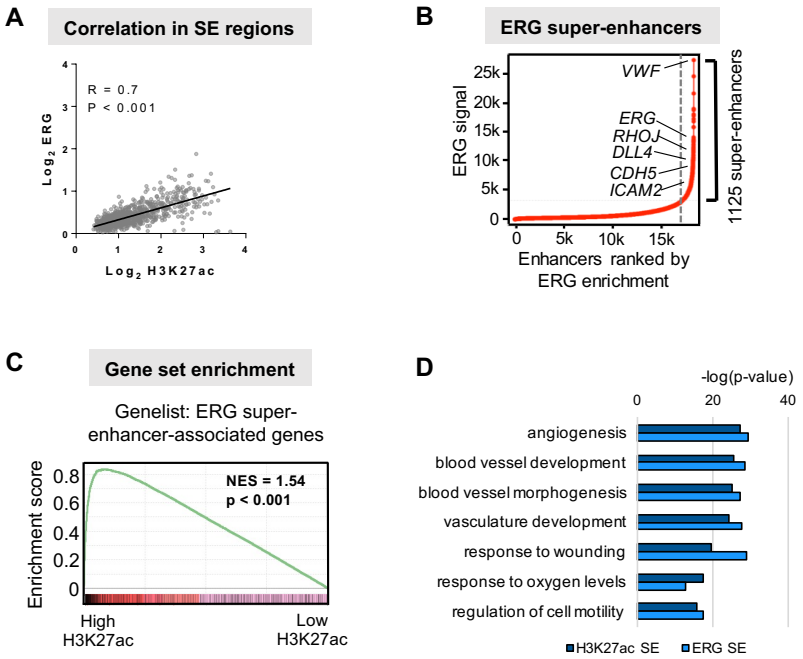


Figure 2.6: Super-enhancers defined by ERG in HUVEC. (A) Scatterplot of the correlation between ERG and H3K27ac occupancy ($\log_2$ transformed) in the 917 super-enhancer regions. Each grey point is the average ChIP-seq signal from one super-enhancer. (B) Super-enhancers ranked by ERG enrichment. Known ERG-target genes proximal to super-enhancers are indicated. (C) GSEA of the top 500 genes associated with ERG-super-enhancers compared with the 917 H3K27ac-super-enhancers. Normalised enrichment score (NES)=1.54, p-value < 0.001. (D) Analysis of gene ontology of both H3K27ac-super-enhancers and ERG-super-enhancers. Significance indicated as -log(p-value).

2.3.5 ERG modulates H3K27ac levels

It has been reported that ERG can recruit the histone acetyltransferase p300 to the DNA (Fish et al. 2017). The co-activator p300 can modify chromatin structure through its intrinsic histone acetyltransferase activity causing the opening of chromatin. To investigate whether ERG was responsible for the recruitment of p300 to deposit the acetyl mark on H3K27 a ChIP-seq for H3K27ac was performed in control and ERG-depleted HUVEC. The samples were prepared together with Dr. Claire Peghaire in the group and the fixed chromatin was shipped to Active Motif for chromatin clean-up, immunoprecipitation and high-throughput sequencing. The depletion of ERG was performed using a siRNA targeted to ERG (siERG) and in a parallel samples were treated with control siRNA (siCtl). Total RNA and protein levels for ERG were significantly reduced (Figure 2.7A and B). In control HUVEC, 56,347 H3K27ac regions were identified in comparison to 47,575 in siERG-treated HUVEC. The H3K27ac ChIP-seq in basal HUVEC produced by the ENCODE/Broad Institute collaboration (ENCODE Project Consortium 2012) correlated significantly with the H3K27ac ChIP-seq data from the siCtl sample generated in this study (Figure 2.7C).

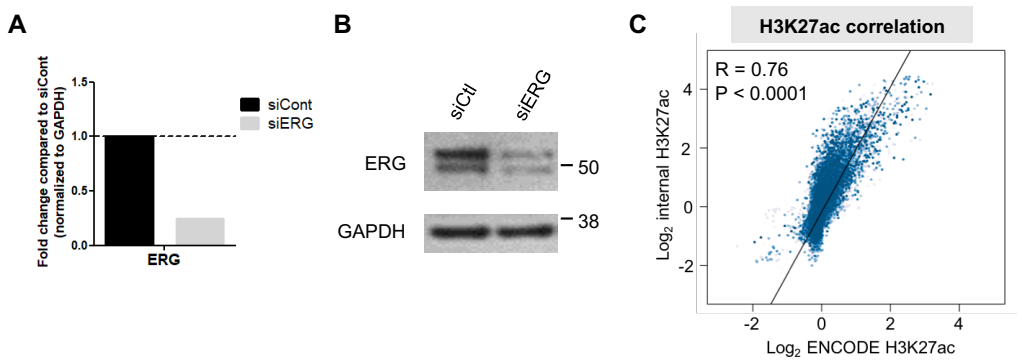


Figure 2.7: ChIP-seq data for H3K27ac in control and ERG-siRNA HUVEC. (A) Validation of ERG total RNA levels following siRNA knockdown of ERG in HUVEC in samples used for H3K27ac and MED1 ChIP-seq. RT-qPCR results shown normalised to GAPDH in siCtl and siERG samples. (B) Validation of ERG protein knockdown in samples used for H3K27ac and MED1 ChIP-seq by immunoblotting in siCtl and siERG samples. GAPDH is the loading control. (C) Correlation of $\log_2$-normalised H3K27ac ChIP-seq data produced by the ENCODE/Broad Institute versus the H3K27ac from the siCtl sample. Pearson correlation coefficient = 0.76, p-value < 0.0001. Each point represents the average ChIP-seq signal in a 50 bp window of the genome.

I performed differential binding analysis to determine which H3K27ac sites showed a significant increase or decrease in acetylation upon the downregulation of ERG. I performed the analysis using all control H3K27ac regions was found H3K27ac levels were modulated globally; 5277 regions were decreased (loss) and 1648 regions were increased (gain) in ERG-deficient cells (Figure 2.8A and B). Enhancer regions identified to be decreased by siERG treatment in the ChIP-seq were validated by ChIP-qPCR. Enhancer regions linked to endothelial genes *CLDN5*, *DLL4* and *VWF* were significantly reduced in siERG HUVEC in comparison to a control gene desert region that is not enriched in either the siCtl or siERG samples (Figure 2.8C).

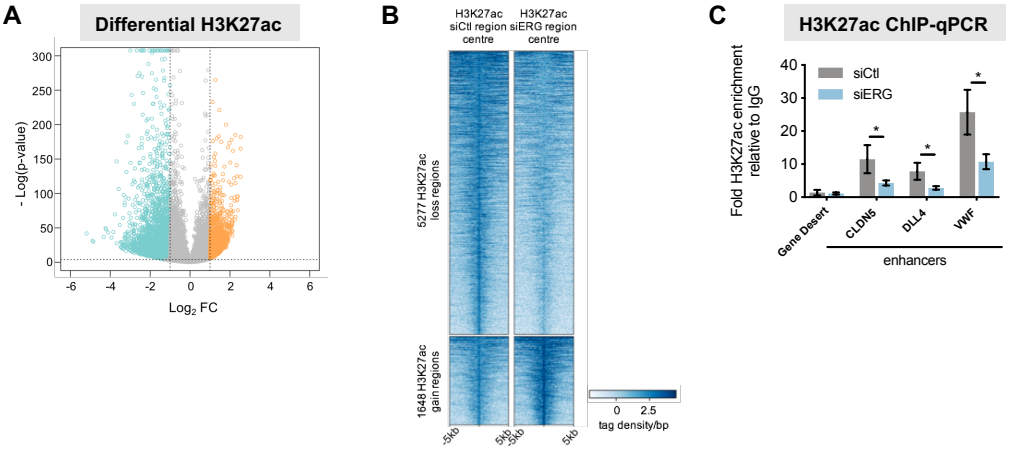


Figure 2.8: Regulation of histone acetylation at enhancers by ERG. (A) Volcano plot of differential H3K27ac analysis. Loss and gain enhancer regions are selected by -1 $\geq$ $\log_2$FC $\geq$ 1; -log(p-value) $>$ 4. (B) Heatmap illustrating the H3K27ac signal normalised to input in siCtl and siERG samples in loss and gain regions. Heatmap is centred on each region $\pm$ 5 kb showing the tag density per base pair. (C) ChIP-qPCR validation for H3K27ac on enhancers of *CLDN5*, *DLL4* and *VWF* with a negative control gene desert region in HUVEC treated with siCtl or siERG. Fold change over IgG presented, n=3, *p-value $<$ 0.05, paired 2-tailed *t*-test.

The functional interdependence between H3K27ac and ERG was investigated by comparing the differentiation in acetylation levels of H3K27 with changes in gene expression by ERG. I used all H3K27ac sites that were significantly modulated by -2 $\geq$ $\log_2$FC $\geq$ 2 and associated them with nearby genes. Gene expression changes with an adjusted p-value $<$ 0.10 were used. There were 335 genomic H3K27ac loci that showed coherent downregulation in acetylation levels and downregulation in gene expression in response to ERG knockdown, yet only 60 loci that showed significant upregulated in acetylation and gene expression (Figure 2.9A). Similarly, genes associated with H3K27ac loss regions were significantly downregulated in comparison to genes where H3K27ac remained constant or showed a gain in levels (Figure 2.9B). Genes in loss and gain groups were analysed on the basis of Gene Ontology (GO) terms using GREAT (Figure 2.9C). The most prominent terms included Notch and integrin-linked kinase signalling in the loss group and TGFβ and BMP receptor signalling in the gain group (Figure 2.9C). This is in line with ERG's positive regulation of target genes in the Notch signalling pathway

(Shah et al. 2017) and ERG-dependent inhibition of canonical TGFβ-SMAD signalling in EC (Dufton et al. 2017).

To investigate whether basal transcriptional machinery was affected upon ERG depletion, a ChIP-seq experiment for MED1 in HUVEC treated as above with either siCtl and siERG was performed. I compared whether the changes in MED1 occupancy were in line with those changes in H3K27ac in response to ERG knockdown. I found that loss and gain in MED1 enrichment coincided with a loss and gain in H3K27ac (Figure 2.9D). This is indicative of either a causative effect of ERG on MED1 occupancy or an indirect effect such as the changes to H3K27ac at distinct genomic loci.

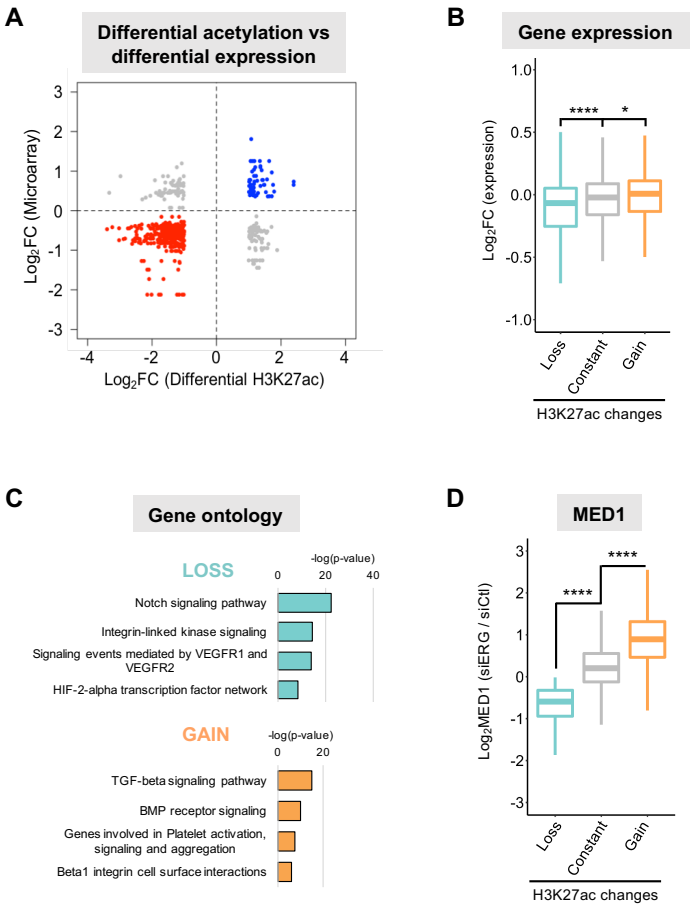


Figure 2.9: ERG-dependent H3K27ac changes associate with transcriptional and functional changes. (A) Scatter plot showing the correlation between differential H3K27ac and differential gene expression. Genes associated with a decrease in H3K27ac and decrease in gene expression are shown in red; conversely genes with an increase in H3K27ac and gene expression are represented in blue. (B) Boxplot demonstrating the distribution of $\log_2$FC values of genes associated with a loss, constant or gain H3K27ac region. P-value < 0.0001, Kruskal-Wallis test and ****p-value $<$ 0.0001; *p-value < 0.05 for post hoc test using Wilcoxon rank-sum test. (C) Pathway analysis of gene ontology terms significantly enriched in loss and gain groups. Top 4 significant terms shown. (D) Boxplot illustrating the distribution of $\log_2$FC in differential MED1 occupancy in loss, constant and gain regions. P-value < 0.0001, Kruskal-Wallis test and ****p-value < 0.0001 with post hoc test using Wilcoxon rank-sum test.

2.3.6 ERG controls a core subset of super-enhancers

The MED1 subunit of the large transcriptional complex Mediator is enriched at cell-type specific super-enhancers. To determine the effect of ERG on the assembly of super-enhancers, I used the ROSE algorithm to define super-enhancers in control and ERG-depleted cells. H3K27ac ChIP-seq analysis in control HUVEC identified 1015 super-enhancer clusters (Figure 2.10A). This subset was representative of the H3K27ac super-enhancers identified using H3K27ac generated by ENCODE in Figure 2.3B. Depletion of ERG using siRNA resulted in a redistribution of HUVEC super-enhancers, associated with several important endothelial genes (Figure 2.10B).

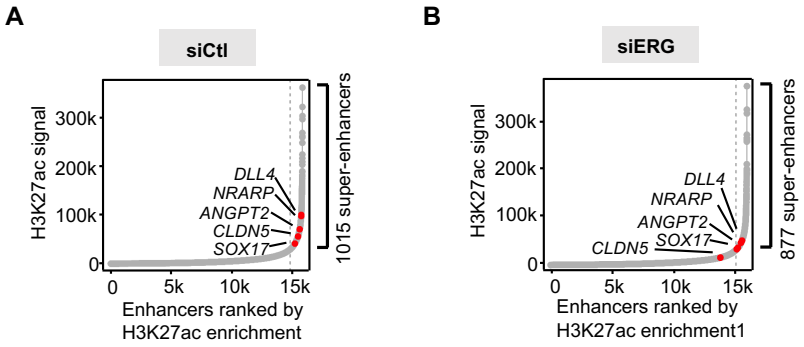


Figure 2.10: H3K27ac-defined super-enhancer distribution is controlled by ERG. (A) Super-enhancer identification in siCtl (left) and siERG (right) HUVEC using H3K27ac. Selected genes important to endothelial function are highlighted in red to indicate their decrease in ranking from siCtl to siERG.

To quantify these changes, I performed differential binding analysis for H3K27ac comparing siCtl to siERG treated cells. Using all 1015 control super-enhancers as regions of interest, I determined the fold change in enrichment between control and ERG-depleted. I found 107 super-enhancers that were significantly downregulated or "decreased" ($\log_2$FC $\leq$ -1.5, -log(p-value) $>$ 4) (Figure 2.11A). These included endothelial genes *CLDN5*, *DLL4*, *NOTCH4* and *NRARP*. In contrast, only 14 super-enhancers were found to be upregulated or "increased" ($\log_2$FC $\geq$ 1.5, -log(p-value) $>$ 4). The vast majority of super-enhancers showed no significant changes following siERG treatment

and were therefore considered and referred to as "constant" (Figure 2.11A).

Comparing the decreased and constant super-enhancer groups I found that changes in MED1 occupancy between control and ERG-siRNA HUVEC were significantly reduced in the decreased group relative to the constant super-enhancer group (Figure 2.11B). I also found that ERG-dependent changes in super-enhancer-associated genes correlated with changes in ERG-regulated genes (Figure 2.11C). Thus, ERG is independently driving a discrete subset of super-enhancers in mature endothelial cells.

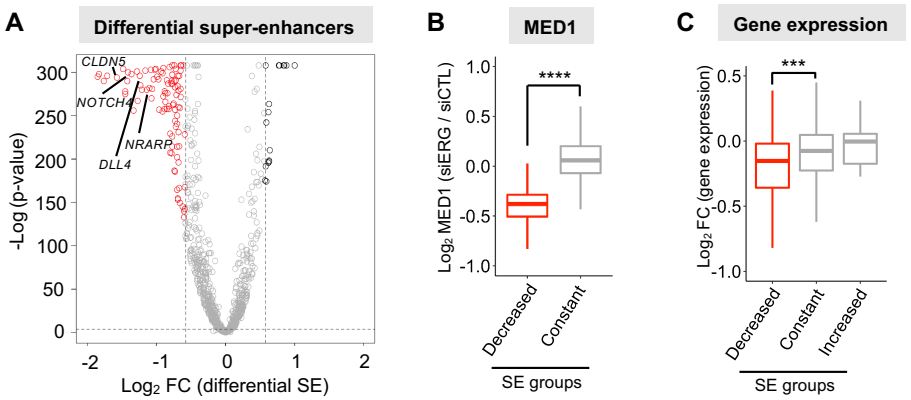


Figure 2.11: ERG regulates a core subset of endothelial super-enhancers. (A) Differential super-enhancers following analysis of changes of H3K27ac levels at the 1015 siCtl super-enhancers. Significantly up or downregulated SEs are selected according to -0.58 $\geq$ $\log_2$FC $\geq$ 0.58 (-1.5 $\geq$ $\log_2$FC $\geq$ 1.5); -log(p-value) $>$ 4. (B) Boxplot representing the distribution of $\log_2$FC in differential MED1 occupancy in decreased and constant super-enhancers. ****p-value $<$ 0.0001 with post hoc test using Wilcoxon rank-sum test. (C) Boxplot showing the distribution of $\log_2$FC values of genes associated with a decreased, constant or increased super-enhancer region. P-value $<$ 0.0001, Kruskal-Wallis test and ***p-value $<$ 0.001 for post hoc test using Wilcoxon rank-sum test.

2.3.7 Prosurvival factor Angiopoietin 1 (Ang-1) enhances ERG genomic occupancy

Proangiogenic factor Angiopoietin 1 (Ang-1) has been shown to be important in blood vessel stability and maintaining vascular maturity. Our group have reported that Ang-1 promotes ERG phosphorylation which potentiates ERG binding at the promoter and

enhancers of the *DLL4* gene, increasing recruitment of the transcriptional co-regulator β-catenin and *DLL4* transcription (Shah et al. 2017).

To determine the effect of Ang-1 on ERG global enrichment Dr. Claire Peghaire prepared and treated HUVEC with control PBS or soluble Ang-1 (250 ng/ml) for 1 hour. Samples were collected and subsequently processed and sequenced by Active Motif. I analysed the output sequencing data and assessed the changes in ERG binding using differential binding analysis at the ERG binding sites in the control PBS treated sample. I found that many ERG sites are enriched following stimulation by Ang-1; 3973 ERG-bound regions are increased, 11817 remain unchanged and only 51 sites decreased ($-0.58 \geq \log_2 FC \geq 0.58$ ($-1.5 \geq \log_2 FC \geq 1.5$)) (Figure 2.12A).

Having previously shown that ERG can be used as a surrogate marker for super-enhancer identification (Figure 2.6B), here I used ERG binding loci to assign enhancers and ERG signal to denote super-enhancers in control PBS and Ang-1 treated HUVEC (Figure 2.12B). The analysis determined that the total number of stitched ERG regions was 5914 in control and 8954 in Ang-1 treated cells. This demonstrates that Ang-1 induced ERG binding at *de novo* sites. Additionally, the ERG ChIP-seq signal is greater following Ang-1 activation indicating an increase in ERG binding. A snapshot of genomic regions with increased ERG occupancy following Ang-1 stimulation are shown at *EGFL7*, *CDH5* and *SOX18* loci (Figure 2.12C). Thus, Ang-1 potentiates ERG occupancy at the genomic level potentially by inducing ERG phosphorylation that enhances its DNA binding activity (Shah et al. 2017).

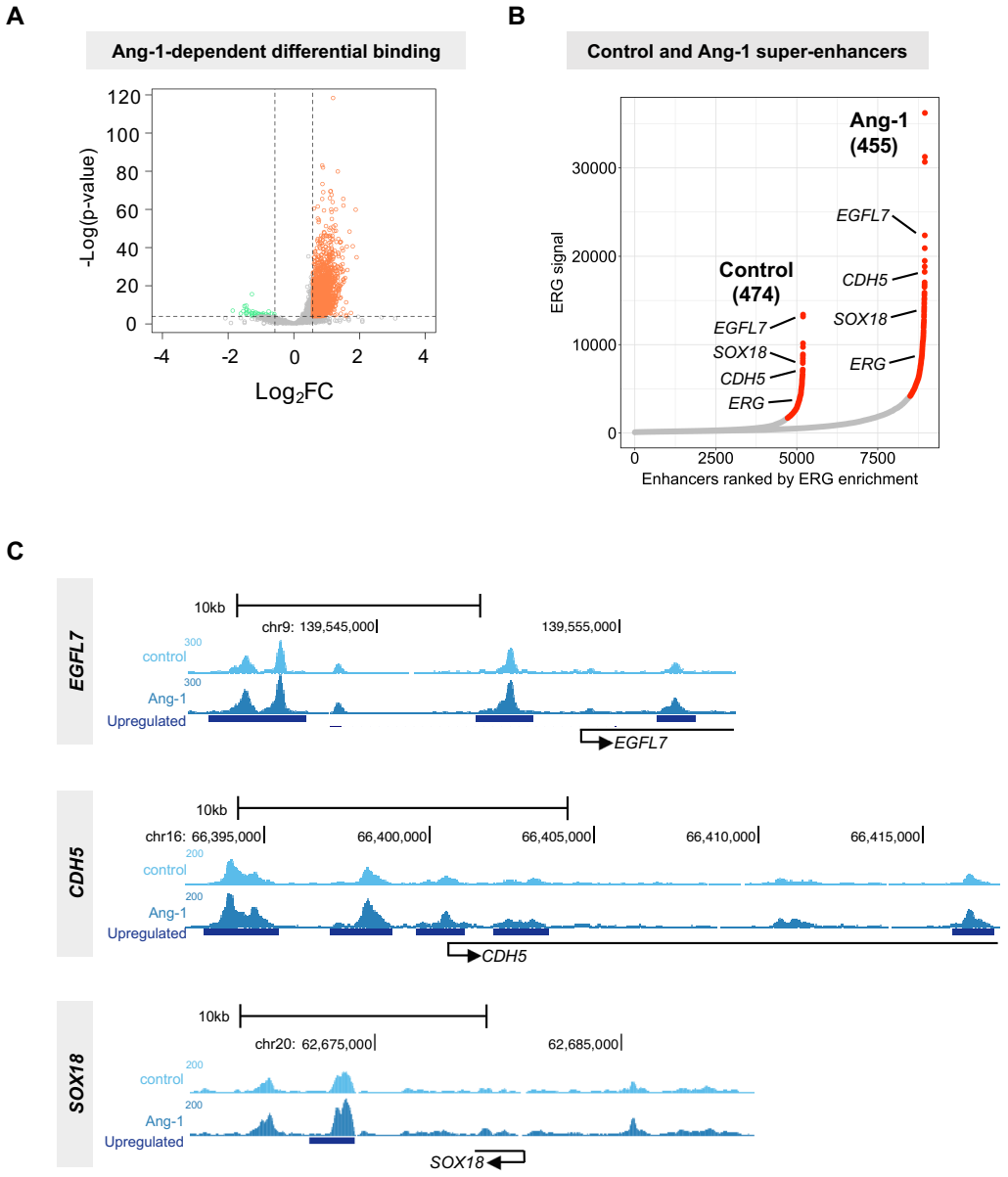


Figure 2.12: Ang-1 stimulation induces ERG occupancy genome-wide. Ang-1 stimulates ERG genome binding globally (A) Volcano plot illustrating the $\log_2$FC in ERG occupancy at PBS control ERG binding regions. Increased regions are highlighted in orange and decreased in green (-0.58 $\geq \log_2$FC $\geq$ 0.58 (-1.5 $\geq \log_2$FC $\geq$ 1.5)). (B) Direct comparison of super-enhancers identified for PBS control and Ang-1 stimulated HUVEC. Total number of super-enhancers noted in brackets. Position of selected endothelial genes indicated for control PBS and Ang-1 treatment. (C) Representative ERG binding loci in PBS and Ang-1 samples with sites of significant ERG binding enrichment indicated (upregulated; blue bars).

2.3.8 ERG genomic loci prioritise cardiovascular and other disease single nucleotide polymorphisms

Disease-associated single nucleotide polymorphisms (SNPs) discovered by genome-wide association studies (GWAS) have been demonstrated to be preferentially enriched at super-enhancer clusters of disease-relevant cell types (Hnisz et al. 2013; Pasquali et al. 2014). It is known that endothelial dysfunction is linked to a number of diseases including cardiovascular diseases (CVD), diabetes, chronic kidney failure and tumour metastasis (Rajendran et al. 2013). I obtained all trait-associated SNPs from the NCBI dbGaP (Mailman et al. 2007) and NHGRI-GWAS (MacArthur et al. 2017) catalogs including SNPs in LD from the 1000 Genomes Project. In total this accounts for 78,247 unique trait-associated SNPs. I examined the enrichment of these SNPs at ERG super-enhancers, ERG-bound enhancers and ERG binding loci. To determine the enrichment of specific disease-associated SNPs, I performed a binomial distribution test by comparing to a subset of background variants obtained from a population of Utah residents with ancestry from Northern and Western Europe (CEU), available from the 1000 Genomes Project (Consortium 2015).

I found that ERG super-enhancers were most enriched with SNPs associated with CVD (p-value = 1.1 x 10^{-14}; Figure 2.13A). Interestingly, ERG super-enhancers were also enriched in SNPs for digestive system (p-value = 5.7 x 10^{-5}), respiratory tract and nervous system diseases (p-value = 6.1 x 10^{-5} and 1.7 x 10^{-3} respectively; Figure 2.13A). This enrichment was not identified in size- and chromosome-matched regions randomly shuffled (permuted) across permissive, open chromatin. Further interrogation of the CVD category revealed strong enrichment for specific CVD traits including abdominal aortic aneurysm (p-value = 1.7 x 10^{-13}), coronary disease (p-value = 3.7 x 10^{-13}), myocradial infarction (p-value = 5.1 x 10^{-9}) and hypertension (p-value = 1.1 x 10^{-8}) but not heart failure (Figure 2.13B). This suggests that CVD variants associated with diseases that have an immediate impact from the vasculature are more prevalent.

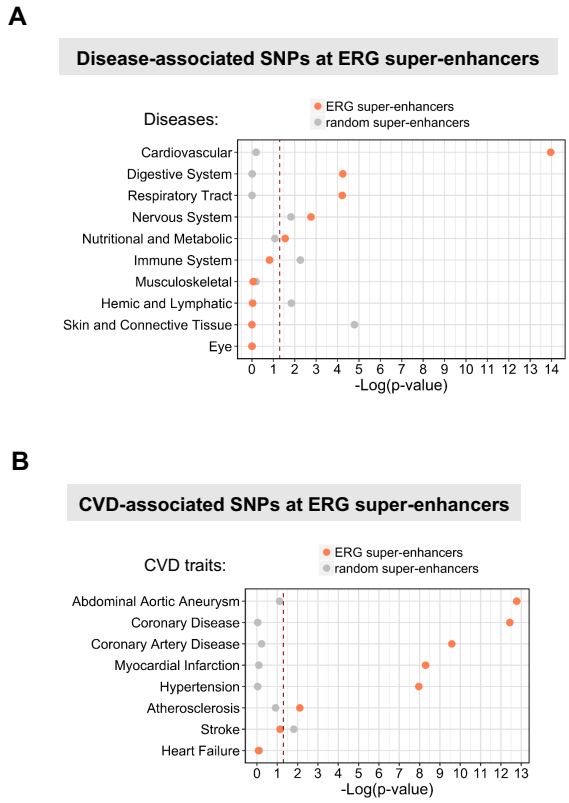


Figure 2.13: ERG super-enhancer regions prioritise CVD variants. (A) Significance of overlap between SNPs associated with a number of disease traits and ERG super-enhancers (orange) and chromosome- and size-matched random super-enhancers (grey). Significance of enrichment is calculated by binomial distribution test. (B) Significance of enrichment by binomial distribution of the overlap between SNPs associated with specific CVD traits and ERG super-enhancer and chromosome- and size-matched random super-enhancer regions (grey). Both plots show significance as -log(p-value) and the red line in all plots represents p-value < 0.05.

The analysis was also restricted to ERG bound loci and ERG-bound enhancer regions. Both ERG sites and ERG-bound enhancers showed highest enrichment for SNPs associated with immune system diseases and CVD (Figure 2.14A and B). Many other variants associated with diseases that have a vascular component were significantly enriched including skin and connective tissue, eye and respiratory tract diseases (Figure 2.14A and B). Collectively, it is evident that variants in active and ERG-occupied regions of the endothelial genome are potentially contributing to a range of complex diseases.

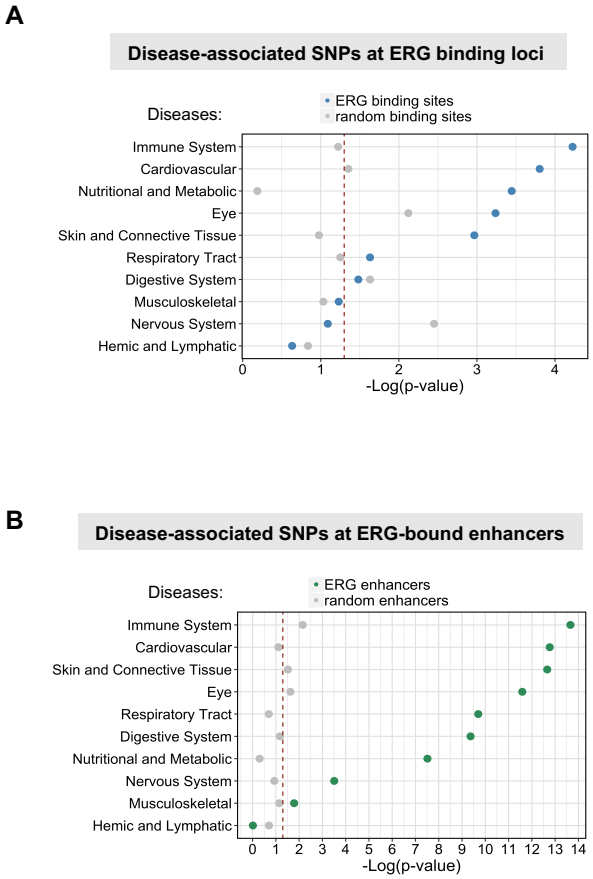


Figure 2.14: ERG-centric regions are enriched for variants with a vascular component. (A) Overlap of disease-associated SNPs and ERG binding sites (blue) with chromosome- and size-matched enhancer regions (grey). (B) Overlap of disease-associated SNPs and ERG-bound enhancers (green) with chromosome- and size-matched enhancer regions (grey). Both plots show significance as -log(p-value) and the red line in all plots represents p-value < 0.05.

2.4 Discussion

In this chapter, I investigate the role of ERG DNA binding and find it to occupy promoters and, predominantly, enhancer regions in endothelial cells which are in an active state. The ETS motif sequence preferentially used by ERG is the core ETS motif flanked by an adenosine (AGGAA). Comparing to the sequence which is flanked by a cytosine (CGGAA) there is a stronger preference for ERG binding to AGGAA at enhancers with less preference for A or C observed at promoter regions (Figure 2.2B). It has been determined that the affinity of ERG binding to CGGAA is 10-fold higher than for the alternate AGGAA sequence (Madison et al. 2018), suggesting that ERG may have stronger binding affinity at promoters potentially for robust regulation of non-specific housekeeping genes.

I identify and characterise endothelial super-enhancers using a number of different surrogate markers. Occupancy of histone modification H3K27ac, transcriptional co-activator MED1 and transcription factor ERG all determine super-enhancers linked to endothelial-specific genes which is indicative of high-density assembly at these regions. The enhancer sites are found to be modulated by ERG depletion and specifically the majority of changes show decreased acetylation indicating ERG is a positive regulator of endothelial-specific enhancers. Importantly, I demonstrate that ERG independently modulates a restricted subset of H3K27ac super-enhancers in EC whilst the majority are not largely affected to ERG depletion. MED1 occupancy was also regulated by ERG in a similar manner as by that observed for H3K27ac histone modification. These changes in lineage-defining super-enhancers were accompanied by changes in ERG-dependent gene expression.

It has been disputed whether enhancers within super-enhancers are exceptional to regular typical enhancers. Here ERG bound more super than typical enhancers and furthermore utilised (ie. bound) the ERG motif sequence more frequently in super-enhancers. However, super-enhancer regions in fact showed a lower frequency of the

ERG motif per kilobase than typical enhancers. Thus, the DNA sequence itself does not discriminate between super and typical enhancers. It is possible that the higher-order structure of chromatin could be a determinant to the localisation of super-enhancers. Recently, formation of phase-separated molecular condensates have been described at super-enhancers (Sabari et al. 2018). These liquid-like condensates are dictated by physiochemical properties whereby the intrinsically disordered regions of co-transcriptional activators (eg. MED1) form phase-separated droplets that are more favourable for transcriptional complex assembly, such as ERG occupancy at endothelial super-enhancers as opposed to typical enhancer regions.

It was surprising to find that many H3K27ac super-enhancers were unaffected in response to the loss of a critical endothelial TF. It is possible that many super-enhancers did not collapse due to the synergy between individual enhancers where more distant ones possibly compensate for the loss of nearby constituent enhancers. This highlights the likely important and non-redundant function of typical enhancers which may be functionally critical in response to genomic and epigenomic insults. In fact, ERG has been implicated in structural chromatin organisation and looping in prostate cells where ERG was overexpressed (Rickman et al. 2012) and in prostate cancer VCaP cells which express ERG due to a chromosomal fusion (Zhang et al. 2019). Overall, higher-order 3D structural conformation of the endothelial genome using H3K27ac HiChIP or promoter-enhancer capture would reveal super-enhancer organisation to a high resolution. Such methods have been adopted in studies using other cell types to uncover chromatin architecture mediating cell identity (Miguel-Escalada et al. 2019; Mumbach et al. 2017). Recently, Nagai et al. (2018) observed the effect of knocking-down ERG in combination with its closest homolog FLI1 and reported a much greater decrease in H3K27ac levels. This supports the notion that other TFs, particularly those that are also part of the ETS family, are involved in cooperative assembly at endothelial enhancers and are functionally compensating for ERG at ETS sites left unoccupied.

Super-enhancers colocalize Mediator and cell type-specific TFs. However, as a cen-

tral integrator of transcription, Mediator complex remains understudied in mammalian cells with no reports of its activity in EC. In other cell types Mediator has been shown to interact with transcriptional and epigenetic regulators including RNA Pol II, cohesin and EZH2 (Fukasawa et al. 2015; Yin et al. 2014). Co-regulators are more ubiquitously expressed and therefore interaction between these and Mediator could be extrapolated to other cell types; however, interactions with cell type-specific TFs require further investigation. The limited studies that have observed interactions between Mediator and TFs that are ESC-specific master factors: Oct4, Sox2 and Nanog. Further afield, Mediator interactions have been verified for ETS factor Pu.1 in pro-B cells, MyoD in myotubes and C/EBPα in macrophages (Whyte et al. 2013). The potential link between the endothelial master TF ERG and non-DNA binding Mediator may shed light into further endothelial chromatin and transcriptional functions. Interestingly, the two proteins share some common functions which include DNA loop formation, interaction with super-enhancer mark BRD4 and general regulation of cell identity (Jang et al. 2005; Whyte et al. 2013; Yin et al. 2014). In this chapter, Mediator complex is mapped in the HUVEC genome via its subunit MED1 in control and ERG-depleted cells with MED1 levels found to be modulated in an ERG-dependent manner. Further work could elucidate the endothelial-specific interactions and functions of the Mediator complex with ERG and other key endothelial regulators. The Mediator interactome was recently deciphered in neural stem cells and revealed a plethora of novel interactions that suggested high enrichment of TFs and co-activators including other complex proteins corroborating the higher binding density at super-enhancers (Quevedo et al. 2019).

ERG was shown to modulate acetylation of H3K27 and in non-ECs has been considered a master TF (Kron et al. 2017) and in other context a pioneer TF (Cai et al. 2013). A pioneer TF has the unique ability to open closed chromatin domains providing access for further TF binding (Heintzman et al. 2007; Zaret et al. 2011). The pioneer TF in endothelial cells is postulated to be ETV2 with several studies demonstrating ETV2 as necessary for EC differentiation. Ginsberg et al. (2012) converted amniotic cells to EC

by transient ETV2 expression with TGFβ inhibition and constitutive ERG and FLI1 expression, Morita et al. (2015) directly converted human adult skin fibroblasts to EC using ETV2 and finally Kanki et al. (2017) used VEGF to convert mouse embryonic stem cells (mESC) into endothelial cells and reported that Etv2 expression peaked early and transiently before rapidly declining. With ETV2 and ERG both classified as ETS TFs there is a possibility that physiologically ETS sequence motifs are progressively unoccupied by ETV2 and consequently adopted by ERG during EC maturation.

The rapid and potent effects of external stimuli on EC has been investigated using growth factors, in particular the central regulator of angiogenesis VEGF. Kaikkonen et al. (2014) identified genomic regions regulated by VEGF and found these to be mediated by lineage-defining TFs. Moreover these regions are coordinated in an organised structure whereby active topologically associated domains (TADs) across chromosomes interact whilst inactive TADs are not involved (Kaikkonen et al. 2014). Here the stimulation by pro-survival ligand Ang-1 globally enhanced ERG occupancy by both inducing *de novo* ERG binding sites and increasing binding at existing sites. It is possible that these new binding sites are regions which had ERG motifs that were now commissioned for binding following the trigger of Ang-1 activation. Abdel-Malak et al. (2008) showed that exposure to Ang-1 in HUVEC increased the DNA binding of AP-1 to a locus on the IL-8 promoter suggesting that the response to Ang-1 is mediating a response on several transcriptional components.

The prioritisation of disease-associated variants at ERG super-enhancers, ERG-bound enhancers and ERG binding loci was performed using a binomial distribution test against background variants from CEU. Random regions were assigned by shuffling size- and chromosome-matched regions in open chromatin. However, there are several ways that this analysis could have been conducted to improve the conclusions. Firstly, the background variants could have been permuted based on more robust stringency. For example, using SNPs with the (1) same mutation allele frequency (MAF), (2) number of SNPs in linkage disequilibrium and (3) distance to nearest gene. Then empirical p-values

could be calculated for each real SNP versus a permuted SNP on every genomic region and a final correction for multiple testing. Such permutation tests would obtain larger p-values than parametric tests reducing the number of false positives. Furthermore, the random shuffled regions used in the original analysis could also have been composed of other endothelial super-enhancers that were not identified by ERG enrichment. Such regions would provide the controls necessary to determine whether any given endothelial super-enhancer is also enriched for the disease-associated SNPs or whether it is only those defined by high levels of ERG.

Non-coding genetic variants have been difficult to prioritise for functional interrogation. Candidates for regulatory variants have commonly been selected based on histone modifications, open chromatin and TFBS. Evidently, TF binding sequences can be disrupted by SNPs, rendering the site ineffective for TF binding. In other interesting cases non-coding variants have reversely created a *de novo* binding site for a TF such as the case for a variant that generated a CCAAT/enhancer binding protein (C/EBP) binding site which enhances *SORT1* gene expression consequently associated with an increased risk for myocardial infarction (Musunuru et al. 2010). Interestingly, here CVD-associated SNPs were the most highly enriched in ERG-defined super-enhancer (Figure 2.13A). Genetic variants in ECs could underlie a number of diseases as the contribution from the vasculature is important for the health of all organs and the enrichment analysis here further prioritises digestive system and respiratory tract associated SNPs in ERG super-enhancers. Beyond organ function, SNPs linked to the immune system are most highly enriched at ERG loci and ERG-bound enhancers. Given the role of ECs in inflammation including participation in the innate and adaptive immune response (Al-Soudi et al. 2017) there are open questions regarding the influence of EC to disease predisposition and progression arising from genetic variants. To address the gap, large-scale studies are collecting expression quantitative trait loci (eQTL) to quantify causal effects on gene expression to nearby (*cis*) genes of a given SNP in different tissues by comparing the common allele expression to that of the alternate allele (Ardlie et al. 2015). However,

such analysis is retrieved from whole tissues and has yet to be extended to a single cell type. For EC this could be possible by collecting ECFC derived from donors to determine their contribution to genetic diseases.

3 CELL-LINEAGE SPECIFICITY BY ERG GENE REGULATION IN HUVEC VERSUS VCAP CELLS

3.1 Introduction

In chapter 2 I described ERG's critical role in the vascular endothelium and identified that ERG independently modulates a subset of key endothelial super-enhancers that define lineage-specific genes. Functionally, ERG controls vascular endothelial homeostasis (Shah et al. 2016). In contrast, ERG is aberrantly expressed and implicated in certain cancer types including Ewing's sarcoma, leukeamia and most prevalently in prostate cancer (Sorensen et al. 1994; Thoms et al. 2011; Tomlins et al. 2005). In prostate cancer ERG has oncogenic potential as a result of a chromosomal translocation resulting in a fusion with androgen-responsive TMPRSS2 (transmembrane protease, serine-2). This fusion is present in approximately 50% of prostate cancers and is linked to enhanced metastasis, increased invasiveness and disease recurrence (Adamo et al. 2016; Hägglöf et al. 2014). The TMPRSS2-ERG fusion protein contains the consensus ETS DNA binding domain of ERG (Adamo et al. 2016). The striking contrast of ERG acting in a homeostatic versus oncogenic fashion has not been investigated. Understanding the underlying molecular mechanisms could prove essential given any potential ERG-based or ERG-related therapy in CVD or prostate cancer.

Cell-type-specific TF recruitment has previously been studied in cancer cells. FOXA1, a pioneer TF, was found to differentially bind to over 50% of chromatin sites of MCF7 breast cancer cells and the LNCaP prostate cancer cell line (Lupien et al. 2008). In this chapter, I use genomic datasets generated in human prostate epithelial cancer cell line VCaP which possess the TMPRSS2-ERG fusion and compare to datasets generated in homeostatic HUVEC. I reveal cell-type-specific chromatin binding and differential regulation of gene expression by ERG. These differences are predominantly linked to distal enhancers as opposed to proximal promoters. I further demonstrate contrasting

super-enhancer profiles between HUVEC and VCaP cells associating this to cell-type-specific molecular machinery responsible for histone mark distribution and functional collaboration with other master TFs.

3.2 Methods and Data

Cell culture of HUVEC and VCaP cell line

HUVEC were cultured as described in chapter 2. The prostate cancer cell line VCaP were obtained from the American Type Culture Collection (ATCC) and cultured in DMEM supplemented with 10% fetal bovine serum (FBS) (Biosera, Labtech).

ChIP-seq data mapping, peak calling and visualisation

Raw fastq sequencing files were obtained from the European Nucleotide Archive (ENA). I performed the ChIP-seq analysis for VCaP using the same pipeline as for HUVEC. Briefly, ChIP-seq reads were aligned to the hg19 reference genome using Bowtie 2 (Langmead et al. 2012), SAM files were converted to BAM using SAMtools (Li et al. 2009) and MACS version 2.1.2 (Zhang et al. 2008) was used to call peaks against input controls. Binding sites were assigned to genes using Bioconductor package ChIPseeker in R (Yu et al. 2015). BigWig files were visualised in the UCSC Genome Browser (https://genome.ucsc.edu).

Analysis of microarray data

Microarray data for ERG transcriptome profiling from TMPRSS2-ERG fusion-positive VCaP prostate cancer cells was obtained from Wang et al. (2014). Differential expression analysis was performed using the Limma package (Ritchie et al. 2015) in R. Meaningful gene expression changes were selected by a Benjamini and Hochberg adjusted p-value $<$ 0.10.

Immunoblotting

Whole cell protein lysates were collected using CelLytic reagent (Sigma). Immunoblotting of cell lysates was performed with primary antibodies anti-ERG (ab133264, 1:1000, Abcam) and anti-GAPDH (MAB374, 1:10000, Millipore). Primary antibodies were detected using fluorescently labelled secondary antibodies: goat anti-rabbit IgG DyLight 680 and goat anti-mouse IgG DyLight 800 (Thermo Scientific). Detection and quantification of fluorescence intensity were performed using an Odyssey CLx imaging system (LI-COR Biosciences) and Odyssey Image Studio v4.0 software.

Gene silencing using siRNA and real-time quantitative PCR (RT-qPCR)

VCaP cells were transfected using 100 nM siRNA targeting exon 7 of ERG (Invitrogen; 5'-ACTCTCCACGGTTAATGCATGCTAG-3') or Stealth negative control siRNA (Invitrogen) using Lipofectamine RNAiMAX (Invitrogen) by Dr Karen Frudd (Imperial College London). RNA was collected 48 hrs post-siRNA treatment and RNA was extracted using the RNeasy Mini Kit (Qiagen). cDNA was synthesised using Superscript III Reverse Transcriptase (Invitrogen) and RT-qPCR was performed on a Bio-Rad CFX96 system using PerfeCTa SYBR Green Fastmix (Quanta Biosciences). Sequences of the oligonucleotides used for RT-qPCR are listed in Table 3.1. Data were plotted using Prism 8.0 (Graph Pad).

Defining super-enhancer

Super-enhancers were defined using the ROSE algorithm as in HUVEC in chapter 2 (Lovén et al. 2013; Whyte et al. 2013). Active enhancers were stitched if within 12.5 kb and promoter regions (± 2 kb from TSS) were excluded. H3K27ac occupancy was used as the enrichment parameter to classify super-enhancers from typical enhancers.

Functional enrichment

Gene ontologies were assigned using the collection from the Molecular Signatures

Name	Orientation	Oligonucleotide sequences
ERG	Forward	GGAGTGGGCGGTGAAAGA
	Reverse	AAGGATGTCGGCGTTGTAGC
GAPDH	Forward	CAAGGTCATCCATGACAACTTTG
	Reverse	GGGCCATCCACAGTCTTCTG
ICAM1	Forward	CTGAAACTTGCTGCCTATTGGG
	Reverse	ACACATGTCTATGGAGGGCCA
SMAD7	Forward	TGGTGTGCTGCAACCCCCATCA
	Reverse	GCACAGCATCTGGACAGTCTGC
IL8	Forward	AGGAACCATCTCACTGTGTG
	Reverse	GGCATCTTCACTGATTCTTG
CDH5	Forward	AGCCAGCCCAGCCCTCAC
	Reverse	CCTGTCAGCCGACCGTCTTTG
CLDN5	Forward	GGGAAAGCCCCTGTGCCACC
	Reverse	TCCAGCCCCGCTCTGAGTCC
HDAC6	Forward	CTAGCAGACACCTACGACTCAG
	Reverse	GCAATAGCCATCCATAAGACTGTG
TMPRSS2	Forward	CGGATGCACCTCGTAGACAG
	Reverse	TCACCACCAGCTATTGGACC
EGR1	Forward	CTTCAACCCTCAGGCGGACA
	Reverse	GGAAAAGCGGCCAGTATAGGT
KLK2	Forward	GGTGGCTGTGTACAGTCATGGAT
	Reverse	TGTCTTCAGGCTCAAACAGGTTG
HES1	Forward	AGAAAGATAGCTCGCGGCATT
	Reverse	CGGAGGTGCTTCACTGTCAT
HLA-DMB	Forward	GCTGGCCACTCTAGTTACACTC
	Reverse	GAGGTCCCCCAAGTTGCTAA
TMEM158	Forward	GGCTGAACCGTAAGCCCATT
	Reverse	CTCCACACCACGATGACCAG

Table 3.1: Primer sequences used for RT-qPCR.

Database (MSigDB) from Broad Institute (Liberzon et al. 2011).

Gene set enrichment analysis (GSEA) was conducted using the GSEA software (Subramanian et al. 2005). The significance of the normalised enrichment score (NES) was based on a p-value adjusted for multiple comparisons and from analysis with 1000 permutations. A nominal p-value for each gene set is also reported to assess the significance of the enrichment score based on the permutation test procedure.

Functional annotation

ChIP-seq data for correlation plots were normalised to input data using the bamCompare tool in deepTools2.0 (Ramírez et al. 2016) and reported as $\log_2$ of the input-normalised ChIP-seq tag density. Resulting bigWig files were used to obtain a multiBigwigSummary in deepTools2.0 from which correlations were plotted in R using the ggplot2 package.

Aggregate plots showing the average ChIP-seq signal were generated by first obtaining tag density profiles by converting ChIP-seq bam files to tag directories using the makeTagDirectory command in the HOMER analysis suite (Heinz et al. 2010). To plot the ChIP-seq signal at super-enhancer regions, the HOMER command 'annotatePeaks.pl' was used with tag density obtained across 10 bp bins. The ChIP-seq signal reported here is the result of normalisation to a standard 10 million mapped tags and to the read depth per base pair per peak.

ChIP-seq binding site distribution for ERG were visualised as a distribution plot (Figure 3.3B) using ChIPseeker function *plotDistToTSS* (Yu et al. 2015).

Heatmaps to visualise gene expression changes were plotted in R. Genes associated with Biological network pathways (extracellular matrix organisation, Notch signalling) were derived from the Reactome database (Fabregat et al. 2016).

Overlap between genomic regions was conducted with intersectBed from BEDTools

(Quinlan et al. 2010). Hypergeometric distribution testing to compare the overlap of promoters and enhancers between HUVEC and VCaP cells used genome-wide occurrences of promoter and enhancer regions as the expected (background) frequencies. Promoter regions were those defined by RefSeq annotations from NCBI (Pruitt et al. 2007) and enhancers were from the FANTOM5 project using Cap Analysis of Gene Expression (CAGE) (Lizio et al. 2015). The hypergeometric distribution statistics were conducted in R.

Motif analysis

To find the most enriched motifs at ERG binding sites I used the 'findMotifsGenome.pl' script from HOMER (Heinz et al. 2010). The analysis was performed on ERG sites ± 200bp from the centre of the binding site. The background is generated from randomly selected sequences in the genome which match the percentage GC content of the query sites.

Statistics

All statistical analyses were performed in R. Fisher's exact test is performed to analyse contingency tables including those that are represented as Venn diagrams. A Shapiro-Wilk test for normality was performed for all data represented in a boxplot and in all cases was found to be significantly not normally distributed and therefore nonparametric tests were used throughout to test for significance between different test groups. A Wilcoxon sum-ranks test was performed when comparing two groups. A Kruskal-Wallis test was used when comparing multiple groups (more than two) followed by post-hoc testing using the Wilcoxon sum-ranks test.

Data

Data used in this chapter is listed in Table 3.2.

Experiment	Factor	Cell type	Condition	GEO ID	Citation / Distributor
ChIP-seq	ERG	HUVEC	untreated	GSM3557980	Kalna et al. 2019
ChIP-seq	H3K27ac	HUVEC	untreated	GSM733691	ENCODE/Broad Institute
ChIP-seq	H3K4me1	HUVEC	untreated	GSM733690	ENCODE/Broad Institute
ChIP-seq	H3K27me3	HUVEC	untreated	GSM733688	ENCODE/Broad Institute
ChIP-seq	ERG	VCaP	untreated	GSM717395	Chng et al. 2012
ChIP-seq	ERG	VCaP	untreated	GSM353637	Yu et al. 2010
ChIP-seq	H3K27ac	VCaP	untreated	GSM1328982	Asangani et al. 2014
ChIP-seq	H3K4me1	VCaP	untreated	GSM353631	Yu et al. 2010
ChIP-seq	H3K27me3	VCaP	untreated	GSM353621	Yu et al. 2010
Microarray		HUVEC	siCtl v siERG	GSE32984	Birdsey et al. 2012
Microarray		VCaP	siCtl v siERG	GSE53994	Wang et al. 2014

Table 3.2: Data used in this chapter.

3.3 Results

3.3.1 ERG binding and gene regulation in VCaP cells

To investigate the chromatin and transcriptional profile of oncogenic ERG in prostate cancer VCaP cells I analysed ChIP-seq data for ERG from existing published datasets. I compared two datasets from Chng et al. (2012) and Yu et al. (2010) to assess the replicability of the ERG ChIP-seq data. The ChIP-seq tag density from the two completely independent datasets showed high concordance (Figure 3.1A). To note, the study by Yu et al. (2010) did not have an input control for the ChIP-seq data and therefore the tag density is not normalised as for Chng et al. (2012). I therefore also used the Chng et al. (2012) dataset for further analysis. Integration of this ChIP-seq dataset with transcriptome profiling in ERG-depleted VCaP cells (Wang et al. 2014) indicated 584 directly activated genes and 589 directly repressed genes, a significant proportion, 70% and 60% respectively, of the total genes activated and repressed by ERG (Figure 3.1B). Functional interrogation by gene ontology analysis of these directly bound regions suggests that in VCaP cells ERG activates cell proliferative pathways, including DNA replication and cell division, yet represses apoptosis (Figure 3.1C).

3.3.2 Cell-type-specific ERG-dependent gene regulation

Given that ERG is expressed in both vascular endothelial cells and in a subset of prostate cancer cells, it is important to delineate the molecular mechanism of the two cell types for possible future interventions. I compared the gene expression profiles regulated by ERG between HUVEC and VCaP. I first assessed the levels of ERG protein in the two cell types. Immunoblotting for ERG protein levels indicated higher levels of ERG in HUVEC compared to VCaP (Figure 3.2A). I performed this in two batches of pooled HUVEC and two different passages of VCaP cells at passage 3 and 18 (Figure 3.2A). I analysed microarray transcriptome profiling following siRNA knockdown of ERG in

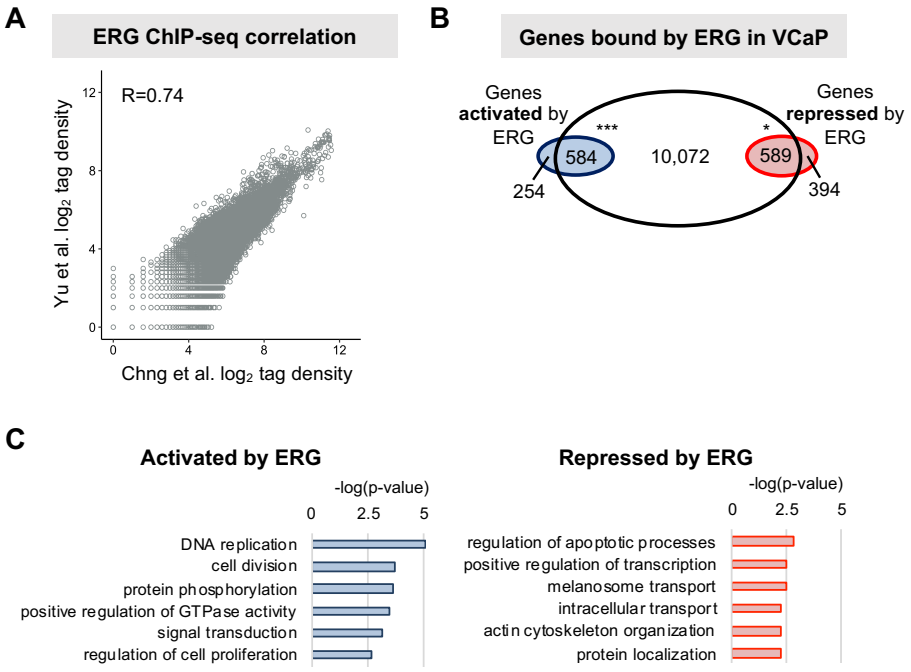


Figure 3.1: Integration of ChIP-seq and transcriptome profiling in VCaP cells. (A) Scatterplot correlation of two independent ERG ChIP-seq experiments represented as $\log_2$ of the tag density. Pearson correlation coefficient = 0.74. (B) Venn diagram illustrating the comparison between genes associated with ERG binding sites and ERG-dependent gene expression. ERG activated and repressed genes are significantly more bound than expected by chance. *p-value < 0.05, ***p-value < 0.001, Fisher's exact test. (C) Gene ontology analysis defines the molecular pathways associated with genes directly bound and activated or repressed by ERG. Significance shown as -log(p-value).

both HUVEC and VCaP; and by only considering the subset of directly ERG-bound genes, I observed that ERG directly regulated many distinct genes between the two cell types (Figure 3.2B). In fact only 88 genes were commonly activated and 53 genes were commonly repressed which is less than 5% of all genes regulated by ERG in HUVEC and VCaP (Figure 3.2C). Interestingly, 84 genes were found to be activated by ERG in HUVEC but conversely repressed in VCaP and 24 genes were activated in VCaP and repressed in HUVEC (Figure 3.2C). The relative regulation between HUVEC and VCaP commonly regulated ERG targets (249 genes in total) is displayed as a heatmap in Figure 3.2D.

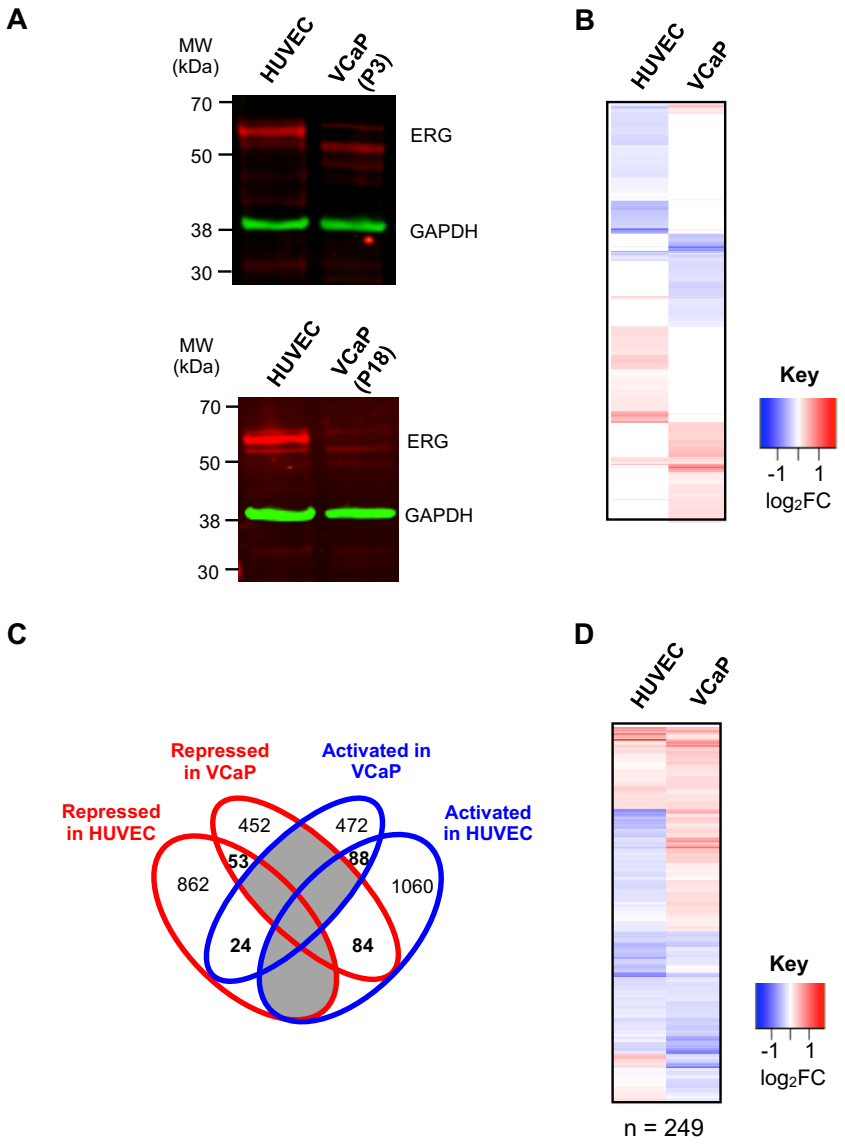


Figure 3.2: ERG controls a lineage-specific gene expression program. (A) Immunoblot indicating the levels of ERG protein expression in HUVEC versus VCaP in two independent batches of mixed-population HUVEC and two different passages of VCaP cancer cells (passage 3 and 18). GAPDH included as a loading control. (B) Heatmap demonstrating the $\log_2$FC of significantly regulated genes from transcriptome profiling in HUVEC versus VCaP. Upregulated (repressed) genes in red and downregulated (activated) in blue. Genes which are not significantly differentially expressed are excluded (white). Gene names excluded. (C) Venn diagram showing overlap of differentially regulated genes bound directly by ERG in HUVEC versus VCaP cells. The genes commonly or conversely regulated in both cell types are shown in bold (total of 249). (D) Heatmap demonstrating the regulation of the 249 genes found to be regulated in the same or opposite direction in HUVEC and VCaP. Upregulated (repressed) genes in red and downregulated (activated) in blue. Gene names excluded.

Furthermore, it is known that ERG controls a number of common pathways of HUVEC and VCaP cells. Extracellular matrix organisation is mediated by ERG during cell migration and plays an important role during angiogenesis. Inhibition of ERG in HUVEC results in impaired cell motility and lumen formation due to disruption of actin and tubulin cytoskeletal dynamics, partly through modulating pathways linked to the small GTPases (Birdsey et al. 2012; Yuan et al. 2011). In prostate cancer, tumour progression is influenced by the oncogenic role of ERG in promoting tumour cell invasion (Cai et al. 2013). Notch signalling is also regulated through ERG in the two cell types as ERG has been shown to control the balance of expression of two Notch ligands Delta-like ligand 4 (DLL4) and JAGGED1 (JAG1) (Shah et al. 2017). In prostate cancer, the Notch pathway regulator genes were shown to be enriched with ERG-dependent enhancer elements (Kron et al. 2017). Using genes associated with these two specific pathways, I found that these were also largely regulated in a lineage-specific manner (Figure 3.3A). Interestingly, ERG targets included the Notch ligand JAG1, which was found to be repressed in HUVEC but activated in VCaP. This concurs with the finding that JAGGED1 is significantly more highly expressed in metastatic prostate cancer as compared to localised prostate cancer cases (Santagata et al. 2004). Whether JAG1's pro-angiogenic role is important in TMPRSS2-ERG-positive prostate cancer as a mediator of metastasis is unknown. However, in ECs JAGGED1 has been confirmed as an antagonist of the Notch-DLL4 signalling axis which hampers angiogenic growth (Pedrosa et al. 2015). Moreover, other genes in these selected pathways are only regulated in one cell type such as oncogene MYC found to be activated by ERG in VCaP and HEY1 which is alternatively repressed by ERG only in HUVEC (Figure 3.3A). Thus, ERG appears to be controlling the homeostatic and oncogenic functions of EC and prostate cancer cells at the transcriptional and chromatin level.

To provide validation of the two microarrays used here for comparative analysis, gene targets with a significant change in gene expression were assessed by RT-qPCR. As for the microarray studies, HUVEC and VCaP were treated with siRNA targeting ERG or a

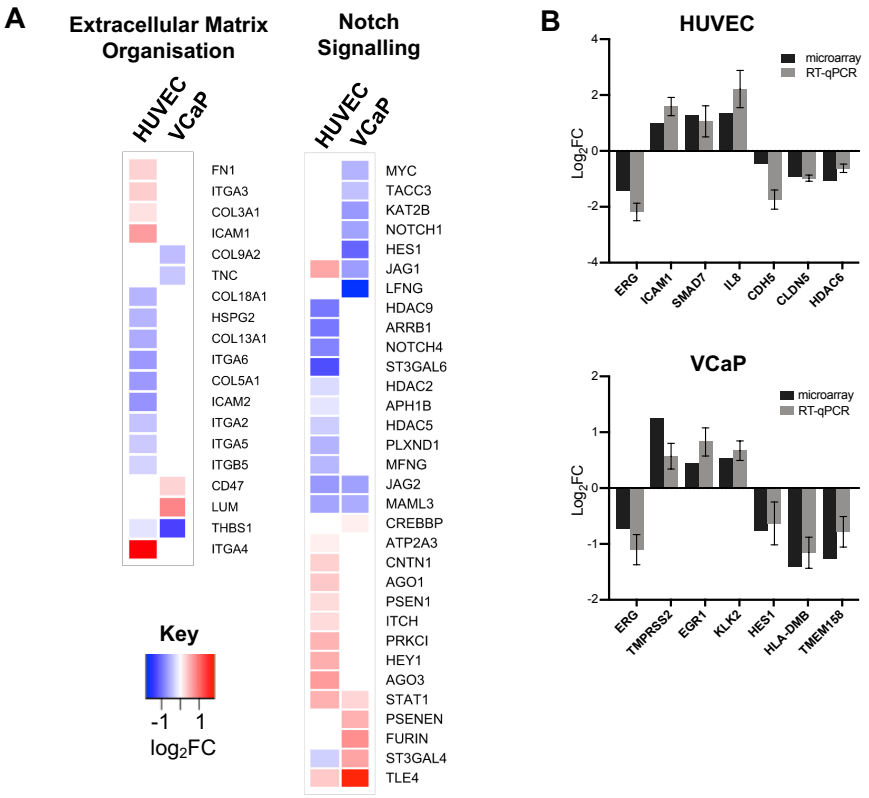


Figure 3.3: ERG directly regulates genes of common endothelial and prostate cancer pathways. (A) ERG target gene regulation linked to pathways important in both HUVEC and VCaP. Extracellular matrix organisation (left) and Notch signalling (right). Gene names indicated. (B) Comparison of gene expression changes by microarray and RT-qPCR in HUVEC and VCaP cells following siCtl and siERG treatment. Gene regulation presented as $\log_2$ FC. RT-qPCR data is normalised to housekeeping gene *GAPDH*. RT-qPCR values are mean ± SEM, n=4.

negative control sequence for RT-qPCR analysis of both up- and down-regulated genes. Indeed, comparing these two gene expression methods showed that selected genes were found to be regulated in the same direction by both microarray analysis and RT-qPCR (Figure 3.3B).

3.3.3 Cell lineage-specific ERG binding landscape

Next, ERG binding profiles were compared in the two cell types to determine occupancy differences in the genome. ERG binding sites shared between HUVEC and VCaP accounted for 23% of all binding sites in HUVEC (Figure 3.4A). Binding sites shared between the two cell types were mostly (70%) distributed in the vicinity of the TSS of genes (Figure 3.4B). The binding sites unique to HUVEC or VCaP were located distal to the TSS (Figure 3.4B). Histone modifications reveal chromatin states and act to modulate gene expression. I therefore classified promoters by selecting H3K4me3-enriched and RefSeq assigned TSS regions and enhancers by co-occupancy of H3K4me1/H3K27ac in the individual cell types. I compared genomic ERG binding loci in these assigned chromatin states and by using a hypergeometric distribution test determined the significance of overlap. I found that ERG promoter binding is shared with 58% of sites in HUVEC in common with VCaP. ERG-bound enhancers were distinct in the two cell types with no significant overlap as only 18% of enhancer sites in HUVEC were shared (Figure 3.4C).

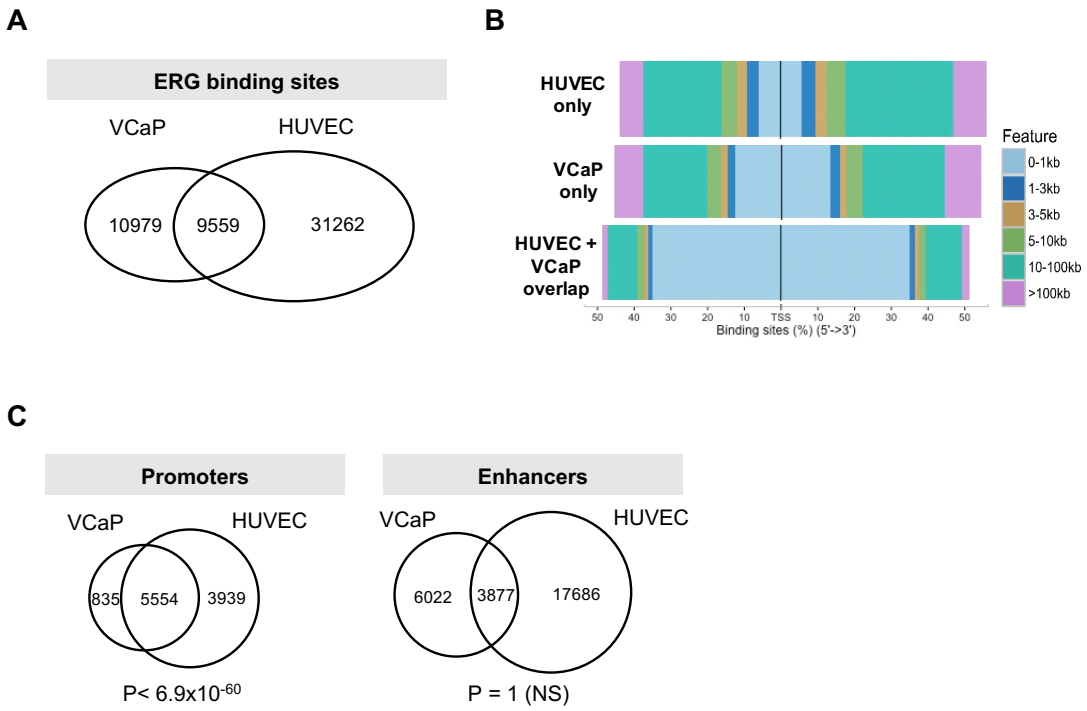


Figure 3.4: ERG binds differentially in HUVEC and VCaP cells. (A) Overlap of ERG binding loci between HUVEC and VCaP. Only 23% of binding sites in HUVEC are shared with VCaP. (B) Visualisation of ERG binding distribution upstream and downstream of the TSS of genes. Percentage of sites occupied in given distance intervals. Sites unique to HUVEC or VCaP are located more distal in marked contrast to sites shared by both cells which are located nearby the TSS. (C) Overlap of ERG binding sites at H3K4me3-enriched RefSeq annotated TSS promoter regions and at enhancers marked by H3K4me1/H3K27ac co-occupancy in HUVEC and VCaP cells. Significance of overlap reported as a hypergeometric p-value. NS = not significant.

3.3.4 Super-enhancer profiling reveals a lineage-specific program

Clusters of enhancers, termed super-enhancers, determine cell fate by controlling cell-type-specific gene expression (Hnisz et al. 2013; Whyte et al. 2013). Therefore, to elucidate the cell-type-specific genes in VCaP cells I identified super-enhancers using ranked enrichment of H3K27ac (Asangani et al. 2014). I identified 208 super-enhancers associated with 199 genes, including several oncogenes as reported in studies of other cancers types (Hnisz et al. 2013; Mansour et al. 2014; Northcott et al. 2014) (Figure 3.5A). The super-enhancer genes have been implicated in prostate cancer diagnosis and progression. Interestingly, as observed in HUVEC, ERG binds the majority of these super-enhancers in VCaP (91%) which is significantly more than the proportion bound at typical enhancers (48%) (Figure 3.5B). Super-enhancer-associated genes in VCaP had significantly higher average gene expression levels than typical enhancers, using expression levels from transcriptome profiling in control siRNA-treated VCaP (Figure 3.5C). Collectively, these results establish a role for ERG in super-enhancer assembly and gene expression in VCaP prostate cancer cells.

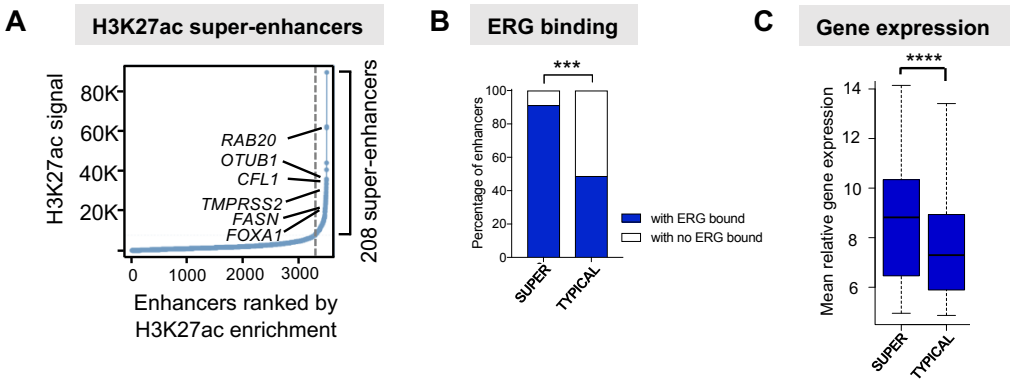


Figure 3.5: H3K27ac-defined super-enhancers in VCaP. (A) Super-enhancers defined by H3K27ac in VCaP. Key prostate pathogenesis genes are indicated. (B) ERG binding at super-enhancers and typical enhancers in VCaP. ERG binds significantly more to super-enhancers than typical enhancers; ***p-value < 0.001, Fisher's exact test. (C) Boxplot distribution of gene expression levels from transcriptome profiling in siCtl-treated VCaP cells. Super-enhancer-associated genes have a significantly higher gene expression level than typical enhancers; ****p-value < 0.0001, Wilcoxon rank-sum test.

3. CELL-LINEAGE SPECIFICITY BY ERG GENE REGULATION IN HUVEC VERSUS VCAP CELLS

To dissect the differences in super-enhancer landscapes between HUVEC and VCaP I first investigated whether super-enhancer-associated genes in VCaP were similar or distinct to those in HUVEC. By using GSEA and ranking super-enhancers in HUVEC, I found there was no significant association between HUVEC and VCaP super-enhancers (Figure 3.6A). Further confirming this, super-enhancer genes in HUVEC and VCaP are linked to distinct functional pathways (Figure 3.6B). As described in chapter 2, HUVEC super-enhancers are functionally most highly associated with genes involved in angiogenesis, a vascular-specific function. Super-enhancer genes of prostate cancer VCaP cells are functionally related to transcription by RNA Pol II providing a means by which other oncogenes could be uncontrollably expressed. Collectively, super-enhancer profiles in these two cell types are notably distinct.

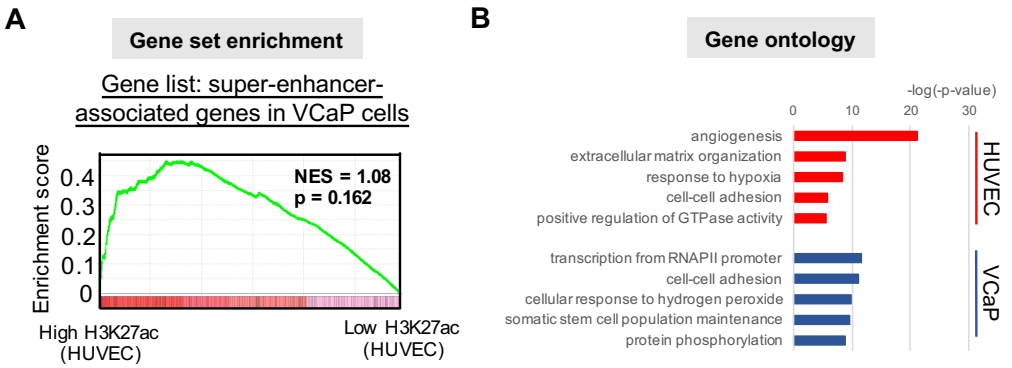


Figure 3.6: Super-enhancers are explicitly independent in HUVEC and VCaP. (A) Gene set enrichment analysis (GSEA) of the 208 VCaP super-enhancer genes compared to the ranked gene list of 917 HUVEC super-enhancers. Both super-enhancer regions were identified by H3K27ac enrichment. Normalised enrichment score (NES) = 1.08, p-value = 0.162. (B) Gene ontology analysis of molecular pathways linked to HUVEC and VCaP super-enhancers. Top 5 scoring pathways displayed. Significance shown as -log(p-value).

It is evident that super-enhancer profiles and ERG occupancy in these two cell types are profoundly contrasting. Super-enhancers at specific endothelial super-enhancer genes (such as *CDH5*, *VWF* and *FZD4*) are occupied by ERG and active mark H3K27ac in HUVEC and comparatively there is limited abundance of ERG and H3K27ac at equivalent regions in VCaP (Figure 3.7). Evidently, at VCaP super-enhancers a similar pattern is observed with high levels of ERG and H3K27ac in VCaP genes (including

TMPRSS2, *RAB20* and *SPON2*) versus very low enrichment in HUVEC (Figure 3.7).

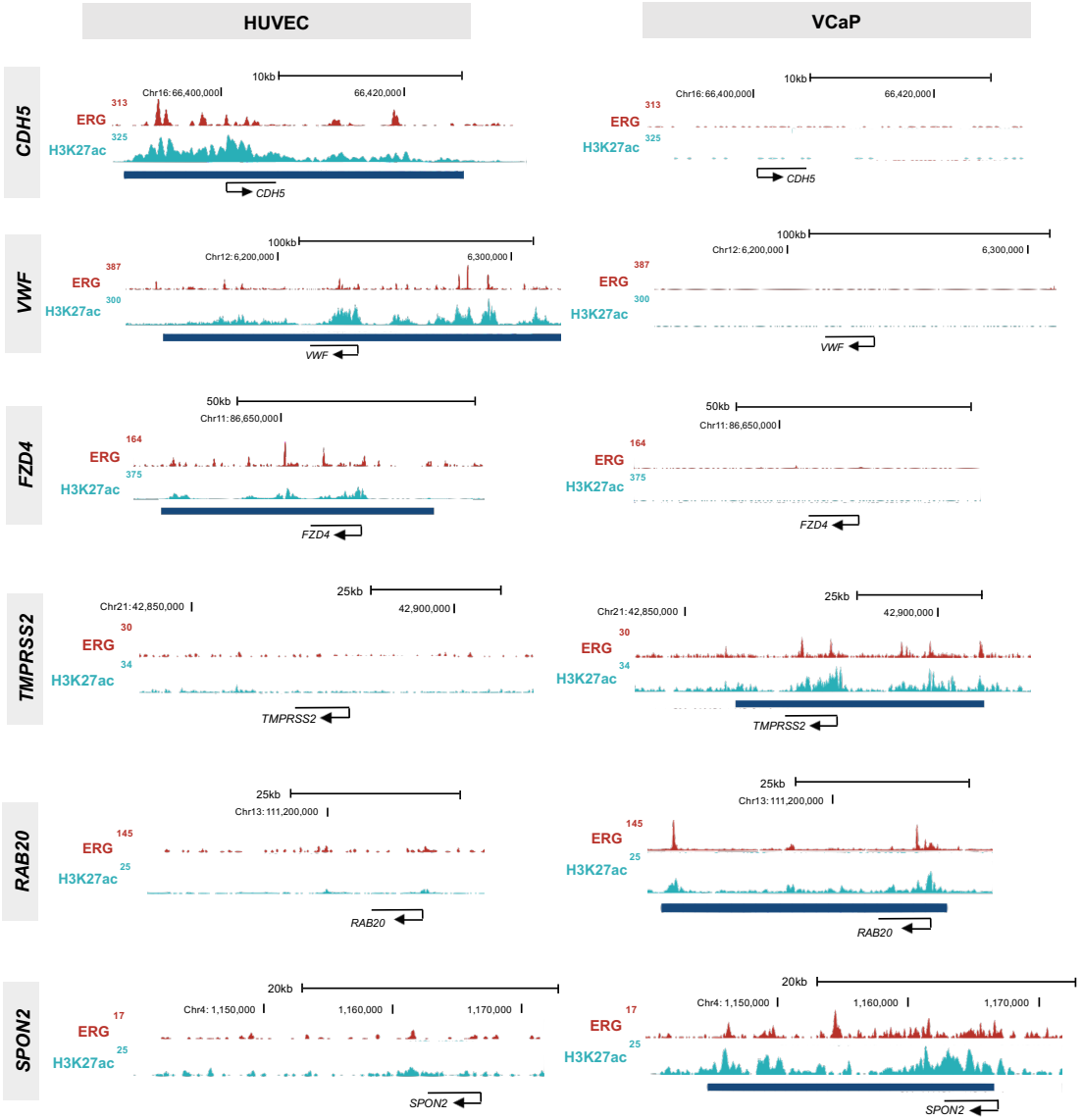


Figure 3.7: Super-enhancers and associated transcription factor ERG are highly cell-type-specific. UCSC genome browser screenshots of ERG and H3K27ac occupancy at super-enhancer genes in the same loci of HUVEC and VCaP. HUVEC super-enhancer-associated genes are *CDH5*, *VWF* and *FZD4* and VCaP super-enhancer-associated genes are *TMPRSS2*, *RAB20* and *SPON2*. The blue bars underneath profiles indicate super-enhancer regions.

3.3.5 Cell-type differences orchestrated by chromatin-remodelling and master transcription factors

Observation of cell-type-specific ERG and H3K27ac occupancy in HUVEC versus VCaP at a few loci prompted a comparative analysis of all super-enhancers in these two cell types. Some chromatin-remodelling complexes have been reported to be cell-type-specific (Ho et al. 2010). I assessed the read-out of chromatin regulator activity by investigating histone modification marks at super-enhancers in the two cell types. I found that histone modifications that commonly mark enhancers and are considered active, H3K27ac and H3K4me1, are enriched at HUVEC super-enhancer regions whereas in marked contrast these regions are poorly occupied by these active marks in VCaP (Figure 3.8A). For repressive mark H3K27me3 the reverse pattern is observed with higher levels in VCaP than HUVEC at HUVEC super-enhancer regions. Conversely, VCaP super-enhancers were enriched in active modifications but not in HUVEC which alternatively showed higher levels of repressive mark H3K27me3 (Figure 3.8A).

To further assess mechanisms of cell-type-specificity in HUVEC vs VCaP, I explore the concept that super-enhancers are hubs for collaborative TF assembly. Sequence recognition motifs are likely spaced in close proximity in *cis*-regulatory elements given the high TF density at super-enhancers: ERG likely co-occupies genomic loci with additional cell-type-specific TFs. To elucidate which TFs are potentially involved in the two cell types, I performed *de novo* motif discovery on sites bound by ERG $\pm$ 200 bp. In HUVEC, the highest scoring motifs alongside ERG were AP1, FOXO1, GATA2 and SOX3 (Figure 3.8B). However, it should also be noted that other members of the ETS TF family are likely to be involved in co-binding with ERG but remain undistinguished in a motif analysis. In VCaP cells a unique and distinct set of master TFs are enriched at ERG loci; FOXA1, HOXB13, NF1-halfsite and GATA2 which is interestingly shared with HUVEC (Figure 3.8B). Collectively, these findings broaden the concept that the combination of epigenetic mechanisms and TF profiles are coordinated in a cell-type-specific manner even in the presence of a master TF such as ERG.

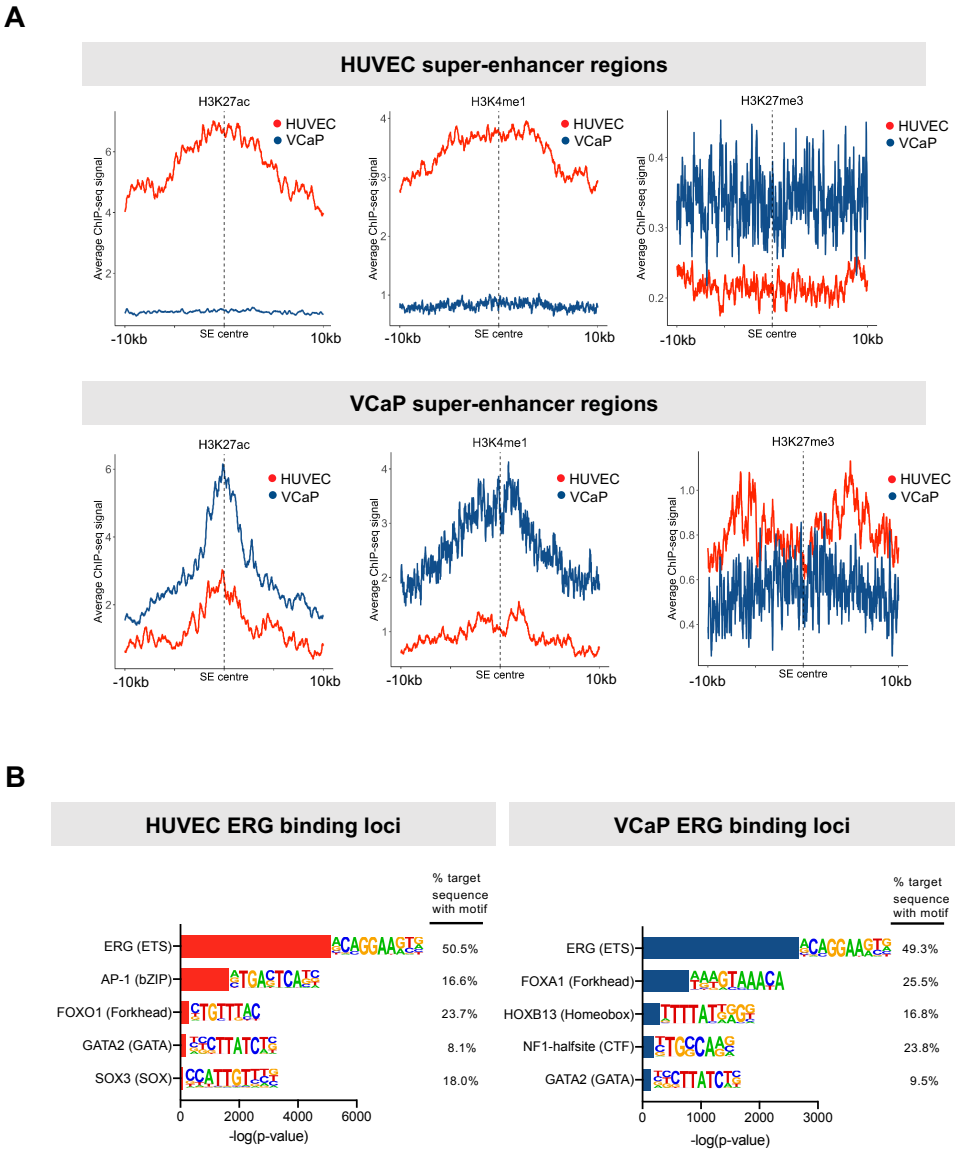


Figure 3.8: Cell-type-specificity partly determined by histone code and master transcription factors. (A) Aggregate plots of active histone modifications H3K27ac and H3K4me1 and repressive H3K27me3 from HUVEC (red) and VCaP (blue) at HUVEC super-enhancers (top) and VCaP super-enhancers (bottom). Plots centred on each super-enhancer centre ± 10 kb. Notably, on average VCaP super-enhancers are smaller than HUVEC super-enhancers. (B) Motif analysis to determine cooperative master TFs at ERG binding loci ± 200 bp in HUVEC and VCaP. Top 5 highest ranked TF motifs indicated with the most highly occurring member of the TF family indicated. Ranking based on significance of enrichment as -log(p-value). Percentage occupancy at the ERG sites is also reported for each TF.

3.3.6 Discussion

In this chapter I identify contrasting transcriptional regulatory networks governed by ERG in HUVEC versus VCaP cells. I demonstrate stark differences in ERG chromatin binding with shared occupancy of promoter regions but independent enhancer binding locations in the two cell types. In addition, functional differences by means of ERG-dependent gene regulation was discovered between the two cell types. The mechanism by which these specificities are achieved is underpinned by the cell-type-specific transcriptional machinery and combinatorial TF assembly.

Super-enhancers have been characterised in many cancer cell types and found to be adjacent to known oncogenes (Chapuy et al. 2013; Hnisz et al. 2013; Zhang et al. 2013). In this chapter, super-enhancers assigned by H3K27ac enrichment in VCaP cells were associated with many oncogenes, a number of which have been implicated a function in prostate cancer. VCaP super-enhancer genes are mediated in cancer metastasis (*SPON2*) (Qian et al. 2012), tumour aggressiveness (*CFL1*) (Lu et al. 2015), cell invasion (*OTUB1*) (Iglesias-Gato et al. 2015) and poor prognosis (*FOXA1*) (Sahu et al. 2011). Interestingly, highly ranked super-enhancers linked to the regulation of long non-coding RNAs (lncRNAs) were discovered and those that have been characterised are overexpressed in prostate cancer (Chakravarty et al. 2014; Chang et al. 2018; Du et al. 2013). They are NEAT1 (nuclear-enriched abundant transcript 1), MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) and PCAT1 (prostate cancer associated transcript 1) which are considered oncogenes critically found in this study to be occupied by ERG in the VCaP cancer genome. The functional roles of NEAT1 and MALAT1 include the binding to active genomic sites in human cells with colocalisation of the two lncRNAs exhibited at many loci (West et al. 2014). The super-enhancer-associated genes in ERG-expressing prostate cancer cells provide candidates to further uncover the mechanisms which underly prostate cancer aetiology and maintenance.

The transcriptional assembly tightly controls gene expression which differs between

different cell types. The DNA binding TFs recruit non-DNA-binding chromatin remodellers and modifiers to maintain a specific chromatin state. I discover opposing active and repressive histone marks in super-enhancer regions of HUVEC vs VCaP cells and given the presence of ERG in both these cell type's super-enhancer elements there is a possible dependence on ERG-associated chromatin remodellers and modifiers to the cell-type chromatin state. From prostate cancer studies investigating the TMPRSS2-ERG gene fusion, ERG has been shown to interact with methyltransferase EZH2, chromatin reader BRD4 and ATP-dependent chromatin remodeller SWI/SNF (Blee et al. 2016; Sandoval et al. 2018; Yu et al. 2010). These interaction have not been explicitly elucidated in EC; however, the interaction between ERG and acetyltransferase p300 has thus far been most widely reported in EC (Fish et al. 2017; Jayaraman et al. 1999). Further studies that investigate the activities between ERG and chromatin regulators would shed light on targetable mechanisms to modulate ERG levels in vascular disease or ERG-expressing prostate cancers.

The other mechanism of cell-type-specificity I explore in this chapter is about differences in cooperative TF binding. I reveal the putative TFs that co-bind with ERG in the two cell types through motif analysis. In HUVEC, I demonstrate the most highly scored TF motif at ERG binding sites to be that for the ubiquitously expressed activating protein 1 (AP-1). This is in line with a study which reported that ETS and AP-1 TF members select endothelial enhancers (Hogan et al. 2017). In prostate cancer VCaP cells, FOXA1 and HOXB13 are the most overrepresented motif sequences at ERG binding sites. Interestingly, the identity of these TFs has previously been established in prostate cancer development with evidence of both TFs physically interacting with ERG and co-binding at enhancer elements in VCaP cells (Kron et al. 2017).

Another key TF implicated in prostate cancer biology is androgen receptor (AR). The motif analysis of ERG binding sites in VCaP did not reveal AR as a highly scoring motif. To some degree this could be a result of AR not recognising its consensus binding sequence consisting of two half-sites as it commonly recognises its single half-site only

(Denayer et al. 2010). The cooperation and regulation between ERG and AR is well established. ERG has been shown to co-opt AR activity by firstly establishing *de novo* AR binding sites via a direct physical ERG-AR interaction (Chen et al. 2013) and secondly ERG directly disrupts AR expression by inducing repressive chromatin formation at the AR locus (Yu et al. 2010). It is possible that ERG recruits AR by "piggybacking" whereby AR does not bind DNA directly but instead forms a protein-protein interaction with ERG. These mechanisms, resulting from aberrant ERG expression, have been suggested to divert AR activity from normal prostate cell differentiation (Mounir et al. 2016).

This chapter further shows evidence to the likelihood that ERG is not a pioneer TF in endothelial cells. Aberrant ERG expression in prostate cancer cells does not reprogram them to ECs as in the likes of ETV2 in endothelial cells as discussed in chapter 2. For prostate cancer cells, it seems likely that a number of TF candidates including those identified through the motif analysis at ERG sites, are acting to prime the chromatin. The influence of FOXA1 on chromatin accessibility is well established: its structurally similarity to the histone H1 allows it and therefore acts to displace this histone at nucleosome-occluded sites (Cirillo et al. 1998; Cirillo et al. 2002). Acting in concert with FOXA1, the TF GATA2 has been identified as a putative pioneer in the prostate cancer genome (Wang et al. 2007). As a facilitator, GATA2 has been shown to recruit AR to certain enhancers but has also been shown to preclude AR binding at other regions (Wu et al. 2014). However, these actions are more explicitly performed by FOXA1 as it has been determined to act upstream of GATA2 and AR (Zhao et al. 2016). From the work presented in this chapter and many thorough studies in the prostate cancer field it is evident that the cancer is transcriptionally dysregulated by a few or all TFs, namely AR, FOXA1, GATA2, HOXB13 and ERG when aberrantly expressed.

4 TRANSCRIPTION FACTOR COOPERATIVITY INFLUENCES GENE REGULATION IN ENDOTHELIAL CELLS

4.1 Introduction

A defining feature of enhancers is the synergistic and cooperative binding of TFs and co-transcriptional complexes (Banerji et al. 1981). TF can directly activate or repress gene expression at enhancers by the recruitment of co-activators and co-repressors respectively. Colocalisation of TFs and other factors at hotspots have been described previously in other cell types (Gerstein et al. 2012; He et al. 2011). In fact, it has been demonstrated that super-enhancers harbour hotspots for colocalisation in the genome (Adam et al. 2015; Siersbæk et al. 2014).

Determining the TF assembly at functional enhancers in the genome requires exhaustive experimental efforts to unravel the connection between DNA motifs and the cooperative direct and indirect DNA binding of TF and co-regulators. Previous work has focussed on the distribution of TF motifs (Grossman et al. 2018), deciphering direct protein-protein interactions (Morgunova et al. 2017) and changes to chromatin structure (Miraldi et al. 2019). More recently studies using physiochemical assays have reported the formation of molecular condensates, protein-DNA concentrates, that promote compartmentalisation of transcriptional machinery at enhancers (Sabari et al. 2018; Shrinivas et al. 2019).

In EC it has been recognised that ETS factors regulate endothelial-specific gene expression in combination with other TFs. Specifically, GATA factors were shown to function synergistically with ETS factors (De Val et al. 2009) and moreover the abrogation of GATA binding protein 2 (GATA2) has been shown to result in the loss of endothelial-specific markers (Kanki et al. 2011). In another study GATA2 was found to

interact with ETS factor ETV2 during vasculogenesis (Shi et al. 2014). More recently, a study implicated the importance of ETS and AP-1 family members for the selection of endothelial enhancers in HAEC (Hogan et al. 2017). It has also been demonstrated that the ETS TFs can physically interact with AP-1 members cFOS and cJUN (Basuyaux et al. 1997). A study by Niskanen et al. (2018) identified EC-specific chromatin compartments in the genome were enriched in TFs GATA2, cFOS and cJUN.

In this chapter I explore the chromatin signatures that define the function of ECs. I demonstrate the combinatorial assembly at endothelial super-enhancers and determine that TF complex assembly is variable across super-enhancers. I specifically define super-enhancers that are highly dependent on ERG and show less extensive cooperativity. Finally, I reveal clusters based on epigenetic features and link these to functional characteristics of the genome.

4.2 Methods and Data

ChIP-seq data analysis

ChIP-seq fastq data files were obtained and processed as described for ChIP-seq data in chapter 2 and 3. ChIP-seq signals from BigWig files are visualised in the UCSC Genome Browser (https://genome.ucsc.edu).

ChIP-seq signals visualised as a heatmap (Figure 4.1) were generated as follows. First data was normalised to input using the bamCompare tool in deepTools2.0 (Ramírez et al. 2016) and enrichment scores at defined genomic regions were calculated from computeMatrix to generate a heatmap.

Defining super-enhancers

As described in chapters 2 and 3, super-enhancers were detected using the ROSE algorithm (Lovén et al. 2013; Whyte et al. 2013). Active enhancers assigned by H3K27ac/H3K4me1

co-occupancy were stitched if within 12.5 kb of another with exclusion of promoter regions (± 2 kb from TSS). These stitched regions were then ranked by the enrichment of GATA2, cFOS and cJUN.

Overlap between super-enhancer regions was conducted with intersectBed from BEDTools (Quinlan et al. 2010). Heatmap displaying the percentage overlap between super-enhancers defined by TFs was generated in Python package seaborn.heatmap (Figure 4.3).

The cumulative ChIP-seq signal in reads per million (rpm) across shared TF-defined super-enhancers was compared and displayed as pairwise scatter plots plotted using ggplots2 in R (Figure 4.4).

ChIP-qPCR

ChIP experiments were performed using the ChIP-IT express kit (Active Motif). HUVEC transfected with control or ERG siRNA were crosslinked with 1 % formaldehyde. Chromatin was sheared for four cycles of 30s on and 30s off using a Bioruptor ICD-200 ultrasound sonicator (Diagenode). Generated fragments of 200 - 1,000 bp were chromatin immunoprecipitated using 3 µg anti-p300 (ab14984, Abcam) antibodies. The respective negative controls were mouse IgG (12-371, Millipore). RT-qPCR were performed using PerfeCTa SYBR Green Fastmix (Quanta Biosciences) on a Bio-Rad CFX96 system for selected genomic regions on immunoprecipitated DNA using primers listed in Table 4.1. Statistical significance was determined using a two-tailed paired sample t-test in R with a statistical significance cut-off p-value < 0.05. Data were plotted using Prism 8.0 (Graph Pad).

Cluster analysis

Open chromatin regions in the HUVEC genome where identified through DNaseI hypersensitivity sequencing (n = 102,515). The input-normalised ChIP-seq signal from 11 datasets (H3K4me1, H3K27ac, MED1, H3K4me3, H3K27ac, H3K27ac, ERG, GATA2,

Name	Orientation	Oligonucleotide sequences
Gene desert	Forward	TGAATAAGCCAATGAAACAATGACA
	Reverse	TGAAACATAGTATGGGTGGCAACT
CLDN5 E1	Forward	CCGGAAGCCAACTTGGAGTTT
	Reverse	GTGCAGAAGAATGCCCGGAA
CLDN5 E2	Forward	TCCTGCATCCCTGACCACTG
	Reverse	CTGGATGCTGCTCACATCGT
DLL4 E1	Forward	GCAGGTTGAGGGTGAATGGT
	Reverse	TGCCCAAGCACCAGAACTTT
DLL4 E2	Forward	CCCCAGGACCTATCCCAAGT
	Reverse	CACCATTTAGCAGAGCCGGA
DLL4 E3	Forward	GTTTCCTGCGGGTTATTTTT
	Reverse	CTTTCCAAAGGAGCGGAAT
DLL4 E4	Forward	CATGTGGGGGACAGGTAGGA
	Reverse	GCTCCCCATCTAGTGCATCA
IL6 E1	Forward	TGACTGAGCAAACCCATTTTCC
	Reverse	TCCTTATGTGGGAAGGTATGGC
IL6 E2	Forward	AGATTCCTCCACATTTGCCCA
	Reverse	GGCAACTCCAAGCCAGAACA
IL6 E3	Forward	GGTCACGCCACAACTGGAAT
	Reverse	CCATTCCTCACACCCACTGTT
PXN E1	Forward	GCCTCTCACCCCTGCTAATC
	Reverse	TTTGTTCGGGTCTCTGTGGG
PXN E2	Forward	GCATCACGTAGCAACAGAGC
	Reverse	GGTGTTGCTGACACATTCCG
PXN E3	Forward	GGTGGAGTAAAGCGTGAGCA
	Reverse	TGGGTGTAATGCCTGCCTTC

Table 4.1: ChIP-qPCR primers used in this chapter.

cFOS, cJUN, EZH2 and CTCF) was used to generate a matrix with signal calculated every 10 bp bins using computeMatrix in deepTools2.0 (Ramírez et al. 2016). To classify the open chromatin regions I performed K-means clustering using Python function sklearn.cluster.KMeans into 6 clusters. I used the combination of histone modifications to define cluster chromatin states or the high enrichment of a particular factor (eg. CTCF).

Biological functions were assigned to regulatory regions using Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al. 2010) with enriched gene ontology (GO) Biological Processes reported. The significance of enrichment was denoted as a false discovery rate (FDR) Q-value.

Motif analysis

To find the most enriched motif sequences present in open chromatin regions defined to a cluster, HOMER script 'findMotifsGenome.pl' was used (Heinz et al. 2010). The background for this analysis is generated from random sequences in the genome that match the GC content of the query sequences. Reported significance p-values were corrected for multiple testing in R to compare enrichment across the different clusters. The -log(q-value)'s were hierarchically clustered to determine similarity of motif representation in the different clusters. This was performed using Python package seaborn.clustermap with the Euclidean distance metric and an average method.

RNA-seq analysis of ERG-depleted HUVEC

HUVEC were treated by Dr Claire Peghaire (Imperial College London) with AllStars Negative Control siRNA (Qiagen) alongside treatment targeting human ERG using 20 nM siRNA as described in chapter 2 and 3. A total RNA extraction was performed by Dr Peghaire using the RNeasy Mini Kit (Qiagen) and high quality RNA was shipped to Active Motif for polyA+ RNA enrichment, library generation and single-end sequencing using Illumina NextSeq 500. The Active Motif bioinformatics team aligned the data

using STAR (Dobin et al. 2013) to the hg19 genome and in order to perform differential expression analysis they performed count-based analysis using featureCounts (Liao et al. 2014) and normalisation and differential expression using DESeq2 (Love et al. 2014). I use the siCtl treated samples for assessing basal gene expression levels in HUVEC and report them as Fragments Per Kilobase per Million mapped reads (FPKM). I used bam files generated following alignment to the genome to calculate FPKM using Salmon (Patro et al. 2017).

Hi-C analysis

I performed the analysis of Hi-C data generated in HUVEC using the HOMER Suite (Heinz et al. 2010). Firstly, I trimmed the fastq files for restriction enzyme MboI sequence 'GATC' using homerTools. Next, I aligned the trimmed fastq files to reference genome hg19 using Bowtie2 (Langmead et al. 2012). Then, sequencing tag directories were generated from sam files with filtering of PCR duplicates. Hi-C interactions were then assigned using the analyzeHiC on the tag directory where normalisation is also performed based on region coverage and distance between the regions. Significant interactions were assigned by setting a threshold on the p-value to 0.001. I overlapped the significant chromatin interactions in HUVEC with genomic regions identified from the cluster analysis using the intersectBed tool from BEDTools (Quinlan et al. 2010). I calculate the significance of overlap using a hypergeometric distribution test where the number of significant Hi-C interactions in a given cluster (from all significant interactions) is compared to the total number of any sequenced Hi-C interactions in a given cluster (from all interactions).

Data

All data used in this chapter is listed in Table 4.2.

Experiment	Factor	Cell type	Condition	GEO ID	Citation
ChIP-seq	ERG	HUVEC	untreated	GSM3557980	Kalna et al. 2019
DNaseI hyper -sensitivity		HUVEC	untreated	GSM816646	ENCODE/Broad Institute
ChIP-seq	H3K27ac	HUVEC	untreated	GSM733691	ENCODE/Broad Institute
ChIP-seq	H3K4me1	HUVEC	untreated	GSM733690	ENCODE/Broad Institute
ChIP-seq	H3K27me3	HUVEC	untreated	GSM733688	ENCODE/Broad Institute
ChIP-seq	H3K27ac	HUVEC	siCtl	GSM3557982	Kalna et al. 2019
ChIP-seq	H3K27ac	HUVEC	siERG	GSM3557983	Kalna et al. 2019
ChIP-seq	MED1	HUVEC	siCtl	GSM3557984	Kalna et al. 2019
ChIP-seq	MED1	HUVEC	siERG	GSM3557985	Kalna et al. 2019
ChIP-seq	GATA2	HUVEC	untreated	GSM935347	Linnemann et al. 2011
ChIP-seq	cFOS	HUVEC	untreated	GSM935585	Linnemann et al. 2011
ChIP-seq	cJUN	HUVEC	untreated	GSM935278	Linnemann et al. 2011
ChIP-seq	EZH2	HUVEC	untreated	GSM1003518	ENCODE/Broad Institute
ChIP-seq	CTCF	HUVEC	untreated	GSM733716	ENCODE/Broad Institute
RNA-seq		HUVEC	siCtl v siERG	N/A	unpublished
HiC		HUVEC	untreated	GSM1551631	Rao et al. 2014

Table 4.2: Data used in this chapter.

4.3 Results

4.3.1 Transcription factor co-assembly at ERG binding loci

Given the central role of ETS TFs in combinatorial protein assembly in ECs, I investigated the TF collaboration at sites bound by the most highly expressed ETS factor ERG in the HUVEC genome. The identification of ERG binding sites across the HUVEC genome provided genomic loci that exhibit co-transcriptional assembly for the direct regulation of endothelial gene expression. By analysing ChIP-seq data for TF GATA2 and for AP-1 family members cFOS and cJUN, I found that these occupied many of the same loci as ERG (Figure 4.1). This indicates that these four TFs are in physical proximity and functionally could be influencing gene expression.

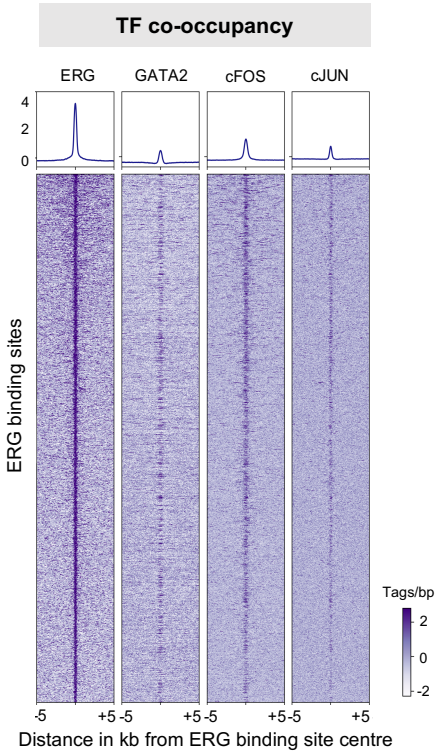


Figure 4.1: Transcription factor enrichment at ERG occupied regions in HUVEC. Heatmap displaying the ERG ChIP-seq tag density per bp and relative co-occupancy of GATA2, cFOS and cJUN TFs from the centre ± 5 kb of all ERG binding sites. Each row is an ERG-bound loci in HUVEC.

4.3.2 Transcription factor identification of super-enhancers in HUVEC

Super-enhancer control of cell-type-specific processes is tightly regulated by the dense occupancy of TFs and co-factors. In chapter 2, I identified super-enhancers in HUVEC using the enrichment of ERG. Here, I further used ChIP-seq signal for collaborative TFs GATA2, cFOS and cJUN and defined super-enhancers by their respective enrichment (Figure 4.2A, B and C). The top ranked super-enhancer-associated gene identified by ERG, cFOS and cJUN was *VWF*, which is almost entirely restricted to ECs. For GATA2-defined super-enhancers the most highly ranked candidate was associated with *TNFSF18*, a gene much less restricted to ECs. The top 10 super-enhancers identified by the four TFs are listed in Table 4.3. Within the top 10 super-enhancer-associated genes, shared gene candidates are more explicit between ERG, cFOS and cJUN whereas GATA2 identifies a more distinct set of top 10 super-enhancer genes. A comparison between all TF-identified super-enhancers identifies 490 shared super-enhancers; accounting for 44 % (490 / 1125) of all ERG-defined super-enhancers (Figure 4.2D).

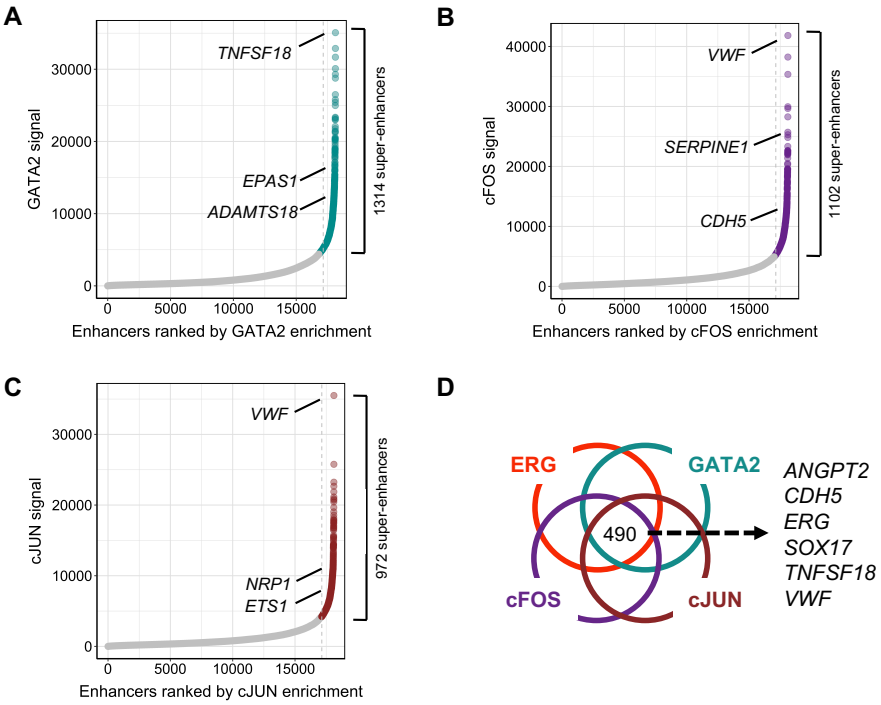


Figure 4.2: Transcription factor defined super-enhancers in HUVEC. Super-enhancers identified by the ChIP-seq signal from TFs (A) GATA2, (B) cFOS and (C) cJUN. Super-enhancers are highlighted by colour and the total number of super-enhancers are indicated in each plot. Select super-enhancer-associated genes highlighted if the gene is the highest ranked among the three defined super-enhancer sets shown here. (D) Schematic of shared super-enhancers defined by ERG, GATA2, cFOS and cJUN. 490 super-enhancer genes were in common between the four TF-defined super-enhancer subsets. Shared and highly ranked super-enhancer genes are listed. ERG super-enhancers were identified in chapter 2.

Rank	ERG (1125)	GATA2 (1314)	cFOS (1102)	cJUN (972)
1	*VWF*	*TNFSF18*	*VWF*	*VWF*
2	*LOC100288911*	*DNMBP-AS1*	*SMAD3*	*SMAD3*
3	*GADD45A*	*TGM2*	*PVT1*	*PVT1*
4	*AFAP1*	*XXYLT1-AS2*	*SMURF2*	*XXYLT1-AS2*
5	*WWTR1-AS1*	*LOC100288911*	*XXYLT1-AS2*	*GADD45A*
6	*XXYLT1-AS2*	*CDK17*	*CAPN2*	*LOC100288911*
7	*SMAD3*	*LINC01541*	*SERPINE1*	*WWTR1-AS1*
8	*TMEM212*	*WWTR1-AS1*	*WWTR1-AS1*	*MTHFD1L*
9	*PVT1*	*LIPG*	*SYNJ2*	*TMEM212*
10	*TGM2*	*SMURF2*	*DNMBP-AS1*	*SMURF2*

Table 4.3: Genes associated with the top 10 super-enhancers identified by TFs ERG, GATA2, cFOS and cJUN. Total number of super-enhancers indicated in brackets.

Moreover, by comparing super-enhancers pairwise I determined the percentage overlap between regions identified by the TFs (Figure 4.3). Unsurprisingly, super-enhancers identified by cFOS and cJUN are highly similar, sharing a total of 81 % of all regions. ERG super-enhancers are also highly concordant with cJUN identified regions, sharing 70 %. Overlap between GATA2 and the other TF-identified super-enhancers showed that no more than 62 % of super-enhancer regions are in common, as observed with cJUN. Direct pairwise correlations of the cumulative TF enrichment at shared super-enhancer regions showed comparable enrichment levels (Figure 4.4). Higher Pearson's correlation coefficients were reported between ERG, cFOS and cJUN at shared super-enhancers, with ERG-cJUN shared super-enhancers showing the highest pairwise correlation of 0.84. GATA2 enrichment displays slightly more divergence at shared super-enhancers with Pearson's correlation values of 0.67, 0.56 and 0.60 for ERG, cFOS and cJUN comparisons respectively. Altogether this suggests that GATA2 identifies a more independent subset of super-enhancers compared with ERG and AP-1 TF members.

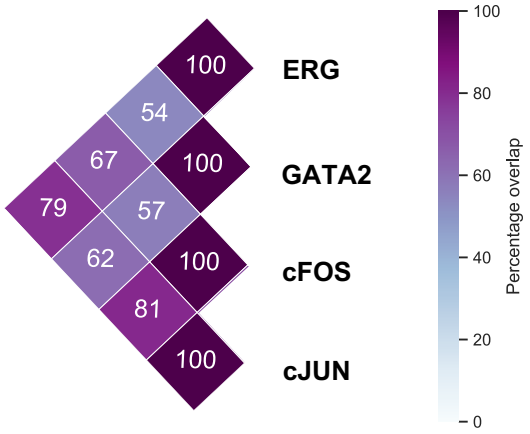


Figure 4.3: Comparison between super-enhancers defined by transcription factor enrichment. Matrix reporting the pairwise percentage overlap between TF-defined super-enhancers in HUVEC. Percentage overlap is determined by the number of overlapping super-enhancers over the total number of super-enhancers identified by the TF in the denominator. ERG super-enhancers were identified in chapter 2.

4.3.3 Transcription factor cooperativity controls transcriptional complex assembly in HUVEC

Genome-wide profiling of histone modification marks and co-activators in HUVEC provides a proxy for which the basal chromatin structure can be assessed. By examining the increasing colocalisation of the EC TFs ERG, GATA2, cFOS and cJUN with the levels of active histone mark H3K27ac and co-activator subunit BRD4, I found a positive correlation. The greater the TF co-occupancy, from 1 to 4 TFs, the higher the H3K27ac and BRD4 levels (Figure 4.5A and B). Here a total of four TFs are included in the analysis; however, a limitation of the analysis is the availability of other TF ChIP-seq data which are not accounted for in a particular TF group. Nevertheless, the trend is indicative of combinatorial assembly whereby the transcriptional complex exhibits a range of compositions at different genomic loci.

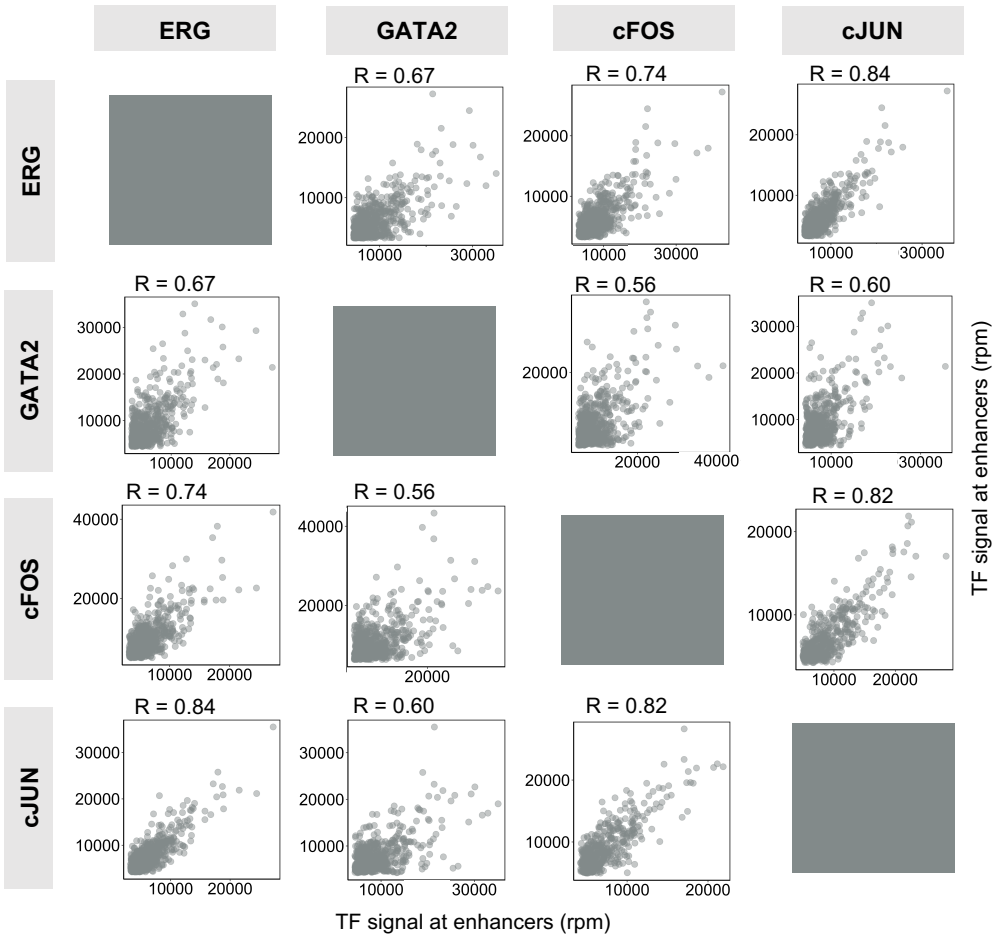


Figure 4.4: Pairwise correlations of shared super-enhancers defined by four transcription factors. TF signal in reads per million (rpm) of the super-enhancers shared between all pairwise comparisons. Each point on the plot is one super-enhancer. Pearson correlation coefficient's are indicated above each plot.

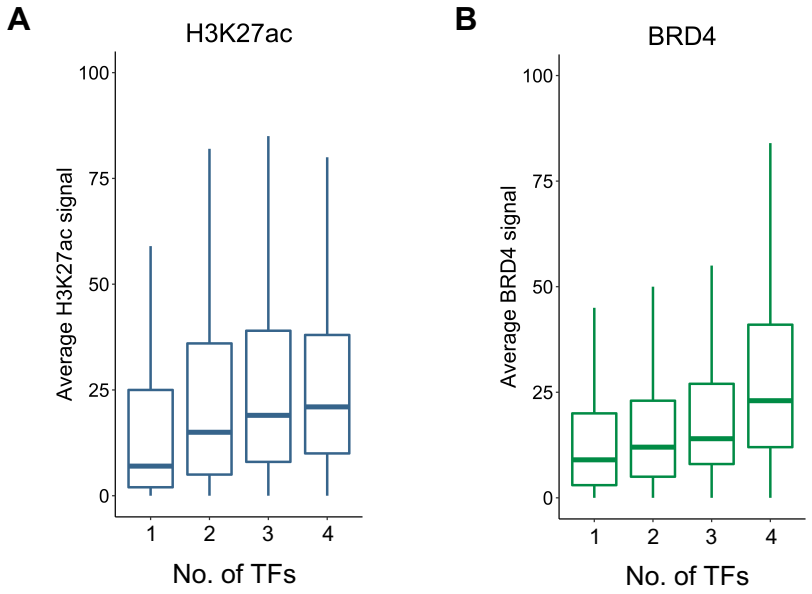


Figure 4.5: Transcription factor cooperativity mediates the binding intensity of H3K27ac and BRD4. Boxplots represent the distribution of the average ChIP-seq signal for (A) H3K27ac and (B) BRD4 in genomic sites that are bound by 1, 2, 3 and 4 TFs of the assessed factors ERG, GATA2, cFOS and cJUN.

4.3.4 ERG regulates a distinct set of endothelial super-enhancers

In chapter 2, I determined that genes activated by ERG were positively correlated with super-enhancer-associated genes (Figure 2.4B). The regulation of gene expression is not directed by ERG alone but by the cooperative network of endothelial TFs. I reported a subset of endothelial super-enhancers that were dependent on ERG (Figure 2.11A). A schematic to illustrate the finding is shown in Figure 4.6A. This subset consisted of 107 super-enhancer that displayed a significant decrease in H3K27ac level when ERG was depleted in HUVEC (*decreased* super-enhancers). However, the majority of super-enhancers remained unmodulated following ERG depletion (*constant* super-enhancers). Therefore I carried out analysis to investigate the distinct properties of this ERG-dependent subset of super-enhancers.

The histone acetyltransferase (HAT) p300 deposits acetyl marks on histone H3K27 and ERG has been shown to recruit p300 in EC (Fish et al. 2017). I investigated

whether the recruitment of p300 by ERG delineates the ERG *decreased* super-enhancers from *constant* super-enhancers. Using ChIP-qPCR, the enrichment of p300 at enhancers across selected *decreased* and *constant* super-enhancers was examined in control and ERG-deficient HUVEC. At ERG targets *CLDN5* and *DLL4*, classified as *decreased* super-enhancers, p300 levels were significantly downregulated following ERG inhibition (Figure 4.6B). However, ERG inhibition did not consistently affect p300 recruitment at *constant* super-enhancers *IL6* and *PXN* (Figure 4.6B). This would suggest that ERG is required for p300 recruitment at *decreased* super-enhancers, whereas ERG is functionally less involved in p300 recruitment at *constant* super-enhancers. Since p300 is recruited by a plethora of other TFs and cofactors (Goodman et al. 2000), including ERG interacting TF cJUN (Kim et al. 2014), the possibility that other TFs could compensate for the role of ERG at *constant* super-enhancers was explored. By analysing the enrichment of ERG, GATA2, cFOS and cJUN at regions of *decreased* and *constant* super-enhancers I determined that the occupancy of all these TFs was significantly higher in the *constant* super-enhancers relative to the *decreased* enhancers (Figure 4.6C).

By inspection of the selected super-enhancer loci of *CLDN5*, *DLL4*, *IL6* and *PXN*, occupancy of key EC TFs show disparate profiles. At the *CLDN5* super-enhancer, collaborative TFs ERG and GATA2 but not cFOS and cJUN are evident; at the *DLL4* super-enhancer ERG, cFOS and cJUN but not GATA2 are present (Figure 4.7A). In comparison, representative loci of *constant* super-enhancers *IL6* and *PXN* are occupied by all four TFs (Figure 4.7B). In addition, the changes to H3K27ac and MED1 occupancy at these super-enhancers further reveals the loci-specific modulations to chromatin following ERG inhibition.

Overall these data suggest a model of two super-enhancer assembly states in HUVEC. The cooperative TF network can be ERG-dependent whereby ERG is functionally critical and loss of ERG is not compensated due to the low cooperativity between ERG and the other TFs. The majority of super-enhancers show high TF cooperativity via the formation of a large and strong transcriptional complex which can compensate function-

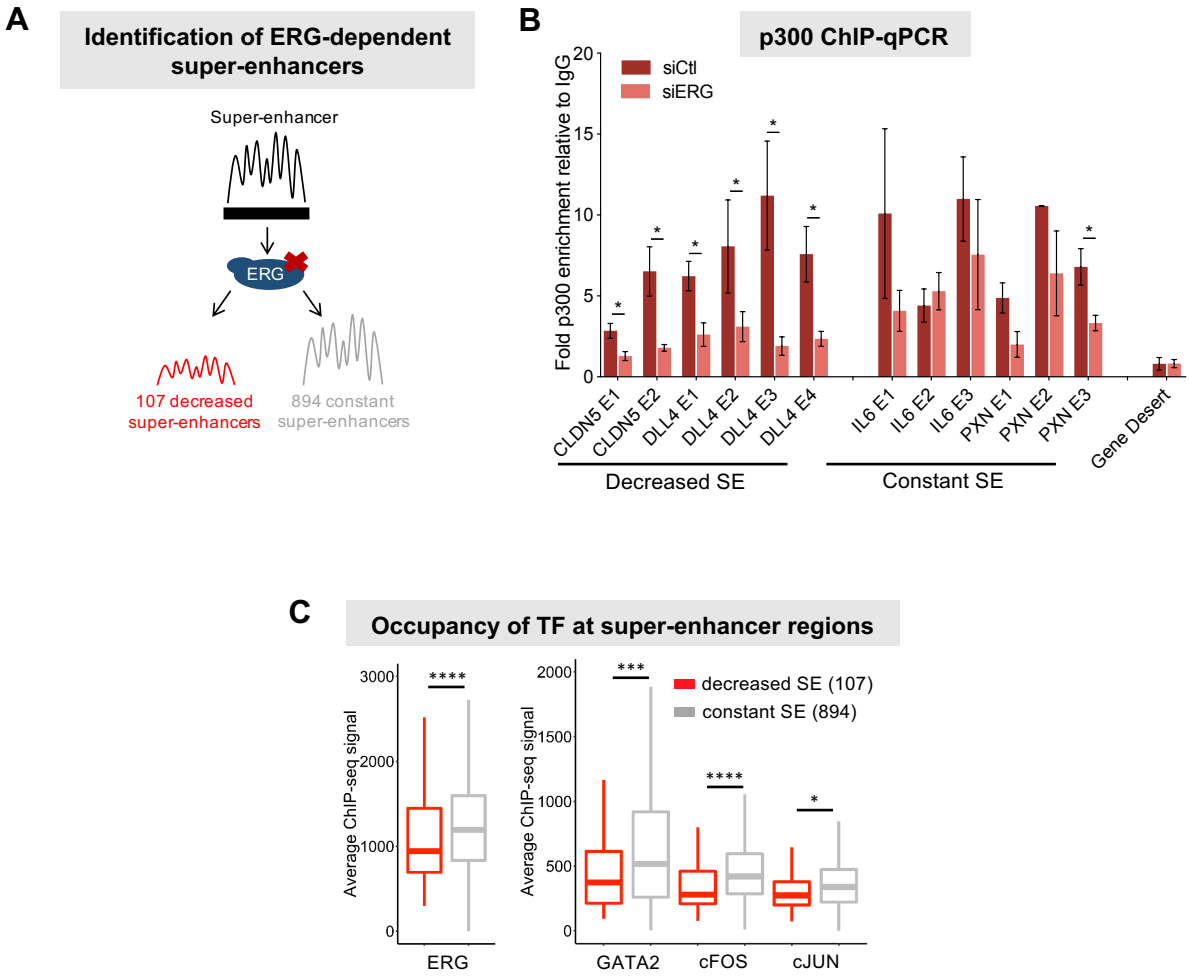


Figure 4.6: Mechanisms of ERG-dependent super-enhancer regulation. (A) Schematic model of the two groups of super-enhancers regulated by ERG. 107 decreased and 894 constant super-enhancers defined by H3K27ac are identified following ERG inhibition. (B) ChIP-qPCR for p300 at constituent enhancer regions of decreased super-enhancers *CLDN5* and *DLL4* and constant super-enhancers *IL6* and *PXN*. Data are represented as fold change over IgG, n=4, *p-value < 0.05, paired 2-tailed *t*-test. (C) Genomic occupancy of ERG and collaborative TFs GATA2, cFOS and cJUN reported as the average ChIP-seq signal across decreased and constant super-enhancers. ****p-value < 0.0001, ***p-value < 0.001 and *p-value < 0.05, Wilcoxon rank-sum test.

ally for the depletion of ERG (Figure 4.8).

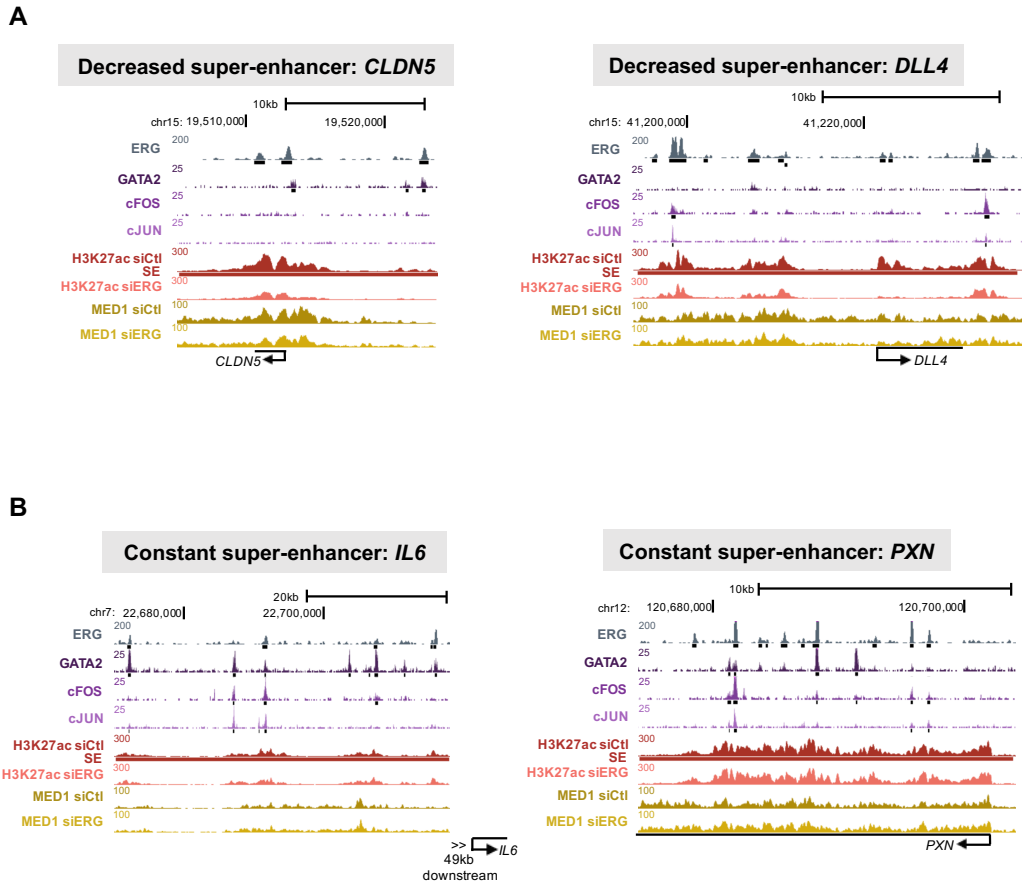


Figure 4.7: Transcription factor and super-enhancer marks at super-enhancer subsets. ChIP-seq profiles of (A) decreased super-enhancers *CLDN5* and *DLL4* and (B) constant super-enhancers *IL6* and *PXN*. Occupancy of TFs ERG, GATA2, cFOS and cJUN and super-enhancer marks H3K27ac and MED1 in control (siCtl) and ERG-deficient (siERG) HUVEC are indicated. Genes are annotated below the profiles. Enrichment of ChIP-seq binding for each factor is highlighted on the profiles.

4.3.5 Unsupervised clustering using ChIP-seq data reveals clusters of chromatin states in HUVEC

The availability and generation of ChIP-seq data in basal HUVEC provides a resource for investigating the combinatorial signature of the transcriptional complex and chromatin modifications. Here I assessed the histone modifications, TFs and co-factors

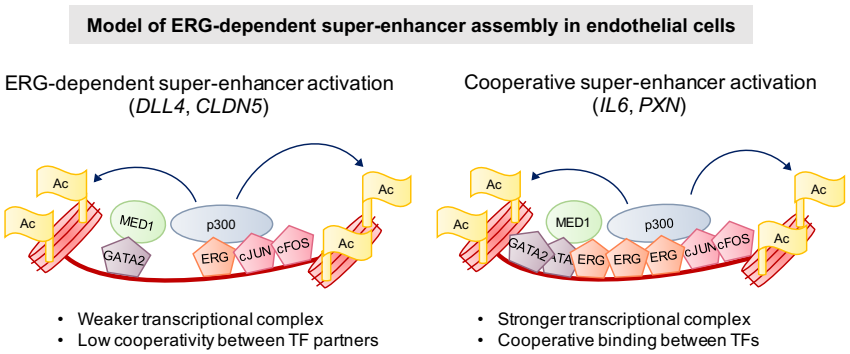


Figure 4.8: Model of ERG-dependent super-enhancer assembly. In a subset of super-enhancers ERG functionality is critically important as reduced cooperativity with other TFs forms a weaker transcriptional complex due to low abundance of all factors. In the super-enhancers which are unperturbed by ERG deficiency a robust transcriptional complex with high TF cooperativity network is established that can compensate functionally by the action of other TFs.

in open chromatin regions to determine functional similarity by adopting unsupervised k-means clustering to segregate the epigenetic states of the chromatin into clusters. The 11 ChIP-seq datasets exploited for this task were histone modifications H3K4me1, H3K27ac, H3K4me3 and H3K27me3, TFs ERG, GATA2, cFOS and cJUN and co-factors MED1, EZH2 and CTCF. Using DNaseI hypersensitivity sequencing from HUVEC, 102,515 open chromatin regions were defined. By setting the cluster number to 6 and centring all ChIP-seq signals on the DNaseI hypersensitive regions ± 1 kb in HUVEC, I found defined chromatin states by annotating the clusters based on the enrichment of particular factors (Figure 4.9). The clusters were defined as follows: (1) class I enhancers: combination of H3K4me1 and high H3K27ac and presence of MED1, (2) promoters: co-occupancy of H3K4me3 and H3K27ac and occupancy of MED1, (3) bivalent chromatin: presence of repressive mark H3K27me3 and of enzymatic PRC2 component EZH2, with low-level H3K4me3 or some H3K4me1, (4) class II enhancers: narrow H3K4me1 with low H3K27ac, (5) CTCF bound: solely bound by CTCF and (6) poised enhancers: weak H3K4me1 with no H3K27ac (Creyghton et al. 2010).

The clusters evidently segregate into these epigenetic states and interestingly reveal both expected and unexpected concepts. All four TFs, ERG, GATA2, cFOS and cJUN, are most enriched at class I enhancers; accordingly, the four TFs show lower occupancy

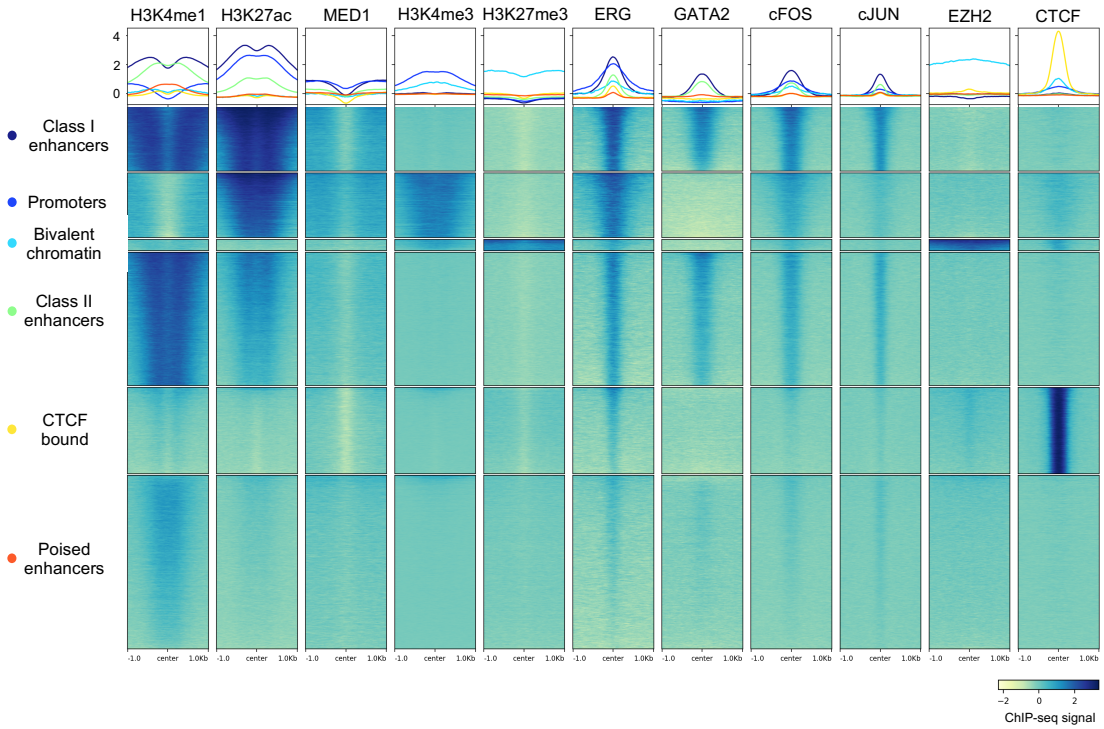


Figure 4.9: Identification of HUVEC chromatin states by unsupervised K-means clustering. ChIP-seq enrichment of 11 factors from HUVEC; 4 histone modifications (H3K4me1, H3K27ac, H3K4me3 and H3K27me3), 4 TFs (ERG, GATA2, cFOS and cJUN) and 3 co-factors (MED1, EZH2 and CTCF) was used to perform k-means unsupervised clustering to determine epigenetic states in open chromatin defined by DNaseI hypersensitivity in HUVEC. The number of clusters was pre-specified to 6. The cluster heatmap is centred on the open chromatin regions ±1 kb.

at class II enhancers as compared to class I enhancers, in line with the considerably lower H3K27ac signal at this cluster. The signal of all four TFs further decreases in the poised enhancer cluster. On promoter regions, ERG and cFOS show notable occupancy; however cJUN shows much lower enrichment and notably GATA2 is absent from promoters.

At CTCF bound sites ERG, cFOS and cJUN show low occupancy whereas GATA2 is depleted. This suggests that ERG and AP-1 may be partly involved in mediating high-order chromatin interactions.

Moreover, in the bivalent chromatin cluster there is evidence of an interplay between EZH2 repressed chromatin (depositing H3K27me3) and CTCF which regulates contact between promoters and enhancers. These are likely to be genes that are held in a looped, poised conformation and marked by H3K27me3 via EZH2 methyltransferase activity of the PRC2 complex, as has been suggested in ESCs (DeMare et al. 2013). This bivalent chromatin is also occupied by low levels of ERG, cFOS and cJUN, indicating a regulatory topology of HUVEC chromatin driven by chromosomal looping, PRC2 repression and TFs.

To determine whether any of the clusters are related to common biological processes, functional analysis of the genes associated with the genomic regions of each cluster would determine gene ontology terms. However, the size of each cluster is very large (eg. class I enhancers have 12,388 regions). This exceeds the limit for meaningful biologically pathway annotation, with more general biological processes likely to be enriched. The exception to this is the bivalent chromatin cluster which contained the smallest number of regions (2,131 regions) and was therefore suitable for gene ontology pathway analysis. Interestingly, the analysis revealed the functions of this cluster to be associated with developmental processes such as regionalisation, morphogenesis and fate commitment (Table 4.4). This epigenetic state is common in ESCs but are far less frequent in differentiated cells (Bernstein et al. 2006). These regions are likely to be poised in a

bipotential state between activation and repression. Thus, HUVEC chromatin contains regions with epigenetic memory related to their differentiation cascade.

Biological Process	FDR Q-Value
regionalisation	1.18E-18
embryonic organ morphogenesis	8.58E-15
skeletal system morphogenesis	8.93E-12
neuron fate commitment	1.10E-10
anterior/posterior pattern specification	3.20E-09

Table 4.4: Biological processes associated with regions linked to bivalent chromatin. Top 5 significant biological processes listed with significance reported as false discovery rate (FDR) Q-value using a binomial test over the genomic regions.

4.3.6 Motif analysis in HUVEC chromatin clusters

To infer the role of other TFs not represented in the clusters, I performed a motif analysis to identify putative TF binding sites in each of the 6 clusters. Enrichment of motifs for the same top 5 putative TF families was evident for all 3 classes of enhancers (class I enhancers, class II enhancers and poised enhancers) with AP-1 and ETS family motifs very highly enriched across the enhancer classes, followed by, SOX, GATA and FOXO family motifs (Figure 4.10). The motifs over-represented in enhancer regions are largely distinct from the other identified chromatin states including the promoter regions (Figure 4.11). The most highly represented TF family in promoter regions is the ETS motif followed by SP1, NFY, NRF and KLF family motifs. Asides from the ETS motif the sequence composition of promoters is largely different from all enhancers (Figure 4.11). The absence of GATA2 signal at promoters coincides with the lack of enrichment of its consensus motif (Figure 4.11). Notably, SP1 and KLF motifs are highly concordant in sequence composition, suggesting the potential for binding redundancy. Finally, at bivalent chromatin the most highly over-represented motif was for TF CTCF which validates the binding of CTCF in this cluster. Unsurprisingly the CTCF motif was the most enriched in the CTCF bound cluster (Figure 4.10 and 4.11). Thus, the chromatin clusters defined by unsupervised methods show distinct DNA sequence composition.

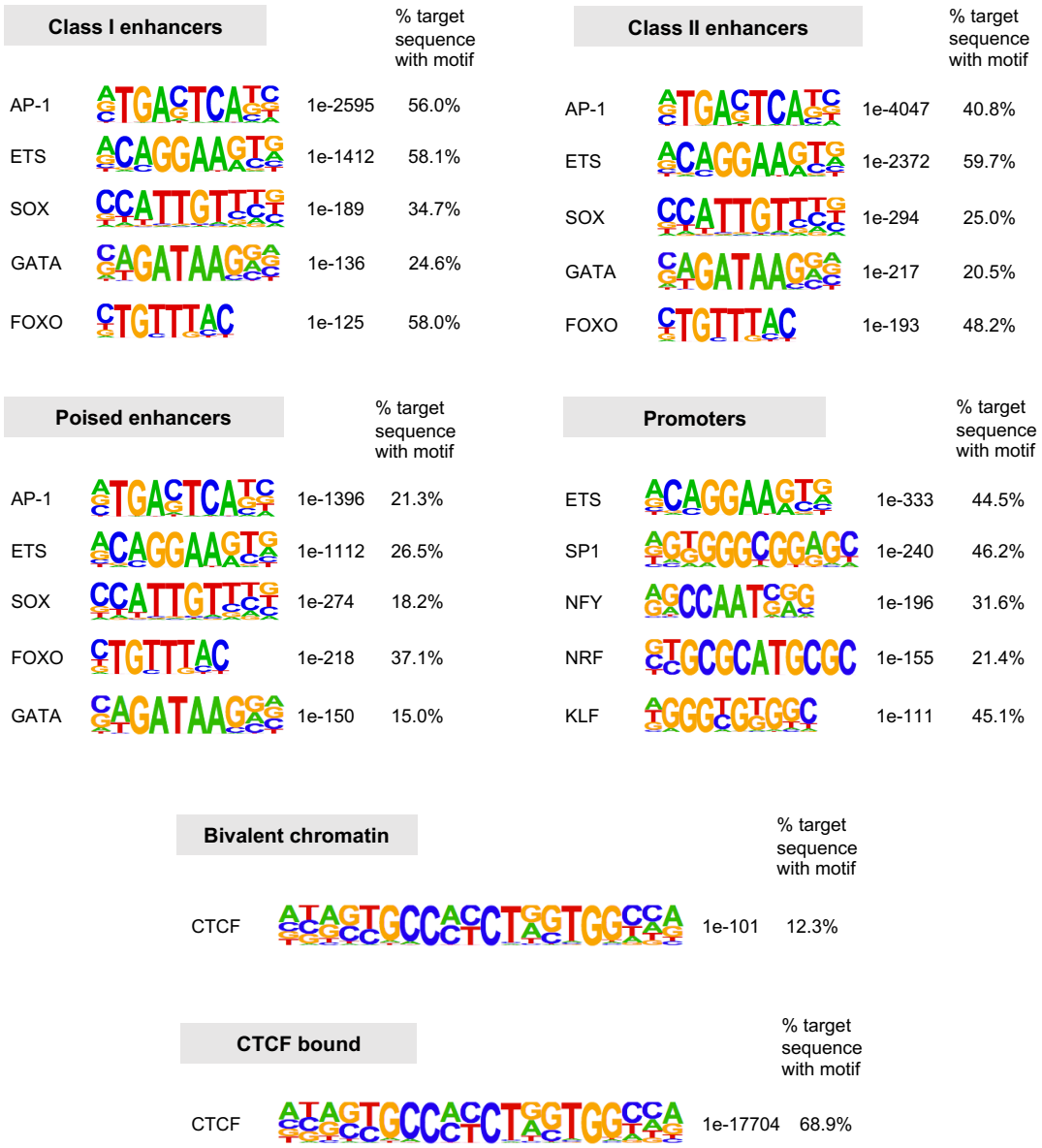


Figure 4.10: Motif enrichment analysis in different endothelial chromatin states. The 5 most over-represented motif families in all 6 clusters are summarised for clusters of enhancer classes and promoters. The single most over-represented motif family is reported for bivalent chromatin and CTCF bound regions as this was exceedingly more enriched than any other motifs. The significance of motif enrichment is reported as a p-value. Percentage of regions with the indicated motif is also reported.

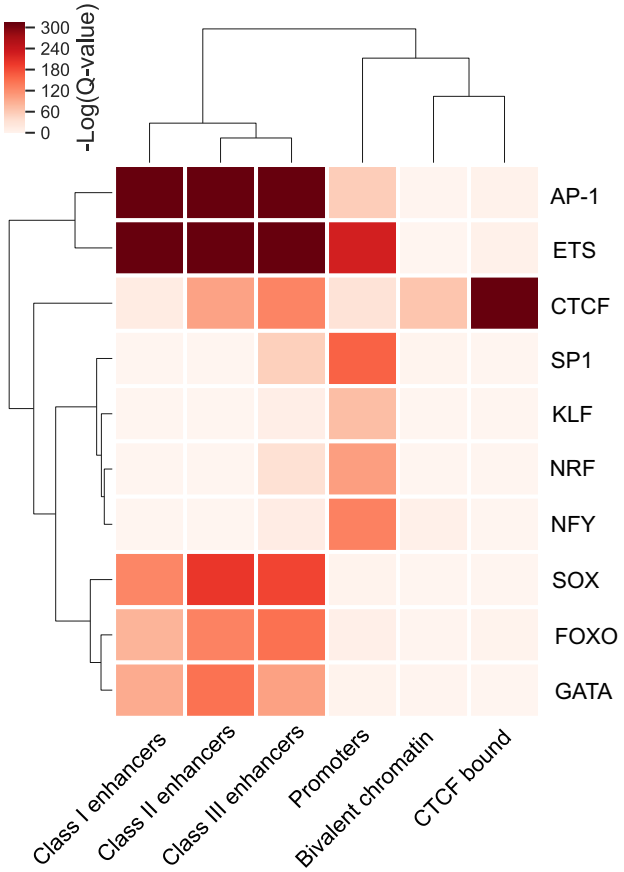


Figure 4.11: Comparative motif analysis of putative transcription factor binding. Matrix of motif enrichment across the 6 identified clusters for TF families found to be in the top 5 of the most significant motifs. Hierarchical clustering shows that motif enrichment in the 3 enhancer classes are most similar. The promoter cluster shows enrichment for an unique set of motifs. The bivalent chromatin and CTCF bound clusters both show over-representation of the CTCF consensus sequence. Enrichment is reported at a -log(q-value) for comparison across the different clusters.

4.3.7 Gene expression and ERG-dependent regulation in HUVEC chromatin clusters

Having identified the putative chromatin state of the different clusters, the functionality of the clusters was addressed. It was postulated that the gene expression levels in the different clusters would correlate with the chromatin state. Gene expression levels obtained from RNA-seq analysis of basal control HUVEC were assessed and the expression levels were identified to be representative of the different chromatin states. The promoter-associated genes demonstrated the highest expression levels, followed by the genes associated with the class I and class II enhancers (Figure 4.12A). In fact, the three enhancer classes show a significant increase in gene expression levels as the class of the enhancer increases from poised enhancers to class I enhancers (Figure 4.12B). The bivalent chromatin cluster was shown to have almost negligible gene expression with an average of 5.0 fragment per kilobase of transcript per million (FPKM) showing that many of these genes have low counts (Figure 4.12A). This is not surprising given the poised and inactive state of bivalent chromatin.

Next, given the occupancy of ERG across all chromatin clusters, albeit at different levels, it was postulated that ERG may specifically drive or repress gene expression in a specific cluster. By assessing the change in gene regulation for all genes associated with each cluster, it was determined that there was no link between ERG gene regulation and any cluster. Each cluster shows a distribution of up- and down-regulated genes (Figure 4.13). Therefore, ERG is likely to be involved in gene activation and repression across the endothelial genome.

4.3.8 Chromatin interactions are mediated by enhancer activity levels

In chromatin organisation enhancers are involved in sub-domain interactions within topologically associated domains (TADs). To evaluate whether the 3 enhancer classes

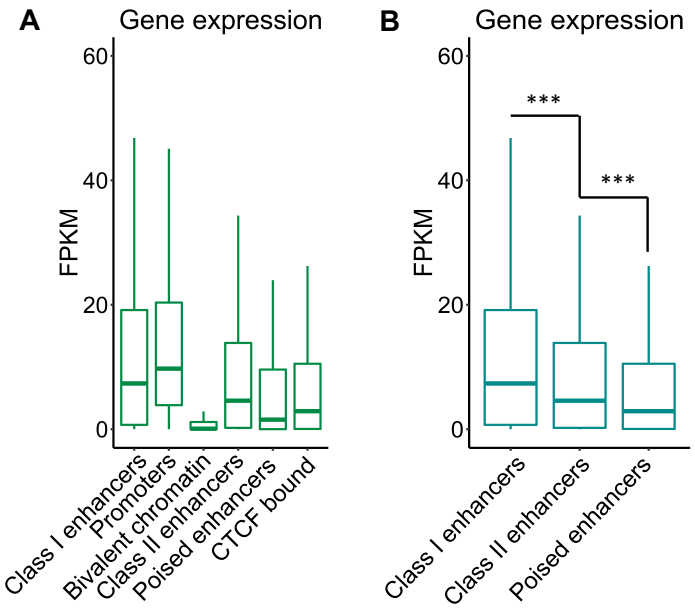


Figure 4.12: Chromatin states are linked to gene expression. (A) Boxplot of gene expression levels acquired from a RNA-seq analysis in control HUVEC in the 6 different clusters. (B) Comparison of the gene expression levels in the 3 enhancer types show significantly higher expression levels in progressively higher enhancer classes. ***p-value < 0.001, Wilcoxon rank-sum test. FPKM = Fragments per kilobase per million

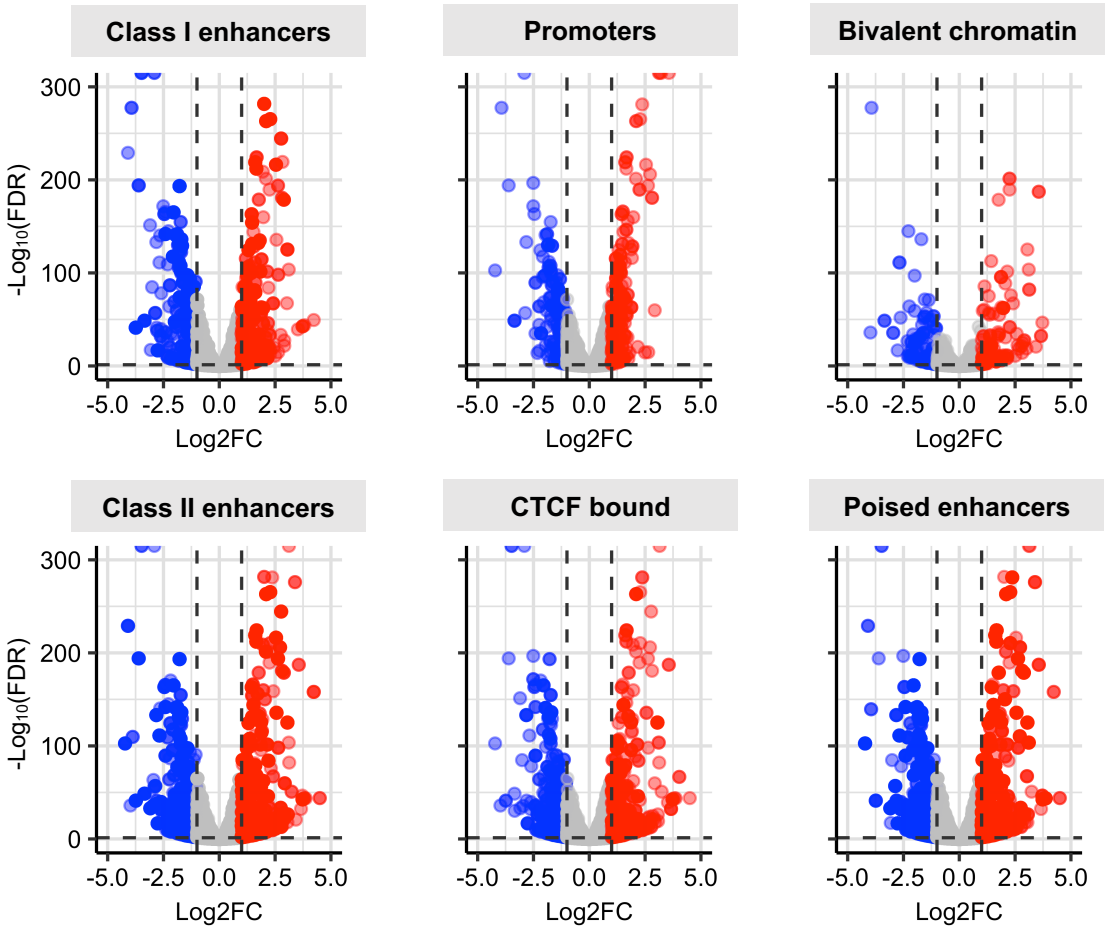


Figure 4.13: Epigenetic states are not associated with ERG-dependent gene regulation. Volcano plots showing the $\log_2$FC of all genes annotated to each cluster following ERG inhibition in HUVEC. Significantly upregulated genes are shown in red and downregulated genes in blue (-1 $\geq \log_2\text{FC} \geq 1$; $-\log_{10}(\text{FDR}) > 1.3$).

are functionally prevalent in chromatin structure assembly, I investigated the chromatin interactions by analysing Hi-C data from HUVEC from Rao et al. (2014). By obtaining the frequency of significant Hi-C interactions (p-value < 0.001) that mapped to the three enhancer classes, I determined that the highest frequency of chromatin contacts were at class I enhancers and the lowest frequency at poised enhancers (Figure 4.14). Notably, interactions in all enhancer classes were significant. This suggests that enhancers in HUVEC are mediating the formation of intra-domain contacts and enhancer activity determines the frequency of interactions.

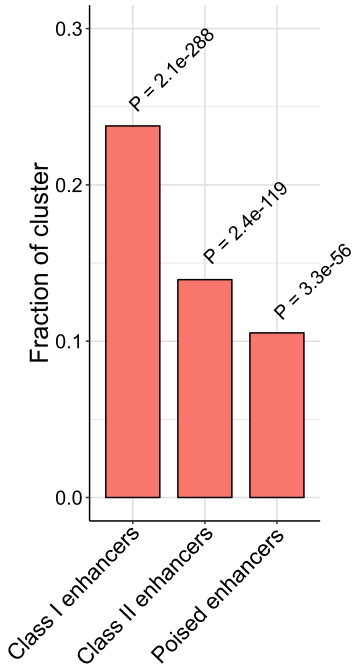


Figure 4.14: Chromatin contacts in different endothelial enhancer classes. The fraction of each enhancer class cluster involved in a significant chromatin interactions as determined by Hi-C analysis in HUVEC. The significance of the frequency of interactions in each enhancer cluster is evaluated by hypergeometric tests with p-values of these tests reported above each bar.

4.4 Discussion

Deciphering the regulatory code of HUVEC requires many high-throughput datasets to build a complete gene regulatory network. In this chapter, I present the identification of endothelial super-enhancers using the enrichment TFs GATA2, cFOS and cJUN to complement the identification of ERG super-enhancers in chapter 2. I compare these super-enhancers to determine the interconnected relationship between the four TFs, and I also determine a correlation between TF co-assembly and cofactor recruitment. I provide a model for the subset of H3K27ac super-enhancers sensitive to ERG loss linked to histone acetyltransferase p300. Finally, I define epigenetic clusters of all open chromatin regions in HUVEC which display distinct combinatorial TF co-assembly and have functional associations to gene expression and chromatin conformation. In this chapter I extend the notion that super-enhancers can be identified by the enrichment of TFs. I defined endothelial super-enhancers with GATA2, cFOS and cJUN and with comparison to ERG super-enhancers I revealed a large proportion of shared super-enhancers. The interdependency between TFs at super-enhancers has previously been described in a model of core regulatory circuitry (Saint-André et al. 2016). This model suggests that core TFs in a cell type have three key properties: (1) they are encoded by a super-enhancer-associated gene, (2) bind their own super-enhancer and (3) form an auto-regulatory loop interconnected with the other TFs. Saint-André et al. (2016) investigate 84 cell types and predicting TF occupancy by sequence motif analysis. The study also included HUVEC, where 18 TFs were indicated as part of the HUVEC core regulatory circuitry including ERG and cJUN. Importantly, cFOS is the one TF of the four that does not have a super-enhancer defined by either H3K27ac, MED1 or any TF from this work. To determine the regulation of expression suggested by the auto-regulation concept of the core regulatory circuitry model, I observed that GATA2 and cJUN were significantly up-regulated from analysis of ERG-dependent gene expression in HUVEC. A schematic of the core regulatory circuitry of the four TFs is shown in Figure 4.15. This suggests that the auto-regulatory relationship between the TFs may serve to dampen the expression

of co-TF partners and thus exert another level of gene expression control.

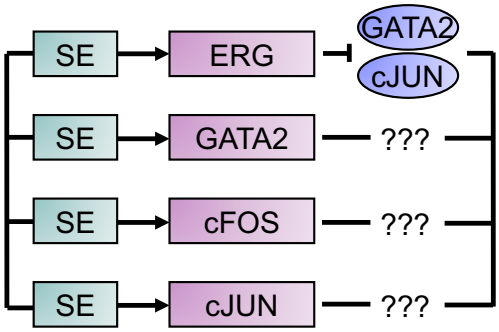


Figure 4.15: Diagram showing the putative regulation of the four transcription factors ERG, GATA2, cFOS and cJUN in HUVEC. From RNA-seq performed in ERG-deficient HUVEC, ERG regulates GATA2 and cJUN by repressing their gene expression levels. Regulation by other TFs is not known. ChIP-seq data shows evidence for binding of all four TFs; ERG, GATA2, cFOS and cJUN, at super-enhancers assigned to these TFs. These are predicted to regulate the levels of each other by core regulatory circuitry.

I further reveal that an increase in TF occupancy is correlated with the occupancy of active histone H3K27ac and co-activator BRD4 (Figure 4.5A and B). This result suggests that TF complex assembly is associated with epigenetic activity of the chromatin. It also supports the billboard model of enhancer activity introduced in section 1.1.2 which states that binding of TFs is additive to enhancer activity.

Furthermore, I propose a mechanism by which ERG-dependent super-enhancers are regulated. Acetylation of H3K27 via p300 and lower TF cooperativity results in the dampening of acetylation upon the loss of ERG as it can not be compensated by other TFs. The deposition of the acetyl group on H3K27 is controlled by the general coactivator p300 in many cell types and in this chapter I suggest that the H3K27ac levels are tightly regulated by the activity of p300. Indeed, other transcriptional co-activators also contribute to the activity of p300. In other cell types p300 has been shown to physically interact with remodelling complex BAF (SWI/SNF) (Huang et al. 2003). ERG in prostate cancer VCaP cells has been shown to interact with BAF (Sandoval et al. 2018). In mouse fibroblast cells, a study reported that ETS motifs were enriched at BAF-dependent H3K27ac enhancers (Alver et al. 2017). There is no evidence of

an interaction between ERG and BAF in endothelial cells and therefore the specific arrangement and function of the transcriptional complex inclusive of ERG, p300 and BAF at endothelial enhancers could be an interesting research avenue.

At *constant* super-enhancers, super-enhancers that were not significantly decreased following ERG depletion, the p300 levels in ERG-depleted cells also showed a trend towards a decrease at constituent enhancers of *IL6* and *PXN*. This seems to suggest that p300 is not the only histone acetyltransferase that is responsible for the deposition of the acetyl group on H3K27. There are other HAT family members that specifically target histone H3. Similarly, the contribution and function of histone deacetyltransferases (HDACs) are not assessed here in relation to the H3K27ac levels at super-enhancers. Moreover, p300 levels are not correlated to their enzymatic activity. For example, recent evidence has implicated the function of the bromodomain of BRD4 for augmenting p300 activity in ESCs (Wu et al. 2018). Thus, the functional enzymes collaborating and competing to regulate histone acetylation in HUVEC are currently unknown but many mechanisms are likely involved.

In order to infer features of HUVEC chromatin I combined ChIP-seq data and performed an unsupervised cluster analysis on all open chromatin regions to reveal several findings of the EC chromatin landscape. The k-means clustering was set to 6 centroids to derive 6 clusters. Prior to the clustering, the Elbow method which is used to predetermine the optimal number of clusters in a dataset was performed. However, there was no clear elbow in the plot possibly due to high dimensionality of the data, another reason to cluster the data into similar groups. I also assessed the optimal number of clusters using a different algorithm; the Silhouette Method. This is performed post clustering to calculate how well a given chromatin region fits to its assigned cluster. This analysis was also inconclusive as there was no optimal cluster number that would fit the data optimally to a set number of clusters. The 6 clusters seemed to show clear segregation of chromatin states and further corroborate with functional features of the chromatin.

These clusters served to identify novel observations regarding chromatin regulation in HUVEC. At the promoter cluster, ERG and cFOS colocalise whereas cJUN displays weak occupancy and GATA2 is entirely absent. Using motif enrichment analysis at the different identified clusters I identified existing and putative TF binding patterns. Indeed, promoter regions do not enrich for the consensus GATA binding motif and therefore GATA2 can be generally classified as enhancer-specific in HUVEC. In contrast, consensus sequences for SP1, NFY, NRF and KLF are found in the promoter cluster (Figure 4.11). SP1, NFY and NRF are ubiquitous TFs at promoters (Benner et al. 2013) with NRF present at promoters of many housekeeping genes (FitzGerald et al. 2004). By deduction it seems that ETS and KLF TF motifs are at endothelial-specific promoters with ETS motifs most highly enriched. In fact, the colocalisation of ETS and KLF motifs at promoters has been reported in *in vivo* studies of cardiac EC (Lother et al. 2018).

The clustering analysis also distinguishes classes of enhancers whose activity correlates with both gene expression in HUVEC and chromatin interaction frequency. The enhancer classes also show a progressive decrease in abundance of each TF included in the clustering (ERG, GATA2, cFOS and cJUN), from the strong class I enhancers to class II enhancers to the inactive poised enhancer cluster. Analysis of these endothelial enhancer clusters showed AP-1 and ETS TF motif families as most highly represented. A recent study exhaustively uncovered the composition of TF assembly on chromatin using a technique called Active TF Identification (ATI) assay for the sensitive detection of DNA-binding. This work unveiled that only a small number of TFs, both cell type-specific and ubiquitous, display strong binding activity and these few TFs are those that control the overall gene regulatory landscape (Wei et al. 2017).

It has previously been determined that ETS and AP-1 members are responsible for enhancer selection and therefore for lineage determination in ECs (Hogan et al. 2017). In fact, ETS-AP-1 composite motif sites have been found to be permissive for the interaction between ERG and AP-1 (Madison et al. 2018). Furthermore, functional characterisation

of these specific motifs was assessed in a study investigating positional distribution of motifs in nucleosome-depleted regions of 47 cell types, including HUVEC. Here, AP-1 members FOS and JUN where classified as TFs that function in "co-activator recruitment and activation" and the majority of ETS members were classified as "cell type-specific" (Grossman et al. 2018). Together, these findings point to EC-specific enhancer selection by ETS factors with AP-1 acting as a possible intermediate for the necessary recruitment of transcriptional co-activators.

In the CTCF bound cluster, ERG and AP-1 members are found to colocalise with CTCF sites. It is plausible that cooperativity exists between CTCF and core TFs to mediate chromatin structure. There is a clear overrepresentation of the CTCF motif in this cluster and no evident enrichment of other TF motifs, implying that CTCF is selecting its binding sites directly. It could therefore be that ERG and AP-1 are localising indirectly to these CTCF sites possibly by protein-protein interactions.

Together these results reveal a cooperative TF network associated with enhancers and super-enhancers that are central to endothelial chromatin structure and function which warrant further experimental investigation.

5 CRISPR-CAS9-MEDIATED MODULATION OF THE *CDH5* SUPER-ENHANCER

5.1 Introduction

The specification of super-enhancers is based on the rich abundance of histone modifications and transcriptional co-activators. To investigate the functional role and structure of super-enhancers a targeted approach using precise genetic editing techniques was adopted. The emergence of the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 nuclease (Cas9) technology has provided an unprecedented opportunity for genome editing in human cells. The functional enzymatic component is the Cas9 endonuclease which induces double-stranded DNA breaks that are subsequently repaired by non-homologous end joining (NHEJ). The precision in the system is achieved by a sequence-specific single guide RNA (gRNA) that targets the genomic locus of interest. To this end, CRISPR-Cas9 can be used to target protein-coding genes for complete gene knockout and for the study of non-coding transcriptional regulatory elements with unknown function.

Due to their functional importance and marked transcriptional activity, enhancer regions have been selectively studied by repurposed CRISPR-Cas9 methods (Brown et al. 2018; Zhang et al. 2016). Activation or repression of enhancer regions by the fusion of effector proteins to a catalytically deactivated Cas9 enzyme (dCas9) has been achieved (Gilbert et al. 2013). CRISPRa (activation) and CRISPRi (interference) methods have provided a powerful tool by fusion of dCas9 to modulator enzymes such as acetyltransferase p300 and methyltransferase Krüppel-associated box (KRAB) for activation and repression respectively.

Previously, the use of CRISPR-Cas9 technology has been implemented in endothelial cells to disrupt gene expression (Abrahimi et al. 2015; Karampini et al. 2019) and allow

clonal expansion by utilising endothelial colony forming cells (ECFC). Super-enhancer composition has been studied in other cell types to elucidate the contribution of single constituent enhancers to the larger super-enhancer by CRISPR-Cas9 deletion and CRISPR-KRAB-dCas9 inhibition (CRISPRi) (Huang et al. 2018, 2016).

In Chapter 2 and Chapter 4, I defined hundreds of super-enhancers in HUVEC. Here I aim to validate the function of one super-enhancer by CRISPR genome editing and deconstruct it by assessing the individual contribution of each constituent enhancer to the super-enhancer gene expression. This super-enhancer-associated gene was selected based on its specific and essential role in endothelial cell biology. The selected super-enhancer candidate was associated with VE-cadherin (*CDH5*), the 13th highest H3K27ac-defined super-enhancer in HUVEC (Kalna et al. 2019). VE-cadherin is a pivotal endothelial adherens junction molecule for cohesion and organisation of intercellular junctions (Vestweber 2008). The TF ERG is known to regulate *CDH5* to mediate endothelial survival and angiogenesis both *in vitro* (Birdsey et al. 2008) and *in vivo* (Birdsey et al. 2015).

In this chapter, I deconstruct the *CDH5* super-enhancer to investigate individual ERG-bound loci on constituent enhancers in endothelial cells using both CRISPR-Cas9 deletion and CRISPRi. Target enhancer elements are identified by ChIP-seq data from HUVEC for precise genome editing by CRISPR.

In order to delete a specific enhancer element in the genome a dual gRNA approach is adopted: one gRNA targets the left side of the enhancer and the other gRNA targets the right side of the enhancer. For increased efficiency the two gRNAs are cloned together into a vector expressing Cas9. Alternatively, for repression of the enhancer elements by CRISPRi only one gRNA is designed to specifically target the centre of the enhancer to allow heterochromatin induction in both directions for effective inhibition by the KRAB enzyme. The contribution of each constituent enhancer to the *CDH5* super-enhancer is confirmed by its effect on *CDH5* expression.

As a primary cell type, HUVEC are difficult to transfect and therefore a lentiviral

transduction method provides an efficient vector delivery method. HUVEC also suffer from short replication capacity and are more averse to clonally expansion. Therefore to expand the potential of successful genetic editing umbilical cord blood-derived ECFC are used. ECFC are superior to HUVEC for their reportedly longer replication capacity and clonability (Medina et al. 2017).

5.2 Methods and Data

Analysis of the *CDH5* genomic region

High-throughput sequencing data for histone modifications and TFs in HUVEC and ECFC was obtained from sources listed in Table 5.8. ChIP-seq data were aligned on the UCSC Genome Browser (https://genome.ucsc.edu) on the hg19 reference genome. Control genomic sequences were selected based on negligible signal across all active histone marks and TF binding ChIP-seq data. Hi-C data for HUVEC was obtained from Rao et al. (2014) and visualised on the 3D Genome Browser (http://promoter.bx.psu.edu/hi-c/). Promoter capture Hi-C data generated for endothelial precursors (ECFC) were visualised on the Capture HiC Plotter (https://www.chicp.org/).

RNA-seq of HUVEC

HUVEC were treated by Dr Claire Peghaire (Imperial College London) with AllStars Negative Control siRNA (Qiagen) alongside treatment targeting human ERG using 20 nM siRNA as described in Chapter 2 and 3. A total RNA extraction was performed by Dr Peghaire using the RNeasy Mini Kit (Qiagen) and high quality RNA was shipped to Active Motif for polyA+ RNA enrichment, library generation and sequencing using Illumina NextSeq 500. I use the siCtl treated samples in this chapter for assessing basal gene expression levels of *CDH5* and associated genes. Active Motif generated bam files after aligning to the hg19 reference genome to calculate Fragments Per Kilobase per Million mapped reads (FPKM) using Salmon (Patro et al. 2017).

CRISPR-Cas9 delivery vectors

For deletion of genomic regions a vector with a lentiviral backbone was used. LentiCRISPR v2 is a high-titre producing vector with Cas9 and gRNA filler cassettes (Shalem et al. 2014). The LentiCRISPR v2 plasmid also confers resistance to puromycin. For CRISPRi, two vectors were used to assess their efficiency in endothelial cells. Lenti-(BB)-EF1α-KRAB-dCas9-2A-BlastR and Lenti-(BB)-hPGK-KRAB-dCas9-2A-BlastR are similar systems that vary by their core promoter, containing EF1α and hPGK promoters respectively. They both have cassettes for dCas9, KRAB and a gRNA filler alongside blasticidin resistance. All vectors possessed a BsmBI restriction enzyme site which recognises an asymmetric DNA sequence and effectively cuts outside of its recognition sequence to generate non-palindromic overhangs. A summary of these specifications are listed in Table 5.1.

LentiCRISPR v2 was a gift from Professor Feng Zhang, Broad Institute (Addgene plasmid # 52961)
Lenti-(BB)-EF1α-KRAB-dCas9-P2A-BlastR and Lenti-(BB)-hPGK-KRAB-dCas9-2A-BlastR was a gift from Professor Jorge Ferrer, Imperial College London (Addgene plasmid # 118154)

Vector	Vector size (bp)	Selective marker	Application
LentiCRISPR v2	14873	Puromycin	CRISPR-Cas9
Lenti-(BB)-EF1α-KRAB-dCas9-2A-BlastR	12935	Blasticidin	CRISPRi
Lenti-(BB)-hPGK-KRAB-dCas9-2A-BlastR	15243	Blasticidin	CRISPRi

Table 5.1: Specifications of plasmid used for CRISPR genome editing in this chapter.

Single guide RNA design

I used http://crispr.mit.edu/ (now shut down) to design 20 nucleotide long gRNAs that include a protospacer adjacent motif (PAM) sequence. Guides for deletions were designed to either side of the target region and guides for CRISPRi were designed to the centre. For each region two sets of deletion guides were designed and for CRISPRi three

guides were designed. All gRNAs were absent of BsmBI recognition sequences. Primer sequences are shown in Figure 5.1A; for deletions the gRNA 1 sequence was flanked by BsmBI overhang and scaffold sequences or for gRNA 2 a BsmBI overhang and H1 promoter sequences. Figure 5.2A illustrates the oligonucleotide design for CRISPRi; primer sequences also included the BsmBI overhang sequences with an additional guanine nucleotide at the 5' end of the forward primer if not already present and its complementary cytosine at the 3' end of the reverse primer for efficient expression from the U6 promoter. The primers for all target regions are listed in Table 5.2.

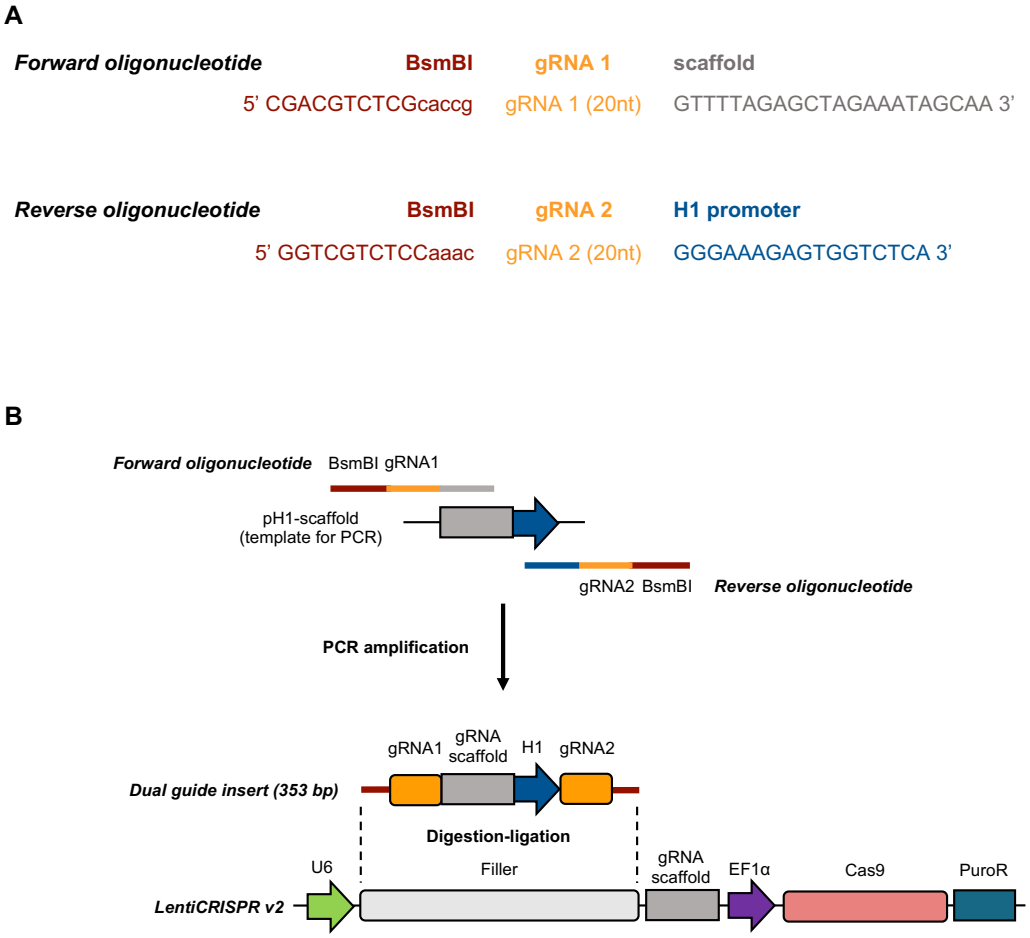


Figure 5.1: Design and cloning of dual gRNAs for CRISPR-Cas9 deletion. (A) Oligonucleotide sequences of the forward primer with gRNA sequence 1 and reverse primer with gRNA sequence 2 (both in yellow) where both primers have BsmBI recognition and restriction sites (dark red). Forward oligonucleotide has the sequence for the gRNA scaffold and the reverse oligonucleotide has the sequence for the H1 promoter required to drive expression of gRNA 2. The forward oligonucleotide is 57 bp in length and the reverse oligonucleotide is 51 bp. (B) Schematic illustrating the two-step protocol for dual gRNA cloning. The step involves a PCR amplification step to generate a 353 bp insert containing both gRNAs, scaffold, H1 promoter and BsmBI restriction sequences. The next step is a one-step digestion-ligation to insert the 353 bp product into the LentiCRISPR v2 vector.

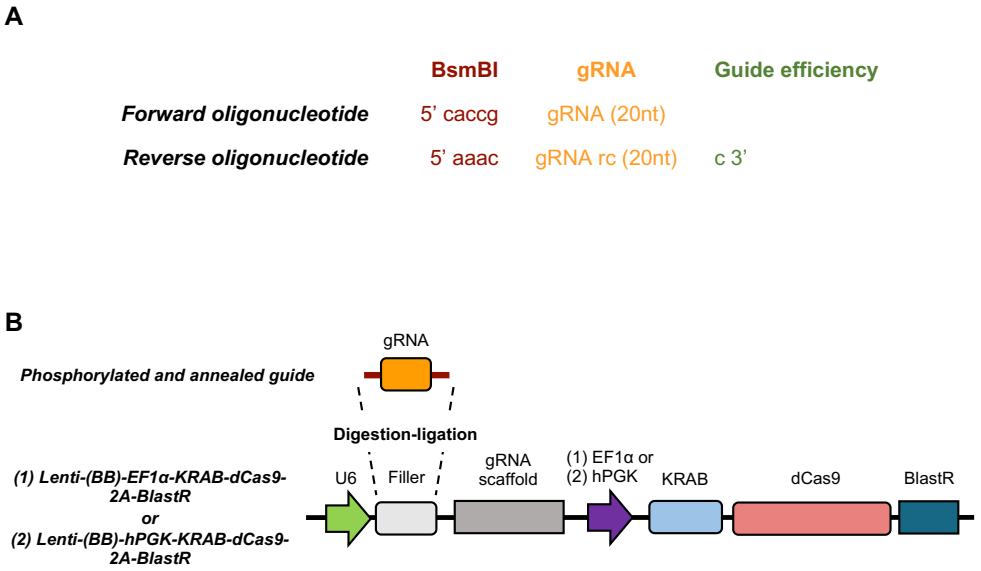


Figure 5.2: Design and cloning of single gRNA for CRISPRi. (A) Forward and reverse oligonucleotide sequences contain gRNA sequence for forward and reverse oligonucleotides are complimentary. The 5' ends of the sequence have the BsmBI sequence alongside an additional guanine nucleotide (if not already present), complemented by a cytosine on the reverse strand necessary to increase the efficiency of expression from the U6 promoter. (B) The first step is phosphorylation and annealing of the primers for the subsequent digestion-ligation step to insert the product into either the Lenti-(BB)-EF1α-KRAB-dCas9-2A-BlastR or Lenti-(BB)-hPGK-KRAB-dCas9-2A-BlastR vectors.

Agarose gel electrophoresis

All PCR generated products were separated by electrophoresis on an agarose gel of (1-2% w/v) dissolved in Tris-Acetate-EDTA (TAE) buffer with GelRed nucleic acid stain (Biotium) at a final concentration of 1μl/ml. PCR products were loaded with 5x loading dye with either a 100 bp or 1 kb DNA molecular weights ladder. Electrophoresis was carried out at 100 volts (constant) until DNA samples had migrated through the gel. DNA was visualised on a BioRad ChemiDoc XRS+ system.

Cloning strategy

For CRISPR-Cas9 deletions, insertion of two gRNAs into the same vector was conducted using a dual gRNA cloning protocol approach (Miguel-Escalada et al. 2019). The vector pScaffold-H1 was used as a template for PCR amplification at 0.25ng/μl with 1X

Q5 reaction buffer, 200μM dNTPs, 0.5μM forward/reverse primer and 0.02U/μl Q5 high-fidelity DNA polymerase (New England Biolabs) with an annealing temperature of 58°C for 30 cycles to generate a 353 bp product containing the two gRNA sequences, H1 promoter and BsmBI recognition sites. This product was purified with the QIAquick Gel Extraction Kit (Qiagen). The PCR product was diluted 1 in 20 and then 1μl was used in an one-step digestion-ligation reaction along with 1X tango buffer, 1μl FastDigest BsmBI, 1μl 0.1M DTT, 1μl 10mM ATP, 0.5μl T7 ligase and 100 ng of LentiCRISPR v2. The digestion-ligation reactions were subjected to 6 cycles of 5 min at 37°C and 5 min at 23°C with a final 5 min incubation at 37°C. The strategy is illustrated in Figure 5.1B.

For CRISPRi, the gRNAs flanked by BsmBI-compatible overhangs were phosphorylated and annealed with 20μM forward/reverse primer, 1X T4 PNK buffer, 1μl 10mM ATP and 0.5 μl T4 PNK with incubation for 30 min at 37°C followed by a 5 min at 95°C and a ramp down at 5°C/min to 25°C. A digestion-ligation reaction was set up to insert gRNAs into BsmBI-digested vectors with 2μl phosphorylated and annealed primers diluted 1:250, 2μl FastDigest BsmBI, 0.2μl 0.1M DTT, 2μl 10mM ATP, 0.5μl T7 ligase and 150ng of of either Lenti-(BB)-EF1α-KRAB-dCas9-2A-BlastR or Lenti-(BB)-hPGK-KRAB-dCas9-2A-BlastR vectors. The digestion-ligation reactions were subjected to 6 cycles of 5 min at 37°C and 5 min at 23°C with a final 5 min incubation at 37°C. A schematic to illustrate the cloning step is shown in Figure 5.2B.

Name	sgRNA1	sgRNA2rc	Sequence 5' to 3'	Deletion size (bp)	Quality score
E1 D-1	AACCTCTCTGTATTCGCTCT		CGACGTCTCGcaccgAACCTCTCTGTATTC GCTCTGTTTTAGAGCTAGAAATAGCAA	1911	82
		GACTGGGGCACATCAAAGAC	GGTCGTCTCCaaacGACTGGGGCACATCA AAGACGGGAAAGAGTGGTCTCA		83
E1 D-2	GTGAAGCTGGAGGATCCTCG		CGACGTCTCGcaccgGTGAAGCTGGAGGA TCCTCGGTTTTAGAGCTAGAAATAGCAA	2247	81
		CGCGAAGCCTTTGCCACATT	GGTCGTCTCCaaacCGCGAAGCCTTTGCC ACATTGGGAAAGAGTGGTCTCA		77
E1 i-1	TGATGCCACAGACACGTTAA		caccgTGATGCCACAGACACGTTAA		88
		TTAACGTGTCTGTGGCATCA	aaacTTAACGTGTCTGTGGCATCAc		
E1 i-2	GAATTCAAACCCCGTTCAAT		caccGAATTCAAACCCCGTTCAAT		88
		ATTGAACGGGGTTTGAATTC	aaacATTGAACGGGGTTTGAATTC		
E1 i-3	GTGCCCCAGCATGTAACCAC		caccGTGCCCCAGCATGTAACCAC		87
		GTGGTTACATGCTGGGGCAC	aaacGTGGTTACATGCTGGGGCAC		
E2 D-1	GACCAGCAAGCCTTATCGAG		CGACGTCTCGcaccgGACCAGCAAGCCTT ATCGAGGTTTTAGAGCTAGAAATAGCAA	1058	95
		AACGGCCTGCAAAGTCTGAA	GGTCGTCTCCaaacAACGGCCTGCAAAGT CTGAAGGGAAAGAGTGGTCTCA		76
E2 D-2	TTCAGACTTTGCAGGCCGTT		CGACGTCTCGcaccgTTCAGACTTTGCAG GCCGTTGTTTTAGAGCTAGAAATAGCAA	1419	85
		AACAGCTTATGGCTGGCATC	GGTCGTCTCCaaacAACAGCTTATGGCTG GCATCGGGAAAGAGTGGTCTCA		81
E2 i-1	CCACAGGAACAGTCGTTTTC		caccgCCACAGGAACAGTCGTTTTC		85
		GAAAACGACTGTTCCTGTGG	aaacGAAAACGACTGTTCCTGTGGc		
E2 i-2	GCTGGGGGTCACCGTAGCAG		caccGCTGGGGGTCACCGTAGCAG		78
		CTGCTACGGTGACCCCCAGC	aaacCTGCTACGGTGACCCCCAGC		
E2 i-3	AAGGCGGCTGGTGGGCTAGC		caccgAAGGCGGCTGGTGGGCTAGC		74
		GCTAGCCCACCAGCCGCCTT	aaacGCTAGCCCACCAGCCGCCTTc		

Name	sgRNA1	sgRNA2rc	Sequence 5' to 3'	Deletion size (bp)	Quality score
E3 D-1	CTGACGCGGCGGATCGCCTG		CGACGTCTCGcaccgCTGACGCGGCGGAT CGCCTGGTTTTAGAGCTAGAAATAGCAA	2038	95
		TCAAGAGTGTGCAACCTGTG	GGTCGTCTCCaaacTCAAGAGTGTGCAAC CTGTGGGGAAAGAGTGGTCTCA		77
E3 D-2	GCGTGAGCCACCGTAGCCGG		CGACGTCTCGcaccgGCGTGAGCCACCGT AGCCGGGTTTTAGAGCTAGAAATAGCAA	1981	87
		GCACTCCTCGTGGGGCTGTA	GGTCGTCTCCaaacGCACTCCTCGTGGGG CTGTAGGGAAAGAGTGGTCTCA		80
E3 i-1	CAATGAGGAAACGTATCAGC		caccgCAATGAGGAAACGTATCAGC		86
		GCTGATACGTTTCCTCATTG	aaacGCTGATACGTTTCCTCATTGc		
E3 i-2	CTGAGCTGATAGATAGGTGC		caccgCTGAGCTGATAGATAGGTGC		82
		GCACCTATCTATCAGCTCAG	aaacGCACCTATCTATCAGCTCAGc		
E3 i-3	GCTGGCGATCAGAGATGTTC		caccGCTGGCGATCAGAGATGTTC		82
		GAACATCTCTGATCGCCAGC	aaacGAACATCTCTGATCGCCAGC		
E4 D-1	GACCATTCCAAGCACCGGAC		CGACGTCTCGcaccgGACCATTCCAAGCAC CGGACGTTTTAGAGCTAGAAATAGCAA	1643	94
		ATACGGATGCATTATTATTT	GGTCGTCTCCaaacATACGGATGCATTATT ATTTGGGAAAGAGTGGTCTCA		81
E4 D-2	ATCAGCAAACTACCCCCTGC		CGACGTCTCGcaccgATCAGCAAACTACCC CCTGCGTTTTAGAGCTAGAAATAGCAA	1935	82
		CGCAGGTCACAGGTTACAGA	GGTCGTCTCCaaacCGCAGGTCACAGGTT ACAGAGGGAAAGAGTGGTCTCA		80
E4 i-1	GCACAATTATCCCACCGGGA		caccGCACAATTATCCCACCGGGA		89
		TCCCGGTGGGATAATTGTGC	aaacTCCCGGTGGGATAATTGTGC		
E4 i-2	CCTTTCAAGGGCTCGCCCAA		caccgCCTTTCAAGGGCTCGCCCAA		88
		TTGGGCGAGCCCTTGAAAGG	aaacTTGGGCGAGCCCTTGAAAGGc		
E4 i-3	ACCACAAGGGGTCGTCACGC		caccgACCACAAGGGGTCGTCACGC		96
		GCGTGACGACCCCTTGTGGT	aaacGCGTGACGACCCCTTGTGGTc		

Name	sgRNA1	sgRNA2rc	Sequence 5' to 3'	Deletion size (bp)	Quality score
TSS D-1	CCGGGAGTAGGGGTTCAAGT		CGACGTCTCGcaccgCCGGGAGTAGGGGT TCAAGTGTTTTAGAGCTAGAAATAGCAA	557	82
		GGGCACAAATGGGCATCTCG	GGTCGTCTCCaaacGGGCACAAATGGGCA TCTCGGGGAAAGAGTGGTCTCA		86
TSS D-2	GCTGAGGCTCAGCGGGTTTA		CGACGTCTCGcaccgGCTGAGGCTCAGCG GGTTTAGTTTTAGAGCTAGAAATAGCAA	893	72
		CCTCGGTGCCTTACCCTTAC	GGTCGTCTCCaaacCCTCGGTGCCTTACC CTTACGGGAAAGAGTGGTCTCA		86
TSS i-1	GCCTGTCAGCCGACCGTCTT		caccGCCTGTCAGCCGACCGTCTT		93
		AAGACGGTCGGCTGACAGGC	aaacAAGACGGTCGGCTGACAGGC		
TSS i-2	TGTCAGCCGACCGTCTTTGG		caccgTGTCAGCCGACCGTCTTTGG		93
		CCAAAGACGGTCGGCTGACA	aaacCCAAAGACGGTCGGCTGACAc		

Name	sgRNA1	sgRNA2rc	Sequence 5' to 3'	Deletion size	Quality score
Control D-1 (proximal)	GCCTCAGCCTATCTAGTAGC		CGACGTCTCGcaccgGCCTCAGCCTATCT AGTAGCGTTTTAGAGCTAGAAATAGCAA	1688	86
		AGTAGGCACACTGTAAGCAG	GGTCGTCTCCaaacAGTAGGCACACTGTA AGCAGGGGAAAGAGTGGTCTCA		78
Control D-2 (proximal)	GGTCCATGAACTTCAAGCCT		CGACGTCTCGcaccgGGTCCATGAACTTC AAGCCTGTTTTAGAGCTAGAAATAGCAA	1187	76
		GAGCGTAAACATGTGTTTAC	GGTCGTCTCCaaacGAGCGTAAACATGTG TTTACGGGAAAGAGTGGTCTCA		88
Control i-1 (proximal)	GCCTCAGCCTATCTAGTAGC		caccGCCTCAGCCTATCTAGTAGC		86
		GCCTCAGCCTATCTAGTAGC	aaacGCTACTAGATAGGCTGAGGC		
Control i-2 (proximal)	CTGCTTACAGTGTGCCTACT		caccgCTGCTTACAGTGTGCCTACT		84
		AGTAGGCACACTGTAAGCAG	aaacAGTAGGCACACTGTAAGCAGc		
Control i-3 (proximal)	GGTCCATGAACTTCAAGCCT		caccGGTCCATGAACTTCAAGCCT		74
		AGGCTTGAAGTTCATGGACC	aaacAGGCTTGAAGTTCATGGACC		
Control D-3 (distal)	GTACGCCTATTTCGGGGATA		CGACGTCTCGcaccgGTACGCCTATTTCG GGGATAGTTTTAGAGCTAGAAATAGCAA	1414	96
		GTGAATCCTTGTGGATTAGC	GGTCGTCTCCaaacGTGAATCCTTGTGGA TTAGCGGGAAAGAGTGGTCTCA		78
Control D-4 (distal)	GCCTAGGTACGCCTATTTCG		CGACGTCTCGcaccgGCCTAGGTACGCCT ATTTCGGTTTTAGAGCTAGAAATAGCAA	1419	96
		AGCTAATCCACAAGGATTCA	GGTCGTCTCCaaacAGCTAATCCACAAGG ATTCAGGGAAAGAGTGGTCTCA		77
Control i-5 (distal)	GTACGCCTATTTCGGGGATA		caccGTACGCCTATTTCGGGGATA		98
		TATCCCCGAAATAGGCGTAC	aaacTATCCCCGAAATAGGCGTAC		
Control i-6 (distal)	GCCTAGGTACGCCTATTTCG		caccGCCTAGGTACGCCTATTTCG		99
		CGAAATAGGCGTACCTAGGC	aaacCGAAATAGGCGTACCTAGGC		

Table 5.2: Guide RNA sequences designed for all regions. For each enhancer and proximal controls, two pairs of deletion gRNAs were designed and three single gRNAs for CRISPRi. For the TSS and distal control, two pairs of deletion gRNAs and two single gRNAs for CRISPRi were designed. For deletion primer pairs, the size of the deletion is indicated and for every gRNA designed the quality score is reported as predicted from the guide design resource from the Zhang Lab at MIT (http://crispr.mit.edu/) which takes into account potential mismatches to non-specific sequences. No gRNAs had off-target sites within other genes.

Bacterial transformations and colony screening

Final ligation products were used to transform Stbl3 chemically competent *E. coli* (Thermo Fisher). 2μl of the final ligation was used to transform 25μl of Stbl3 *E. coli* after which transformations were plated. Colonies were picked after overnight incubation at 37°C and used to inoculate 5 ml of LB broth supplemented with 100μg/ml of ampicillin for selection. Bacterial cultures were used to screen for positive clones with vectors harbouring gRNAs. PCR amplifications were performed with primers designed to the U6 and EF1α promoters for LentiCRISPR v2 and Lenti-(BB)-EF1α-KRAB-dCas9-2A-BlastR vectors or U6 and hPGK promoters for the Lenti-(BB)-hPGK-KRAB-dCas9-2A-BlastR vector. These primers are listed in Table 5.3. Bacterial cultures with amplicons of the correct size were selected for plasmid purification using the QIAprep Spin Miniprep Kit (Qiagen) and confirmation of correct insertions were validated by sequencing of the plasmid using the U6 promoter primer for all plasmids. Bacterial cultures for plasmids with confirmed gRNA insertion were scaled-up and purified for larger concentrations using the MaxiPrep HiSpeed Kit (Qiagen).

Name	Oligonucleotide sequences (5' to 3')
U6 promoter	ACCGAGGGCCTATTTCCCATG
EF1α promoter	TACACGACATCACTTTCCCAG
hPGK promoter	ACGGACGTGAAGAATGTGC

Table 5.3: Primers for screening bacterial clones to identify sequences with inserted gRNAs.

Lentiviral production

HEK293FT cells were used for lentiviral production. HEK293FT were cultured in DMEM media supplemented with 10% FBS with 0.1 mM MEM non-essential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine, 1% penicillin-streptomycin and 500μg/ml of geneticin. For transfection, 293FT were seeded at 75,000 cells/cm^2 in T75 or T175 flasks in complete medium without any antibiotics. After 24h cells were approximately 80% confluent and were transfected with gRNA-containing CRISPR vec-

tors with third generation packaging plasmids pMDLg/pRRE, pRSV-Rev and pMD2.G using PEI-Pro (Polyplus-transfection) diluted in DMEM. A 1:1 ratio of total DNA μg to μl of PEI-Pro was incubated at room temperature for 15 min prior to drop-wise addition to cells. The following day media was replaced by fresh antibiotic-free media and then cells were left to incubate for 48h. After 48h, lentiviral particles were collected and concentrated by centrifugation for 5 min at 1500 rpm, filtered using Steriflip-HV, 0.45μm, PVDF filters (Millipore), supplemented with 1mM $MgCl_2$ and treated with 1μg/ml DNaseI for 20 min at 37°C. Viral particles were concentrated overnight by incubation with Lenti-X Concentrator (Clontech) at 4°C. The following day, viral particles were collected by centrifugation for 45 min at 2500 rpm at 4°C, resuspension in PBS and aliquoting and storage at -80°C.

pMDLg/pRRE, pRSV-Rev and pMD2.G were a gift from Didier Trono (plasmids #12251, #12253 and #12259, Addgene).

Culture of HUVEC and ECFC

Pooled HUVEC were purchased from Lonza and cultured in EGM-2 media (Lonza) on 1% gelatin. HUVEC were used at passage 2 or 3. ECFC were isolated by Dr Gaye Saginc (Imperial College London) from cord blood and cultured in complete EGM-2 supplemented with 10% fetal bovine serum (FBS) (HyClone, Thermo Scientific) on type I rat tail collagen (BD Biosciences). ECFC were used for CRISPR engineering at passage 3.

Transduction of HUVEC and ECFC

HUVEC or ECFC were seeded in a format of either 100,000 cells per 12-well or 200,000 cells per 6-well. Next day, concentrated lentiviral supernatant was thawed and used to transduce cells with 10μl per 12-well or 20μl per 6-well in 1ml or 2ml of media respectively. Next day media was replenished and 24h later antibiotic selection was commenced with 1μg/ml puromycin for cells infected with LentiCRISPR v2 plasmids

and 10μg/ml blasticidin for CRISPRi plasmids. Media was replenished every 48h until all negative control cells were dead which took approximately 7-8 days after which cells were allowed to reach confluency in full growth media.

Validation of deletions

Following transduction and antibiotic selection, cells transduced with LentiCRISPR v2 were collected in bulk for genomic DNA (gDNA) extraction post-transduction and post-selection. Genomic DNA was isolated using the AllPrep DNA/RNA/Protein MiniDNA extraction Kit (Qiagen) and used in PCR to verify the presence of intended genomic deletions using the reaction mix in Table 5.4 and PCR protocol in Table 5.5. The primers used to amplify the regions around listed in Table 5.6. Primers designed inside and outside of the deletion allowed samples to be examined for the presence/absence of a deletion. DNA polymerase extension time in the PCR was 2 min to allow products greater than 2 kb to be produced.

Master mix for PCR	µl
Forward primer	1
Reverse primer	1
Genomic DNA (10ng /µl)	2
ReadyMix TaqMan	5
ddH_2O	1
total	10

Table 5.4: Reaction mix for PCR amplification.

Stage (cycles)	Temp (°C)	Time (min:sec)
1 (x1)	95	1:00
2 (x35)	95	0:15
	60	0:15
	72	2:00
3 (x1)	72	5:00
	4	∞

Table 5.5: PCR cycling conditions.

Name	Orientation	Oligonucleotide sequences (5' to 3')	Amplicon size (bp)	Amplicon size in deletion (bp)
E1 In	Forward	ATGCCCATGCTAACACACTCA	309	No amplicon
	Reverse	TGCCCTGTGGTTACATGCTG		
E1 Out	Forward	CTGGGGCTTCTGGTTCCAAG	2200	~ 289
	Reverse	TCACCAATGTTTTATGTTCCTTTGT		
E4 In	Forward	CCTGACTCTCAACACAAATGCG	433	No amplicon
	Reverse	CTGTAAACAAAGGCTCCAGGC		
E4 Out	Forward	GGACCGTGTGCTTTCAGAGT	2268	~ 333
	Reverse	CATGAGGAGGTACTAAGGCCC		
TSS In	Forward	CAGGAAACCATCCCAGGGGG	305	No amplicon
	Reverse	AGTGGGACGATGCCCATTCTA		
TSS Out	Forward	TGGGATCACCACAGAAAATCAA	987	~ 430
	Reverse	CAGTAAGGGTAAGGCACCGAG		

Table 5.6: Primers for screening transduced endothelial cells for CRISPR-Cas9 mediated deletions. Two primer designs were used to identify deletions; one pair inside and pair outside of the intended deletion sequence (see Figure 5.6A). The size of the primer amplicon in non-deleted and deleted cells is indicated.

RT-qPCR analysis of gene expression

To determine the effect of gene editing in cells transduced with CRISPRi lentiviral particles RT-qPCR was used to assess gene expression of *CDH5*. RT-qPCR was also used to determine the effect on gene expression in cells carrying genomic deletions. Cells were harvested post-transduction and post-selection as for gDNA collection. Total RNA was extracted using the RNeasy Mini Kit (Qiagen). 500ng of total RNA was used when possible for reverse transcription with Superscript III Reverse Transcriptase (Invitrogen). RT-qPCR was performed on a Bio-Rad CFX96 system using PerfeCTa SYBR Green Fastmix (Quanta Biosciences). Relative gene expression was quantified by normalising to the housekeeping *GAPDH*. RT-qPCR primers are listed in Table 5.7. All results were plotted using Prism 8.0 (GraphPad Software).

Cell cloning by limiting dilution

Limiting dilution was performed to obtain clonal populations of ECFC. Cells were counted and diluted in order to seed cells at a density of 6, 3 and 1.5 cells per well in a

Name	Orientation	Oligonucleotide sequences (5' to 3')
CDH5	Forward	ACAGAGCTCCACTCACGCTC
	Reverse	CATGAGCCTCTGCATCTTCC
GAPDH	Forward	CAAGGTCATCCATGACAACTTTG
	Reverse	GGGCCATCCACAGTCTTCTG

Table 5.7: RT-qPCR primers used in detecting gene expression changes.

48-well plate in complete EGM-2 media. Media was replenished every 72h. After 7 days, colonies were visible and wells were screened for successfully derived clones. Clones were harvested when confluent or when cells ceased growth in RTL buffer and were screened directly by RT-qPCR.

Data

Data used in this chapter is listed in Table 5.8.

Experiment	Factor	Cell type	Condition	Dataset accession	Citation / Distributor
ChIP-seq	ERG	HUVEC	untreated	GSM3557980	Kalna et al. 2019
ChIP-seq	GATA2	HUVEC	untreated	GSM935347	Linnemann et al. 2011
ChIP-seq	cFOS	HUVEC	untreated	GSM935585	Linnemann et al. 2011
ChIP-seq	cJUN	HUVEC	untreated	GSM935278	Linnemann et al. 2011
ChIP-seq	H3K27ac	HUVEC	untreated	GSM733691	ENCODE/Broad Institute
ChIP-seq	H3K4me1	HUVEC	untreated	GSM733690	ENCODE/Broad Institute
ChIP-seq	H3K27ac	ECFC	untreated	EGAX00001291797	Blueprint Consortium
ChIP-seq	H3K27ac	ECFC	untreated	EGAX00001326489	Blueprint Consortium
Hi-C		HUVEC	untreated	GSM1551631	Rao et al. 2014
Promoter capture Hi-C		ECFC	untreated	EGAD00001002268	Javierre et al. 2016
RNA-seq		HUVEC	siCtl v siERG	N/A	Unpublished

Table 5.8: High-throughput sequencing data used in this chapter.

5.3 Results

5.3.1 Deconstructing the *CDH5* super-enhancer

To validate the function of these large intensely marked clusters, the *CDH5* super-enhancer was deconstructed to individual constituent enhancers. H3K4me1/H3K27ac defined enhancers in HUVEC were generally large, each around 5 kb in size. It was considered more feasible to select smaller genomic regions of the enhancers for editing as an inverse relationship between genomic deletion size and frequency of deletion has been demonstrated (Canver et al. 2014). It was therefore decided to target the ERG binding loci of the *CDH5* enhancers. Using ChIP-seq data the most highly ERG-occupied sites were chosen as optimal candidate regions for CRISPR editing. These included four enhancers; two downstream of the TSS and two upstream intronic enhancers (Figure 5.3A and Table 5.9). Genomic editing of the ERG-bound TSS region was utilised as a positive control to assess the maximum regulation of *CDH5* expression (Figure 5.3A and Table 5.9).

To determine whether enhancers classified in HUVEC were comparable to those in ECFC, ChIP-seq data for H3K27ac performed in two individual adult ECFC donors by the Blueprint Consortium (Stunnenberg et al. 2016) were aligned with ChIP-seq data from HUVEC. ERG-bound HUVEC enhancer regions showed distinct H3K27ac signals in both ECFC donors (Figure 5.3A).

The *CDH5* super-enhancer locus also provides a model to investigate TF cooperativity. The different combinations of ERG with GATA2, cFOS and cJUN at the four selected ERG-bound enhancers would suggest that E4, bound by all four TFs, may be the strongest enhancer and potentially classified as a hub enhancer (Huang et al. 2018) and might contribute most to *CDH5* expression. E2 is also bound by all four TF to a lower level than E4 but may also contribute significantly to gene expression (Figure 5.3A). E3 is bound by ERG, GATA2 and cFOS whereas E1 is solely ERG bound and

Region	Genomic location	Size (bp)	Distance to TSS (bp)
E1	chr16: 66,393,607-66,395,346	1740	-5179
E2	chr16: 66,397,891-66,398,796	906	-1729
TSS	chr16: 66,400,235-66,400,893	659	0
E3	chr16: 66,410,029-66,411,635	1607	-1572
E4	chr16: 66,415,881-66,417,164	1284	2674
Proximal CTL	chr16: 66,385,979-66,387,686	1708	12822
Distal CTL	chr16: 67,385,041-67,386,454	1414	940055

Table 5.9: Genomic loci selected for CRISPR genomic editing.

hypothetically may contribute less to levels of *CDH5* expression (Figure 5.3A). Therefore the selected genomic regions provide a model to study combinatorial TF assembly by precise genomic editing.

Control regions were chosen relative to the *CDH5* super-enhancer locus. A region proximal to the *CDH5* super-enhancer locus was selected to serve as a negative control. This region, located 12 kb downstream of the *CDH5* TSS in a non-coding and inactive transcriptional region (Figure 5.3B left), would be expected not to change *CDH5* gene expression despite being in vicinity of the *CDH5* gene. Additionally, a distal control was designed to be used in parallel during genomic editing experiments to consistently replicate the treatment of cells with gRNA-containing lentiCRISPR plasmids. This control was used to assess the gene expression changes versus the *CDH5*-targeting CRISPR plasmids. This region is on the same chromosome as *CDH5* but located approximately 1 Mb upstream of the *CDH5* TSS. It is a non-coding region on the leucine-rich repeat-containing (*LRRC36*) gene that is transcriptionally inactive in endothelial cells (Figure 5.3B right).

CRISPR genome editing at the selected *CDH5* locus could affect other genomic loci by long-range chromatin interactions. To identify the spatial chromatin organisation, I studied Hi-C data generated in HUVEC (Rao et al. 2014), to define domain boundaries and the topologically associated domain (TAD) in which the *CDH5* locus lies (Figure

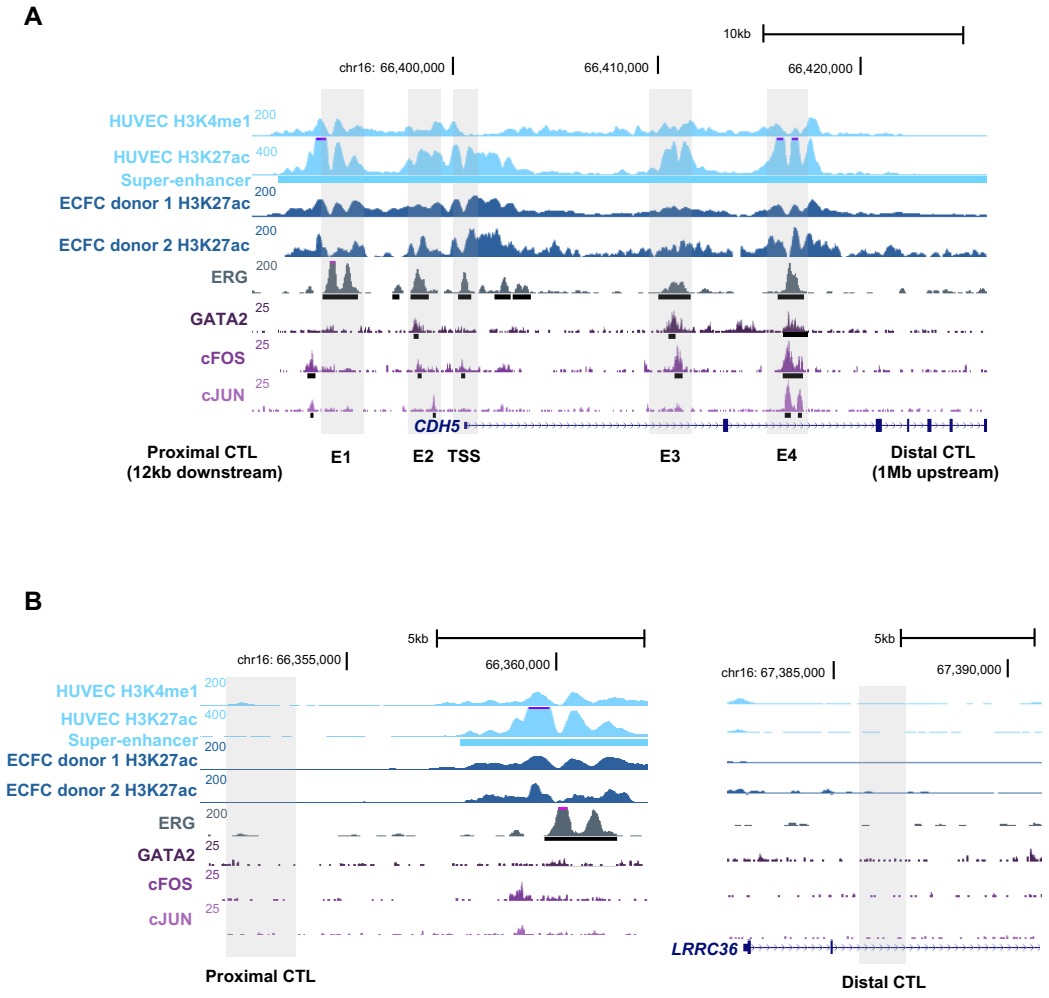


Figure 5.3: Locus of the *CDH5* super-enhancer selected for genomic editing. Genome browser view of (A) selected enhancer regions; E1, E2, E3 and E4 and the TSS region and (B) proximal and distal control (CTL) regions highlighted in grey. ChIP-seq signal for histone modifications in HUVEC, H3K4me1 and H3K27ac, and TFs ERG, GATA2, cFOS and cJUN with H3K27ac from two ECFC donors are shown.

5.4A). *CDH5* is located to the leftmost of a TAD containing other genes with low expression in endothelial cells (Figure 5.4B). Expression of all genes in this TAD would have to be assessed following CRISPR-editing to confirm that *CDH5* enhancers do not co-regulate other TAD-associated genes. Promoter-capture Hi-C performed in ECFC (Javierre et al. 2016) demonstrates that the *CDH5* promoter interacts with regions downstream of the four enhancers investigated here on the *CDH5* transcript but to no other coding or non-coding genomic regions (Figure 5.4C).

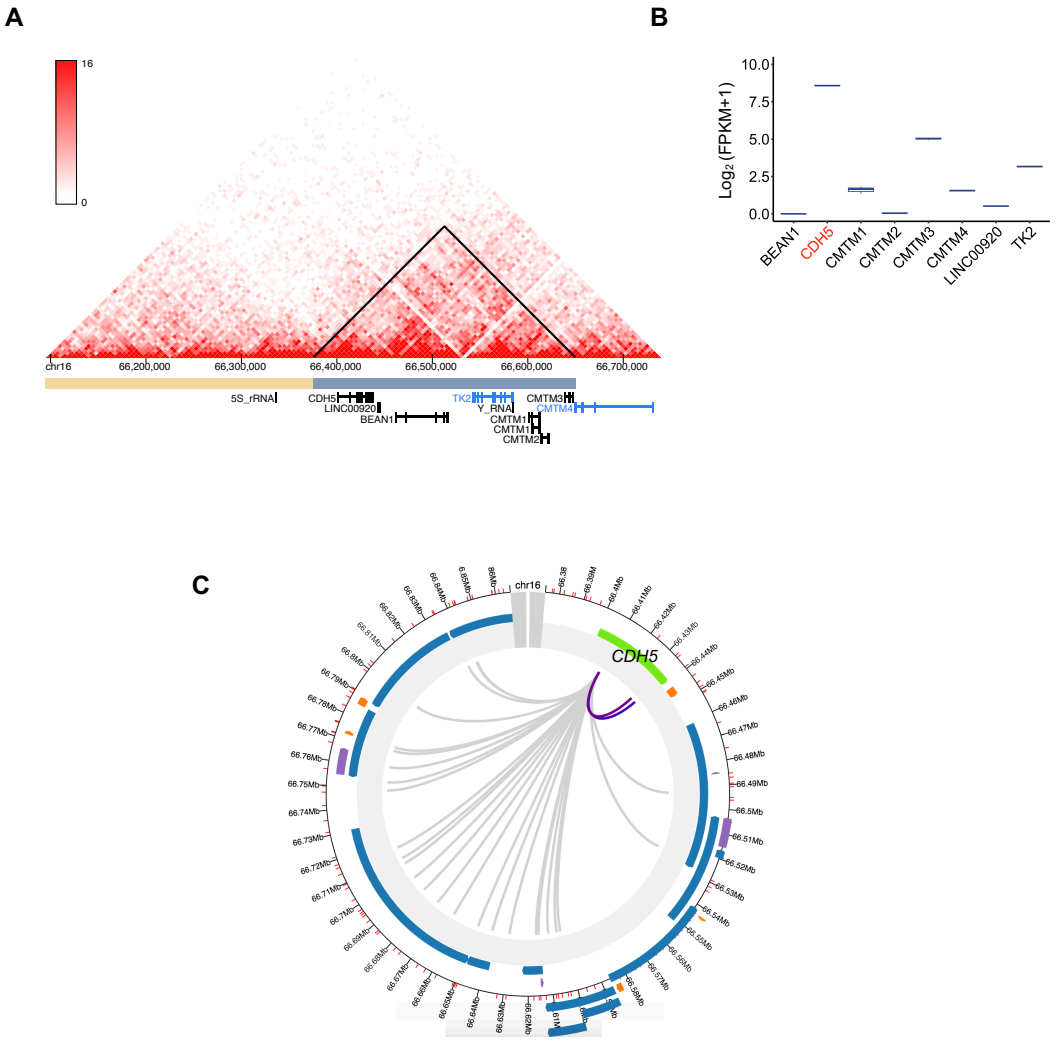


Figure 5.4: Spatial chromatin organisation at the *CDH5* super-enhancer in endothelial cells. (A) Hi-C interactions and TAD arrangement at *CDH5* shows the other genes in the shared TAD. Hi-C matrix is at a 10 kb resolution displayed on the hg19 reference with forward transcripts in black and reverse in blue shown in the 3D Genome Browser. (B) Boxplots of HUVEC gene expression in the *CDH5* TAD shows *CDH5* is most highly expressed. Expression shown as log of Fragments Per Kilobase per Million mapped reads (FPKM) retrieved from RNA-seq data of control siRNA treated HUVEC. (C) Promoter capture Hi-C for the *CDH5* locus from chicp.org in ECFC. Lines in colour show significant interactions only to regions at the 3' end of the *CDH5* gene.

5.3.2 Cloning to generate deletion CRISPR vectors with dual gRNAs

To precisely delete ERG-bound enhancers of the *CDH5* super-enhancer, a dual gRNA cloning protocol was adopted to co-express two distinct gRNAs with one targeting constitutive Cas9 to the left side and the other to the right side of the target enhancer region.

The success of paired gRNA cloning was determined by PCR amplification around the LentiCRISPR v2 vector insert site in bacterial clones with positive clones generating an 829 bp amplicon. Correct assembly was confirmed by sequencing genomic DNA. In general, success of dual gRNA cloning was lower than single gRNA cloning (strategy in Figure 5.1 versus strategy in Figure 5.2). Success of positive bacterial clones containing both dual gRNAs regardless of the gRNA design, was 1/12 for E1, 1/20 for E4, 1/8 for TSS and 1/12 for proximal controls giving an overall efficiency for dual gRNA cloning of 7.7%. Considering this, it was decided to take forward only one gRNA design per enhancer or control region. Successful clones and inserted gRNA sequences for E1, E4, TSS and proximal control are shown in Figure 5.5.

5.3.3 Genomic deletion of *CDH5* enhancer constituents in HUVEC

For genomic deletion of *CDH5* enhancer regions lentiviral packaged CRISPR deletion plasmids were used to transduce HUVEC. The resulting antibiotic-resistant bulk HUVEC population was examined for genomic deletions using one set of primers internal to the deletion ("In") and another set of primers upstream and downstream of the gRNA cleavage sites ("Out") (Figure 5.6A). In the absence of a deletion, the amplification by the "Out" primers may be too large to efficiently amplify despite a modified PCR cycle.

The deletion of E1 in a bulk population of HUVEC was evident by the presence of a 289 bp band that would be expected if deletion had taken place at this locus (Figure 5.6B). The "Out" primers did not amplify the large 2.2 kb band in the non-deleted cells;

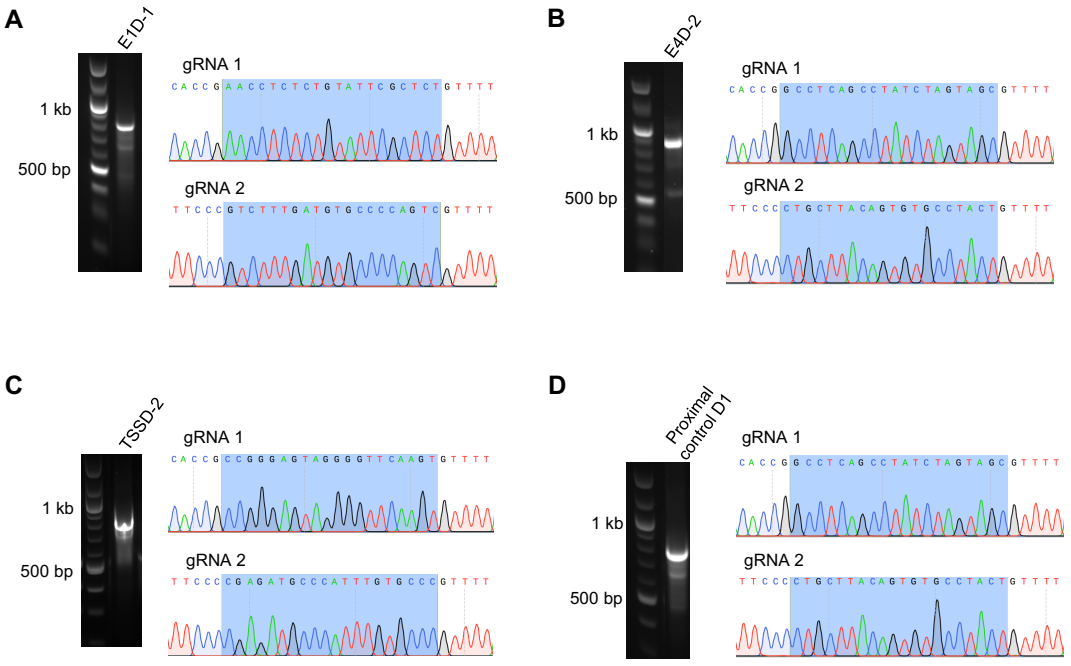


Figure 5.5: LentiCRISPR v2 vectors containing gRNA pairs. Gel representing the correctly sized amplicon and sequenced dual gRNAs shown for regions (A) E1, (B) E4, (C) TSS and (D) proximal control.

however with the "In" primers, the presence of a 309 bp band indicated that many of the E1 deletion HUVEC remained unedited (Figure 5.6B).

The deletion of E4 was similarly successful in a bulk population of HUVEC; a 333 bp band was observed with the "Out" primers which is visible in the gel (Figure 5.6C). A band of 433 bp with "In" primers demonstrated that a larger population of cells remained unedited as expected (Figure 5.6C). The amplification of the 2.3 kb band in both control and E4 deletion HUVEC was absent in this case (Figure 5.6C).

Finally, deletion of the *CDH5* TSS was not evident in the majority of cells. Firstly, "In" primers show a band at around 305 bp (a shift during the gel run was observed making the band run lower than the 300 bp marker) and secondly, "Out" primers produce a band around 1 kb indicating that many cells remain unedited (Figure 5.6D). However, in addition to the larger "Out" 1 kb band there is also the faint presence of a band at 430 bp indicating that a small subset of cells potentially harbours the intended deletion.

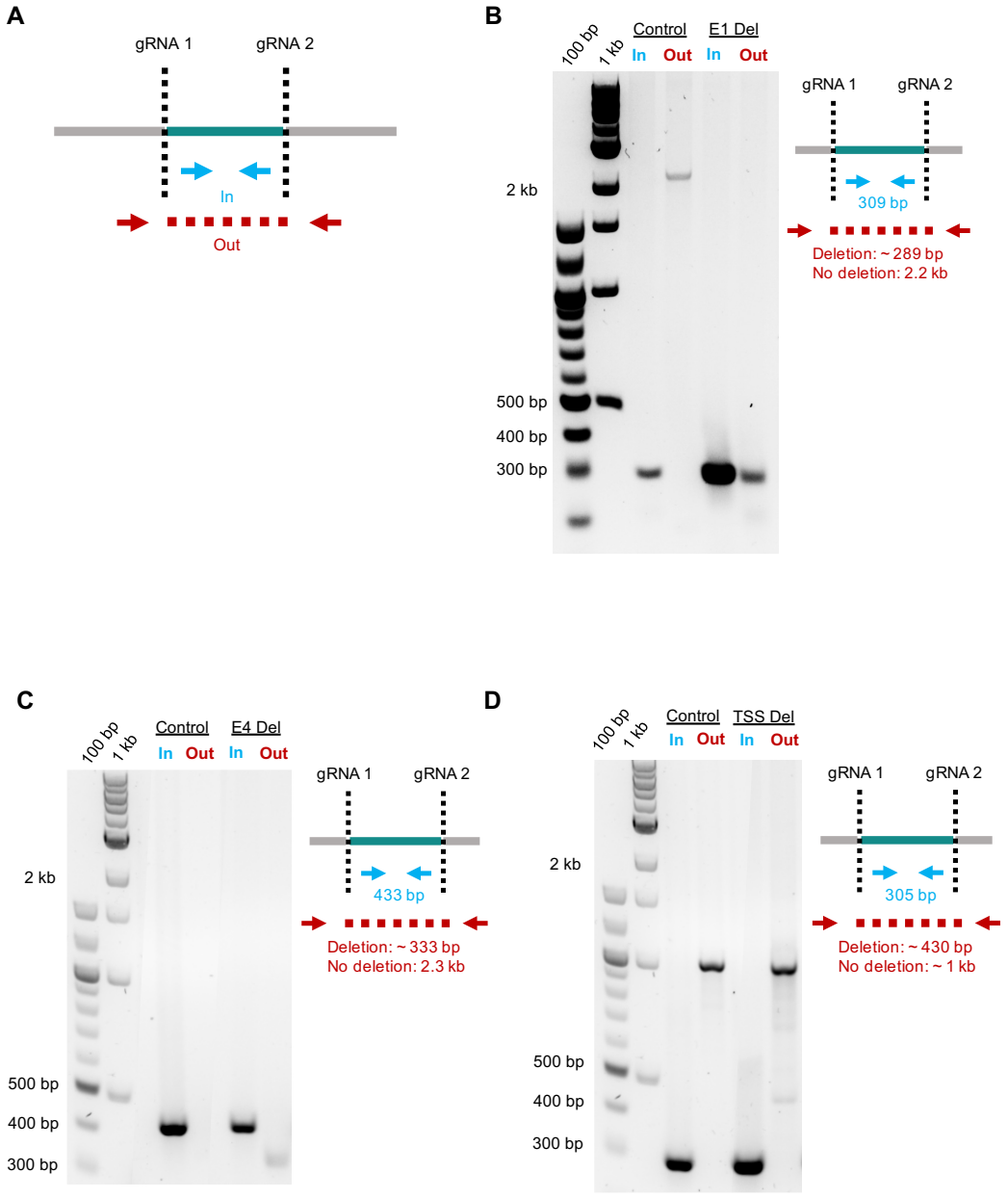


Figure 5.6: Screening for CRISPR-Cas9-mediated deletions. (A) Schematic of PCR-based deletion screening strategy. One primer pair anneals to the inside of the deletion region ("In") and the other pair anneals outside the deletion ("Out") designed at least 100 bp from the gRNA site. (B) Gel identifying amplicons with "In" and "Out" primers in control untreated HUVEC and transduced HUVEC with gRNAs targeting (B) E1, (C) E4 and (D) TSS. Schematic of expected band sizes for each gel is illustrated next to the gel. All screens were performed on a bulk population of cells with evidence in each deletion of a successfully edited proportion.

To evaluate whether editing the *CDH5* enhancers in a relatively small population of HUVEC would have an overall effect on *CDH5* expression in the bulk population, RT-qPCR analysis of the bulk populations was performed. The changes in *CDH5* expression in the bulk population of HUVEC subjected to CRISPR-Cas9 enhancer and TSS modifications showed very minimal changes compared to HUVEC subjected to CRISPR-Cas9 modification of the distal control region. This is likely due to the large proportion of unedited cells (Figure 5.7B).

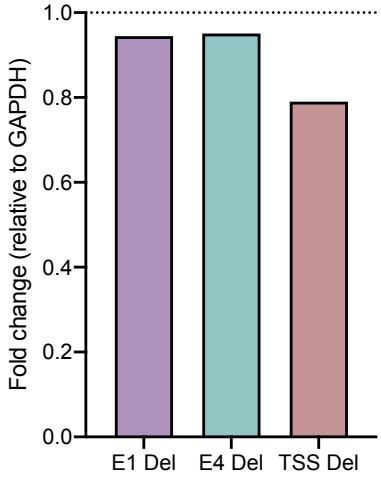


Figure 5.7: Gene expression changes in CRISPR-Cas9-modulated *CDH5* enhancers in bulk HUVEC. Relative expression of *CDH5* in CRISPR-Cas9-mediated deletions in bulk HUVEC for one replicate. Each deletion is compared to the distal control. Overall gene expression effect is minimal in this bulk population. *GAPDH* is the housekeeping gene.

5.3.4 Transduction of endothelial colony forming cells (ECFC) with enhancer deletion constructs for single-cell clone selection

Since there was evidence of a deletion for E1, E4 and TSS in a proportion of HUVEC, it was decided to use the gRNA-containing CRISPR plasmids in ECFC derived from cord blood. The rationale was to leverage the capacity of ECFC to proliferate for longer and yield clonal progeny that could be expanded as a pure population possessing the intended deletion. Deletion of enhancer, TSS or control regions was conducted in ECFC

at passage 3 as described for HUVEC. A PCR screen of transduced bulk HUVEC and ECFC was performed in parallel to determine deletion efficiency (Figure 5.8). Deletions in bulk HUVEC for E1, E4 and TSS were evident as before (Figure 5.6B-D) with effective deletions shown by the presence of fainter bands less than 500 bp with "Out" primers. However, these bands were not seen in deletions conducted in bulk ECFC using the concentration of gDNA obtained from these samples (Figure 5.8).

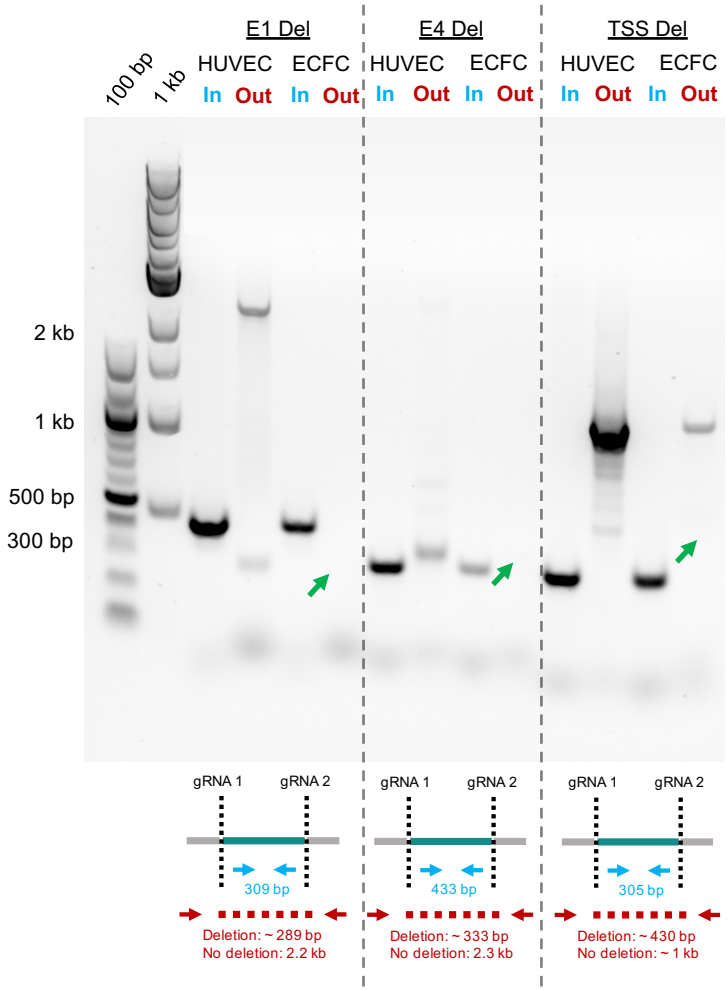


Figure 5.8: Deletion screening in bulk ECFC. "In" and "Out" primer amplifications are shown for E1, E4 and TSS deletions in both bulk HUVEC and ECFC for comparison. Evidence of deletion bands observed in HUVEC were not obvious in the ECFC population (green arrows indicate where deletion bands were expected). Schematic of expected band sizes is illustrated below the gel. Less gDNA in the ECFC sample could explain the absence of a faint deletion band as seen for HUVEC.

Given the possibility that there was a population of edited ECFC in the bulk that was simply not detected in the PCR due to the quantity of gDNA, I pursued with dilution cloning to obtain single-cell ECFC clones. For each deletion of E1, E4, TSS and distal control six 48-well plates were plated with either 6, 3 or 1.5 cells per well from the antibiotic-selected bulk ECFC to attempt to derive single cell clones. Presence of ECFC colonies was observed 7 days post seeding; yet the number of wells with colony formation was extremely small. After 10 days, appearance of ECFC colonies derived was observed and only five E1 colonies (Figure 5.8A), two TSS colonies (Figure 5.8B) and one distal control colony was obtained. No colonies were obtained for the ECFC subject to an E4 deletion. Cclones varied profoundly in morphology and proliferative capacity. Colonies such as E1 clone 2 and TSS clone 1 appeared healthy and cobblestone-like, whereas majority of cells in E1 clone 4 looked small and cells from TSS clone 2 appeared stretched (Figure 5.9A and B).

After 16 days of maintaining clones in a 48-well format, the ECFC ceased to proliferate and were collected for analysis by RT-qPCR directly. Due to the very low yield of clonal cells, only gene expression analysis was feasible. *CDH5* expression changes relative to the distal control clone showed no clear change in expression in three E1 clones and two TSS clones (Figure 5.10). In two E1 clones, a more than a two-fold increase in gene expression was found; yet these were the clones which yielded the lowest RNA concentrations, with C_T values reaching above 30 for both housekeeping gene *GAPDH* and *CDH5*. Therefore, the calculation of fold change in *CDH5* expression relative to the other clones may be unreliable. The dual gRNA CRISPR-Cas9 deletion approach was successful on a small proportion of ECs and ideally a clonal population with the intended deletion would have determined the contribution of each *CDH5* enhancer on *CDH5* expression.

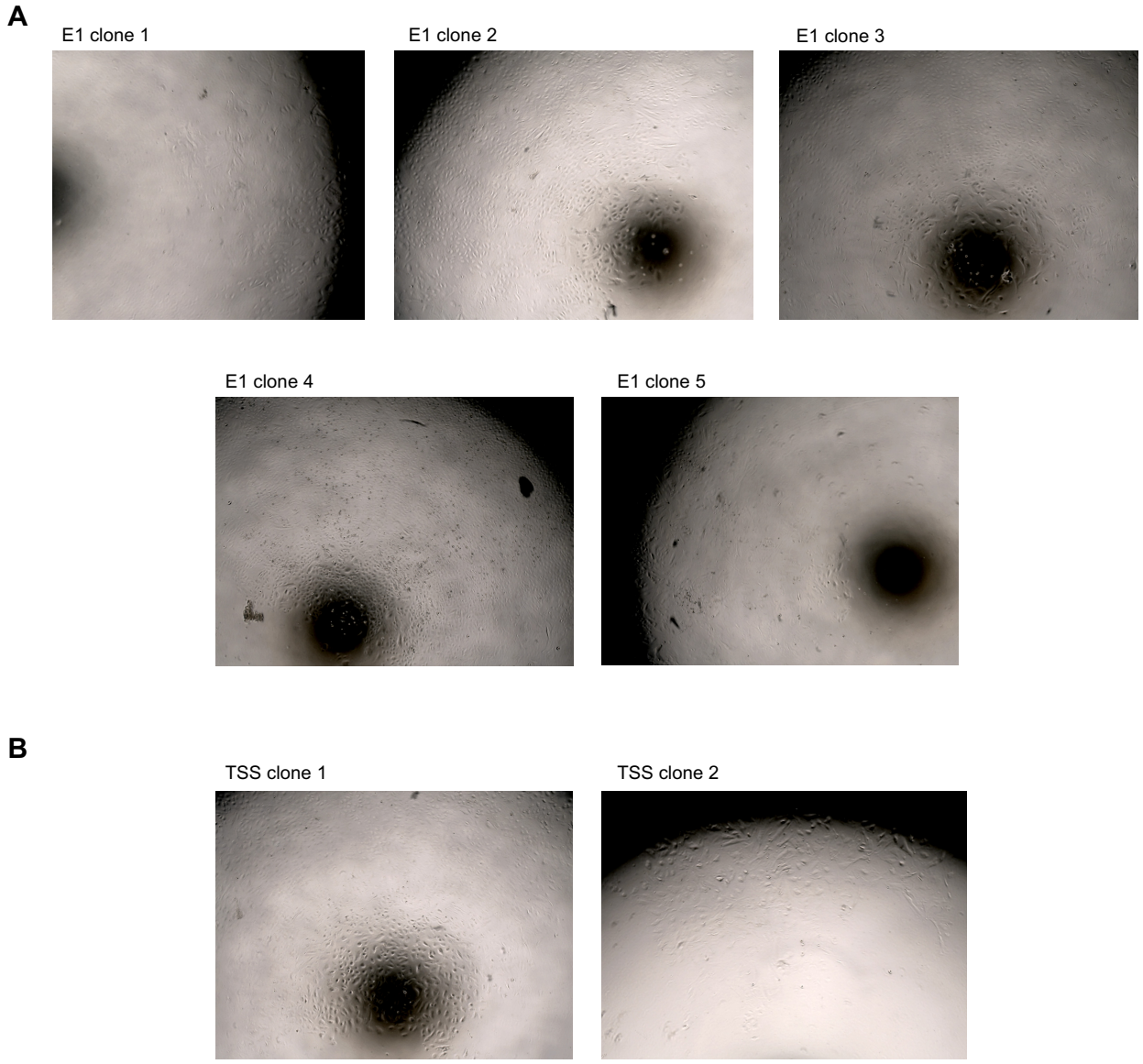


Figure 5.9: Clonal populations of CRISPR-Cas9 transduced ECFC. Brightfield images of the successful colonies 10 days after plating for single cell clones. Five colonies were visible for (A) five E1 deletions and for (B) two TSS deletions.

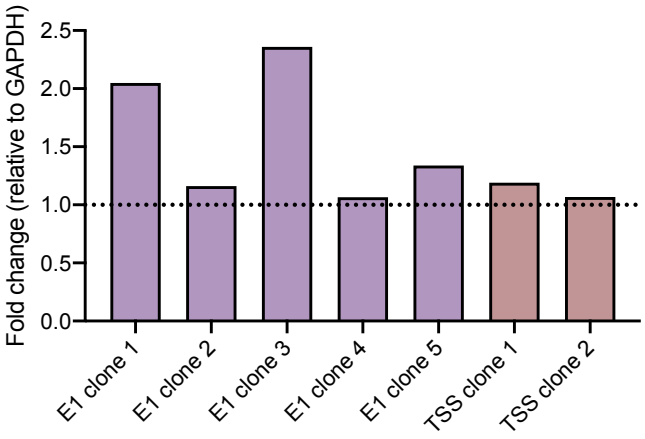


Figure 5.10: Gene expression changes in CRISPR-Cas9-modulated ECFC clones. Relative expression of *CDH5* in the seven ECFC colonies show no gene expression change with large increases in expression likely a result of high C_T values.

5.3.5 Cloning to obtain CRISPRi-modulation of *CDH5* enhancers in HUVEC

RNA-guided DNA targeting using a dCas9 protein has enabled specific genomic modulations with the advantage of having more minimal off-target effects than activated Cas9. Here I utilised a CRISPR plasmid expressing KRAB along with single gRNAs to target the centre of *CDH5* enhancer regions for transcriptional repression by heterochromatin formation. Single gRNAs were cloned into two CRISPRi plasmids to optimise high expression of KRAB-dCas9 under two different promoters; EF1α and hPGK. Screening of bacterial clones was successful, with the presence of an amplification product of ~479 bp and ~461 bp for EF1α and hPGK plasmids respectively. Cloning one gRNA was more efficient than dual gRNA cloning, with the rate of positive clones being 9/36 for E1, 8/40 for E4, 12/24 for TSS and 17/21 for proximal controls giving an overall success rate of 35%. Correct bacterial clones identified by PCR were confirmed by sequencing across the insert site of E1 (Figure 5.11A and B), E4 (Figure 5.12A and B) and the TSS (Figure 5.13A and B).

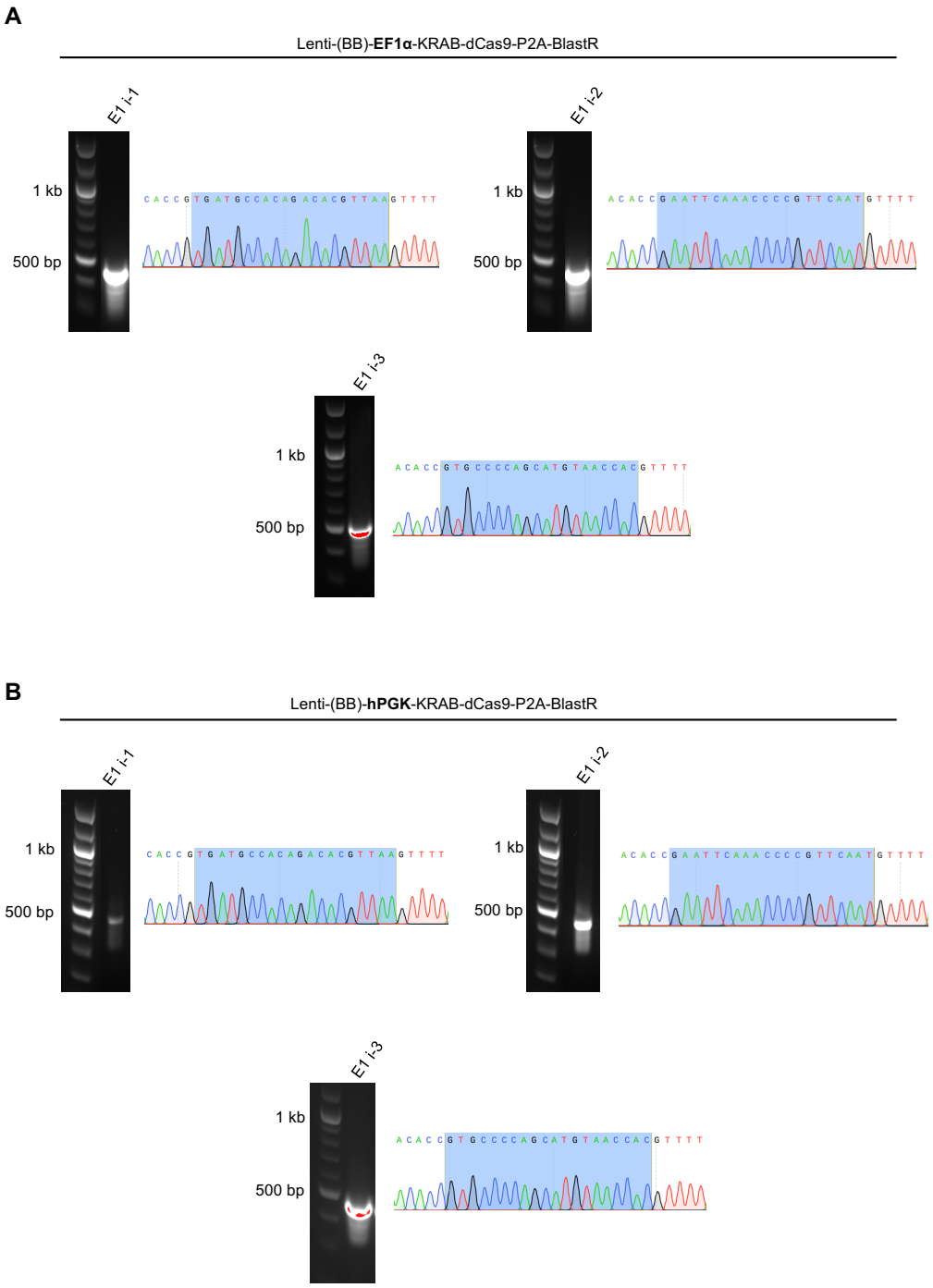


Figure 5.11: CRISPRi cloning to inhibit *CDH5* constituent enhancer E1. Gels showing correctly sized amplicons with respective gRNA sequence for all three gRNA designs for E1 in (A) EF1α-KRAB-dCas9 and (B) hPGK-KRAB-dCas9 plasmids.

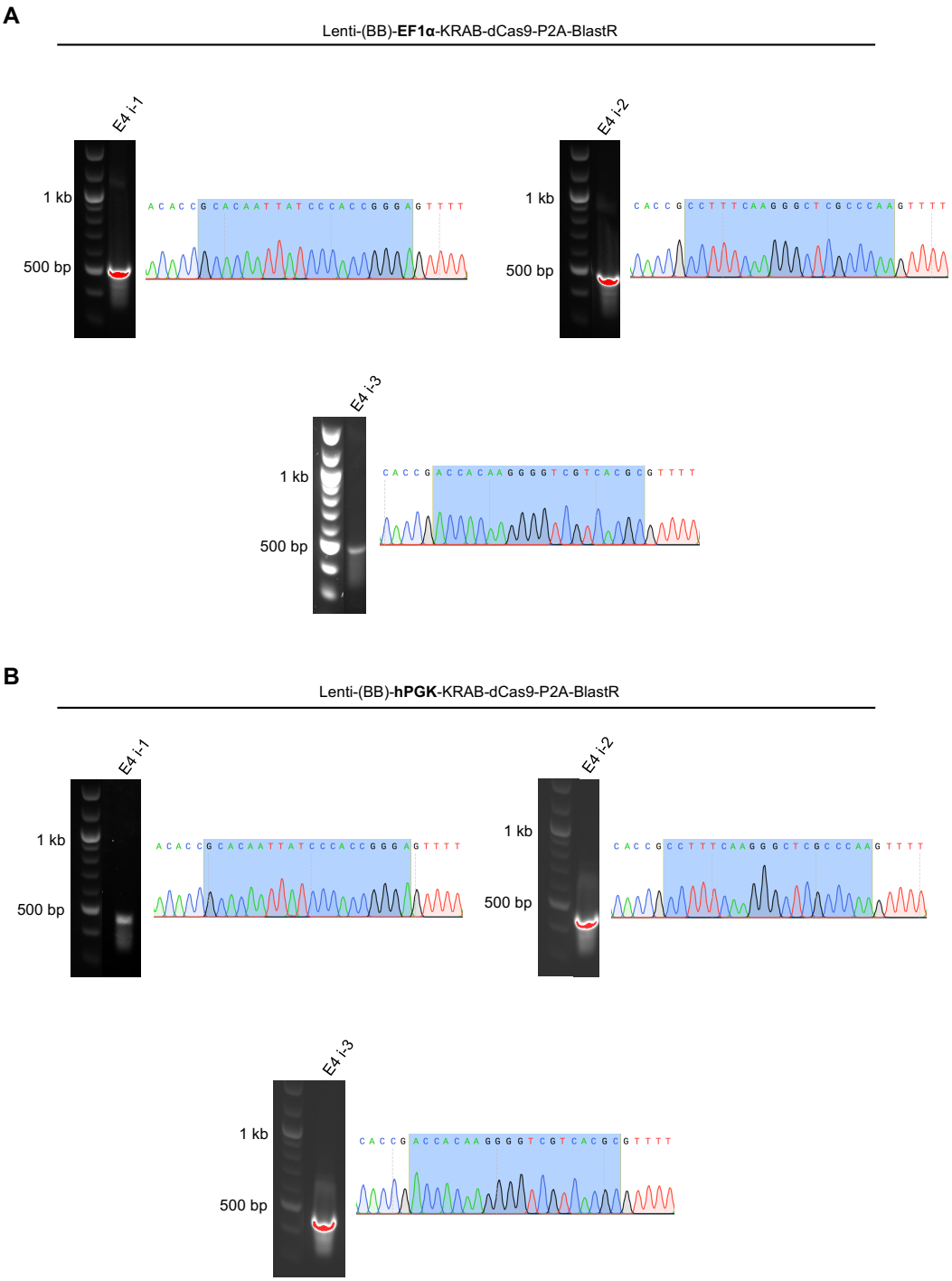


Figure 5.12: CRISPRi cloning to inhibit *CDH5* constituent enhancer E4. Gels and sequenced gRNA for both CRISPRi plasmids (A) EF1α-KRAB-dCas9 and (B) hPGK-KRAB-dCas9.

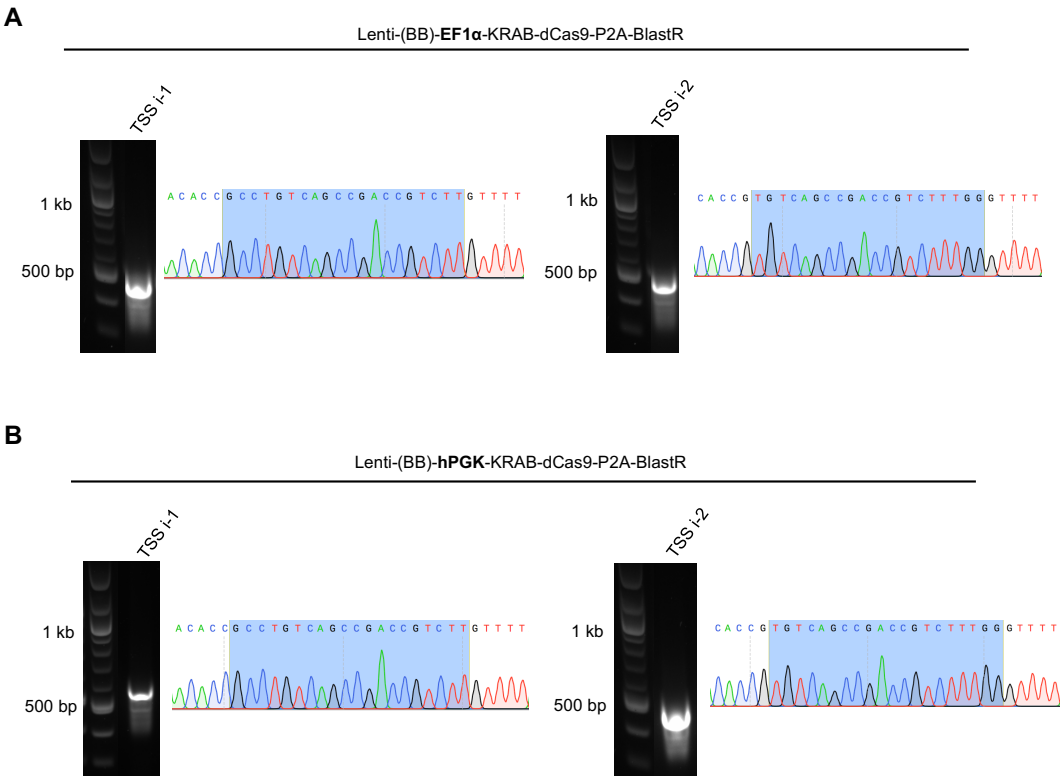


Figure 5.13: CRISPRi cloning to inhibit the *CDH5* TSS. Gels and sequenced gRNA for both CRISPRi plasmids (A) EF1α-KRAB-dCas9 and (B) hPGK-KRAB-dCas9.

5.3.6 *CDH5* regulation in CRISPRi transduced HUVEC

To test whether a bulk population of HUVEC transduced with gRNA-containing CRISPRi vectors modulated the *CDH5* constituent enhancers, the TSS targeting CRISPRi served as a positive control to measure CRISPRi activity in ECs. Gene expression changes in *CDH5* were monitored; CRISPRi with either the EF1α or hPGK promoter driving KRAB-dCas9 vectors on the TSS region did not shown the expected decrease in *CDH5* expression. In fact, gene expression seemingly showed an increase (n=1) (Figure 5.14B, red bars). Subsequently there was no effect on *CDH5* expression with E1 and E4 targeting CRISPRi plasmids (Figure 5.14B, purple and green bars). Therefore, there was no indication that the induction of heterochromatin at the enhancer or TSS regions was effective.

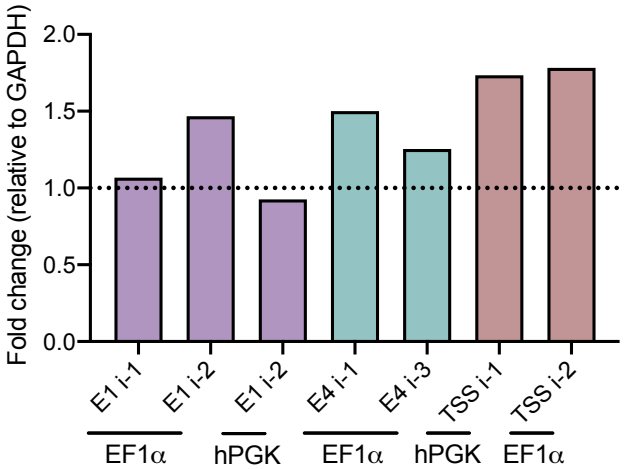


Figure 5.14: CRISPRi-mediated effect on *CDH5* expression in bulk HUVEC. *CDH5* expression changes in the CRISPRi-modulated regions with either EF1α or hPGK promoters. The bulk population showed no indication of a decrease in *CDH5* levels.

5.4 Discussion

CRISPR-Cas9 genome-editing technologies have provided the most advanced tool to study non-coding regulatory elements. In this chapter I adopted the system in pri-

mary ECs to deconstruct the structure of a super-enhancer identified in HUVEC. ERG-bound enhancer regions and the TSS of the *CDH5* super-enhancer locus were targeted to either delete enhancers, using Cas9 double-stranded breaks and dual gRNAs, or transcriptionally silence enhancers using KRAB-dCas9 inhibition without complete enhancer disruption. The outcomes of success were variable due to both technical and biological constraints highlighting the challenge that CRISPR-Cas9 editing systems pose in different contexts.

The selection of enhancer regions on the *CDH5* super-enhancer were restricted to ERG-bound regions of larger constituent enhancers. This allowed the ERG-occupied component of the enhancer to be investigated and simultaneously provided a more feasible deletion size. The locus was assessed for potential long-range chromatin interactions that may serve to regulate *CDH5*. Observations of Hi-C data in HUVEC indicates strong physical interactions within the *CDH5* gene itself. Furthermore, promoter capture Hi-C in ECFC revealed promoter interactions to the 3' end of the *CDH5* gene itself. This region has very low H3K27ac and H3K4me1 signals and no TF binding, potentially indicating that such an interaction would not be significantly involved in controlling *CDH5* levels. Ultimately, this indicated that chromosomal conformation is not likely to be involved in targeting the enhancer regions to other genes than *CDH5*.

CRISPR-mediated deletions in HUVEC appeared to be a success in both enhancers, E1 and E4, and the TSS. PCR of a bulk cell population identified fragments corresponding to the expected deletion size, although in small amounts, suggesting that a small proportion to the whole population was successfully edited. It is difficult to confirm the percentage of cells edited in this bulk population although it may seem straightforward to compare the deletion band to the non-deletion band int he same PCR. However, this comparison was confounded by PCR amplification inefficiency of the large non-deleted band, particular those that are 2 kb and larger. These PCR products were often either absent or inefficiently amplified by the "Out" deletion primers. In the samples with the deletion targeting the TSS, where the deletion band is clearer, the proportion of suc-

cessfully edited cells appears to be less than 10% (Figure 5.8). This is despite antibiotic selection to select for a pure population infected with the gRNA-harbouring CRISPR-Cas9 plasmids. Therefore, it is evident that transduction and expression from the vector is sufficient but target specificity from the designed gRNAs is ineffective. Furthermore, it is also plausible that in bulk mixed-cells population the surviving antibiotic-resistant cells, i.e. those without the intended deletion, survive better and possibly outcompete cells possessing the specific deletion. Such a circumstance would also decrease the total deleted proportion and reduce the probability of identifying correctly deleted ECFC clones.

The deletion sizes intended for the study of these enhancer elements are not unattainable. Deletions up to 1 Mb have been achieved yet at a very low deletion frequency (Canver et al. 2014). Importantly, deletion frequency has been found to inversely correlate with deletion size (Canver et al. 2014). However, double-stranded DNA breaks are not the only possible outcome of CRISPR editing. There is the possibility of Cas9 targeting the deletion junctions and making local indels in conjunction with NHEJ, without deleting the intended intervening region. Such "scarring" of DNA is a phenomenon reported in other dual gRNA studies (Canver et al. 2014).

The use of ECFC for CRISPR engineering was preferred for replicative and clonogenic capacity; however a limitation exists in enhancer selection. Here I compared H3K27ac ChIP-seq data generated in a mixed-cell population of HUVEC (approximately 10 donors contribute to the pooled HUVEC used for ChIP-seq generated for ENCODE) to that of single-donor adult ECFC generated by the Blueprint Consortium. However, I had to extrapolate ERG binding sites in ECFC from a HUVEC ERG ChIP-seq dataset. It is unknown whether the same ERG binding sites are present in ECFC; the experimental setup would be improved by performing a ChIP-seq for ERG binding in ECFC. However, since individual ECFC donors were used in this work it is likely that different donors will possess different chromatin patterns of histone modifications and TF occupancy. To this end, some ECFC donors may retain minor chromatin structural alterations or more

possibly genetic variants at the *CDH5* locus that prevent specific gRNA targeting and render the CRISPR-editing ineffective. Such chromatin structural variations have been highlighted by a number of previous CRISPR-Cas9 studies (Chakrabarti et al. 2019; Jensen et al. 2017; Scott et al. 2017).

The outcome of single-cell cloning from ECFC resulted in the formation of very few colonies. This could have been reflected in the potential cell viability effects of altering VE-cadherin levels as it plays a critical role in endothelial cell biology. VE-cadherin maintains vascular integrity without which organisation of the endothelial cytoskeleton would be dysfunctional (Giannotta et al. 2013). Endothelial cell stability could therefore be compromised making them more susceptible to cell death. For future prospectives the bulk population of CRISPR-targeted cells could be enriched for low VE-cadherin expression at the cell surface by cell sorting using fluorescence-assisted cell sorting (FACS). This would select the population of cells in which alterations have occurred.

CRISPRi has more recently emerged as a sequence-specific non-mutagenic tool for genome regulation. Its ability to tune gene expression and transcriptional elements in a reversible manner has made it a functionally favourable genetic editing alternative (Qi et al. 2013). The fused KRAB domain facilitates the spreading of heterochromatin in the form of H3K9me3 to silence chromatin which can propagate as far as 15 kb from the KRAB binding site (Groner et al. 2010) which would be expected to severely reduce the specificity of direct *CDH5* enhancer targeting. However, using CRISPRi-mediated silencing here showed no apparent modulation in *CDH5* expression highlighting potential constraints in gRNA specificity, chromatin structure or low KRAB-dCas9 efficacy.

There are several steps in the protocol which could have been adopted to improve the efficiency of genome-editing. For CRISPR deletions in primary cells more recent methods have proved to be effective. These methods adopt a protein-delivery vector, specifically CRISPR ribonucleoproteins (RNPs) which contain the already expressed Cas9 protein in combination with the gRNAs. These are commonly delivered directly into primary

cells using electroporation (Mangeot et al. 2019). This has further been shown to limit the off-target effects of Cas9. For CRISPRi, the limitation in the ability to determine the effect of KRAB-dCas9 in HUVEC could have somewhat been determined by quantifying levels of KRAB protein expression to confirm the presence of KRAB-dCas9 in the cells.

6 SUMMARY AND CONCLUSIONS

In the work presented here I have demonstrated how the TF ERG acts as a master regulator in ECs via super-enhancer. By employing multi-level omics datasets I reveal genomic mechanisms that are employed in EC under a homeostatic state. In chapter 2, I identify endothelial super-enhancers and reveal dense occupancy of ERG at these regions. A core subset of endothelial super-enhancers were regulated by ERG, showing a role for ERG in epigenetic regulation in EC. I further report the enrichment of genomic variants associated with cardiovascular diseases at ERG-defined super-enhancers.

In chapter 3, I compare transcriptome and chromatin profiles between EC and ERG-positive VCaP prostate cancer cells revealing distinct ERG occupancy and super-enhancer landscapes. The cell-type differences seem to be defined by transcriptional complex composition and combinatorial TF assembly. I explore the TF cooperativity in EC in chapter 4, defining super-enhancers with TFs GATA2, cFOS and cJUN. I suggest a mechanism whereby p300 co-activator recruitment by ERG control the regulation of the core subset of ERG-dependent super-enhancers. By combining ChIP-seq data in HUVEC for histone marks, TFs and cofactors I partition the chromatin using unsupervised clustering and link the clusters functionally to gene expression levels and structural chromatin interactions.

Finally, in chapter 5, I home in to dissect the super-enhancer regulating VE-cadherin in ECs. I adopt two CRISPR-based methodologies to assess the contribution of constituent enhancers to VE-cadherin expression; I attempt to disrupt enhancers using CRISPR-Cas9 and silence enhancers using CRISPR-dCas9-KRAB. In this summary and conclusions I will provide an overview of the main results and discuss the implications of the findings. I will consider the future work in this area to further address and uncover the molecular mechanisms of chromatin and gene regulation in ECs.

6.1 Experimental limitations

The work presented in this thesis demonstrates the central role of ERG in maintaining an endothelial identity via super-enhancers. However, several experiments present limitations that preclude strong and definite conclusions. A key limitation is the lack of replicates in all the ChIP-seq experiments presented here. It may be argued that the use of pooled HUVEC from 10 individual donors provides some biological replication however increasing replicates would account for technical noise of ChIP-seq and hence increase the reliability of binding site identification. Moreover, three replicates are required to perform differential binding analysis to attain true statistical significance and account for potential false positives sites.

Further limitations of this work exist in the study of cell homeostasis which by definition is the balance of the several factors that maintain a healthy cell. In this thesis there is only one experiment to challenge HUVEC with prosurvival factor Ang-1 with a extracellular stimulus to activate downstream mechanisms and investigate the recovery of these cells back to homeostasis. One potential experiment could have been the stimulation with proinflammatory cytokines, such as TNF-α, which would model chronic disease. Such a study would investigate the changes to the transcriptomic and chromatin landscape upon this stimulation. Further interesting experiments could have modelled the physiological state of endothelial cells by exposing them to different flow rates, both physiological and disturbed flow which occurs in different vascular beds.

6.2 Specificities of transcription factor binding directed by sequence composition

TFs bind DNA to activate or repress gene transcription. TF combinations compete and collaborate to regulate common genes. From integration of ERG ChIP-seq and ERG-dependent gene expression profiling in HUVEC, I found genes that were directly

regulated by ERG. Yet, there were also many ERG binding sites that were not contributing to gene regulation; a phenomenon reported for other studied integrating ChIP with gene expression profiles (Gorski et al. 2011; Hull et al. 2013). It has been postulated that these TF binding events could be binding loci for enhancer "priming". This would facilitate subsequent activation to a stimulus for instance (Alasoo et al. 2018) or, otherwise, may serve as placeholders to prevent nucleosomal repositioning.

In fact, I found that ERG showed preferential occupancy of its core consensus motif flanked 5' by an adenosine, AGGAA, at enhancers as opposed to approximately equal occupancy of AGGAA and CGGAA at promoters (chapter 2). Biochemical studies of these sequences have shown a considerably lower affinity for AGGAA, reflecting the fact that ERG is interacting transiently at enhancers. There was no indication that either the presence of adenosine (A) or cytosine (C) was associated with the modulation status of a gene. This could be a result of other TFs and their assembly with ERG at any given locus. Indeed, Boeva (2016) has studied the motif sequence of another ETS factor Spi1/PU.1 in combination with KLF family motifs and determined that motif orientation and spacing partly dictate Spi1/PU.1 regulated genes suggesting cooperative interactions are partly responsible for gene regulatory mechanisms.

In chapter 4, I determined chromatin states by k-means clustering and interrogate the DNA sequence for putative TF binding. I use motif analysis and reveal a notable difference in the candidate TFs that harbour promoter and enhancer clusters. Similarly, a study investigating multiple cell lines, tissues and primary cells using *in silico* methods also reported this difference (Vandel et al. 2019). Although the underlying DNA sequence facilitates the binding position of TFs, these preferences are also guided by interplay from chromatin structural conformation and biophysical properties.

6.3 Transcription factor combinations in chromatin accessibility

Current models suggest that a small subset of TFs, pioneer TFs, can remodel condensed chromatin for accessibility (Heinz et al. 2015). Pioneer TFs can recruit chromatin remodelling factors and are best characterised in development. Previous reports, in non-ECs, have be inconclusive in determining whether ERG is a pioneer TF. Here, I show that ERG regulates a subset of H3K27ac sites; a histone modification determined to correlate highly with open regions as defined by ATAC-seq or DNase-seq (Lara-Astiaso et al. 2014). Therefore, the modulation of select H3K27ac sites by ERG means that it may have pioneer properties in certain contexts. In non-ECs, ERG was determined to show pioneer action on a cryptic enhancer of the gene *SOX9* (Cai et al. 2013). Recent reports have implicated the action of collaborative action of TFs in chromatin accessibility. A study showed that factors Oct4, Sox2 and Klf4 were required in combination to open enhancers linked to pluripotency in embryonic stem cells (Chronis et al. 2017). Similarly, Foxa2 can lead to changes in accessibility when co-bound with additional TFs at enhancers required for endoderm differentiation (Cernilogar et al. 2019). Finally, another study proposed that pioneer TFs can bind closed chromatin, yet it is the action of cooperative non-pioneer TFs that drives chromatin opening (Mayran et al. 2019). It has been postulated that TFs can synergise to recruit multiple chromatin remodelling complexes. Indeed, in chapter 4, I showed enhanced recruitment of H3K27ac and cofactor BRD4 as the number of TFs increased. Future work should elucidate the combinatorial logic to better understand pioneer and non-pioneer TF binding and chromatin recruitment in a cell type-specific manner.

6.4 Validity of the super-enhancer concept

The distinction of super-enhancers using the criteria established in the seminal reports requires two critical steps: (1) enhancer clustering and (2) cluster ranking by enrichment (Hnisz et al. 2013; Whyte et al. 2013). Both steps display bias to a degree. In the case of (1), enhancers are clustered within specifically 12.5 kb of one another, a default distance, raising the issue that enhancer clustering may be represented more accurately by a different distance. The interesting concept for (2) is that several clusters consist simply of only one enhancer with very high enrichment of the factor used for ranking. In my super-enhancer analysis (chapter 2), I identified super-enhancers associated with key endothelial genes and also general housekeeping genes, contradicting the idea that super-enhancers genes are almost entirely cell type-specific. Conversely, it was reported that many highly transcribed genes are not under the control of a super-enhancer (Moorthy et al. 2017). To overcome some of these issues, other methodologies for detection of regulatory elements linked to cell identity genes have been established (Madani Tonekaboni et al. 2019). In chapter 5, I attempted to dissect the super-enhancer of the highly endothelial specific gene VE-cadherin endogenously to evaluate whether these act either additively or redundantly to regulate gene expression as these concepts remain widely disputed (Hay et al. 2016; Moorthy et al. 2017)

6.5 Harnessing disease-associated variants for prioritisation

One of the most interesting challenges in the post genome-wide association studies (GWAS) era is elucidating the molecular mechanisms of the large number of disease-associated variants identified. In chapter 2, I reveal the high enrichment of variants associated with risk for CVD at ERG-defined super-enhancers. Future work should further prioritise these variants to distinguish which are likely to disrupt endothelial

genes causal for CVD, given the fact that vascular dysfunction predisposes to a number of CVD traits (Rajendran et al. 2013). Samples sequenced for GWAS are obtained from blood cells, thus uninformative for tissue-specificity in genetic disease. To uncover tissue-specific contributions, resources such as the Genotype-Tissue Expression (GTEx) project contain donor-level information from over 50 different tissues including quantitative trait loci (QTLs), mainly expression QTLs (GTEx Consortium 2013). QTLs are locus with a variation that correlates with a trait, such as expression level. However, this still does not delineate EC and other specific cell types. One possibility is generating models of disease variants using CRISPR-Cas9 single nucleotide editing in EC as performed for other cell types (Gupta et al. 2017). Further work should integrate the single-cell profiling including single-cell (sc)RNA-seq and scATAC-seq with GWAS and associated QTLs. Interestingly, a study combining scRNA-seq and GWAS showed that a number of CVD traits (coronary artery disease, high blood pressure, systolic blood pressure and diastolic blood pressure) had the strongest association with ECs compared to other cell types (Watanabe et al. 2019). Indeed, analysis of specific CVD traits at ERG super-enhancers in EC similarly showed significant enrichment for variants associated with coronary artery disease and hypertension (high blood pressure). This demonstrates that integration of disease variants with functional genomics datasets lead us closer to understanding causal mechanisms of genetic disease such as CVD.

6.6 ERG as a master regulator of endothelial cell super-enhancers

ERG is a TF involved in inter-dependent functions at the genome level. Whyte et al. (2013) showed profound loss of super-enhancer-associated genes relative to other genes when levels of TF Oct4 were reduced in embryonic stem cells. In chapter 2, I demonstrate a correlation between endothelial super-enhancer-associated genes and ERG-driven genes. However, in chapter 4, I show that only a core subset of super-

enhancers are significantly regulated by ERG and present a model of ERG-dependent super-enhancer assembly in EC whereby TF complex formation governs super-enhancer activation. Previous work has identified VEGF-regulated compartments in EC and postulated that genes coregulated by VEGF are under the control of a few highly activated enhancers (Kaikkonen et al. 2014). This may be the case for the core subset of ERG-dependent super-enhancers as chromatin interactions allow for the regulation of multiple genes from a single super-enhancer. The model in chapter 4 also highlights differences in TF cooperativity at different sites with GATA2 and AP-1 members cFOS and cJUN. In this work, other ETS factors are not considered, yet are likely to be important. Regions bound by multiple FOX family members have been shown to be associated with higher functionality than a single member (Chen et al. 2016). The TF cooperativity in EC is also likely to vary according to EC subtype. In murine EC from brain, liver, lung and kidney ERG expression at mRNA level was consistently high with other TFs showing tissue-restricted expression (Sabbagh et al. 2018).

In conclusion, this study provides novel evidence on the transcriptional and epigenetic mechanisms which ERG controls in EC to drive lineage-specific gene expression. These associations provide valuable insights for investigating the role of ERG-dependent regulatory programs in maintaining endothelial homeostasis and protecting against vascular diseases.

6.7 Future work

The work presented in this thesis integrates a number of datasets from previously published studies. There are other interesting datasets that would be useful to extend the findings in the current work. These include a timecourse experiment with VEGF activation at 0 h, 1 h, 4 h and 12 h for assessing the binding kinetics of ETS TF ETS1, coactivator p300 and enrichment of H3K27ac by ChIP-seq (Zhang et al. 2013). Another valuable dataset that could have been utilised is a global nuclear run-on coupled to deep

sequencing (GRO-seq) which identifies nascent RNA transcripts (Kaikkonen et al. 2013). This data would allow active enhancers in the HUVEC genome to be validated by quantifying transcriptional activity at these regions by GRO-seq.

The current interest in the field of vascular biology has been elucidating the tissue and organ specificity of endothelial cells to delineate vascular-bed-specific functions. This requires experiments that allow much higher resolution than those utilised in this present work yet due to advances in single-cell omics techniques this is being widely achieved. Following this work, it would be interesting to observe the changes at the transcriptome and epigenome level in response to ERG knockout by single-cell (sc)RNA-seq and scATAC-seq respectively. Such tissue-specific experiments would have to use mouse models where integration of ATAC and RNA at the single-cell level could be used to correlate TF expression levels with chromatin accessibility. The heterogeneity at the level of chromatin could further be studied using single-cell ChIP-seq. Overall, the changes at the single-cell level in response to ERG depletion would be interesting as it may find a subset of cells most sensitive to the loss of ERG in a particular tissue.

REFERENCES

Abdel-Malak, N. A., C. B. Srikant, A. S. Kristof, S. A. Magder, J. A. Di Battista, and S. N. A. Hussain (2008). "Angiopoietin-1 promotes endothelial cell proliferation and migration through AP-1-dependent autocrine production of interleukin-8". *Blood* 111.8, pp. 4145–4154.

Abrahimi, P., W. G. Chang, M. S. Kluger, Y. Qyang, G. Tellides, W. M. Saltzman, and J. S. Pober (2015). "Efficient Gene Disruption in Cultured Primary Human Endothelial Cells by CRISPR/Cas9". *Circulation Research* 117.2, pp. 121–128.

Adam, R. C., H. Yang, S. Rockowitz, S. B. Larsen, M. Nikolova, D. S. Oristian, L. Polak, M. Kadaja, A. Asare, et al. (2015). "Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice." *Nature* 521.7552, pp. 366–370.

Adamo, P. and M. R. Ladomery (2016). "The oncogene ERG: a key factor in prostate cancer". *Oncogene* 35.4, pp. 403–414.

Adams, R. H. and K. Alitalo (2007). "Molecular regulation of angiogenesis and lymphangiogenesis". *Nature Reviews Molecular Cell Biology* 8.6, pp. 464–478.

Aird, W. C. (2007). "Phenotypic heterogeneity of the endothelium: I. Structure, function, and mechanisms." *Circulation research* 100.2, pp. 158–173.

Alasoo, K., J. Rodrigues, S. Mukhopadhyay, A. J. Knights, A. L. Mann, K. Kundu, C. Hale, G. Dougan, and D. J. Gaffney (2018). "Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response". *Nature Genetics* 50.3, pp. 424–431.

Alver, B. H., K. H. Kim, P. Lu, X. Wang, H. E. Manchester, W. Wang, J. R. Haswell, P. J. Park, and C. W. M. Roberts (2017). "The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers". *Nature Communications* 8.1, pp. 14648–14658.

An, J., S. Ren, S. J. Murphy, S. Dalangood, C. Chang, X. Pang, Y. Cui, L. Wang, Y. Pan, et al. (2015). "Truncated ERG Oncoproteins from TMPRSS2-ERG Fusions Are Resistant to SPOP-Mediated Proteasome Degradation". *Molecular Cell* 59.6, pp. 904–916.

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, et al. (2014). "An atlas of active enhancers across human cell types and tissues". *Nature* 507.7493, pp. 455–461.

André, P., C. V. Denis, J. Ware, S. Saffaripour, R. O. Hynes, Z. M. Ruggeri, and D. D. Wagner (2000). "Platelets adhere to and translocate on von Willebrand factor presented by endothelium in stimulated veins." *Blood* 96.10, pp. 3322–3328.

Ardlie, K. G., D. S. Deluca, A. V. Segre, T. J. Sullivan, T. R. Young, E. T. Gelfand, C. A. Trowbridge, J. B. Maller, T. Tukiainen, et al. (2015). "The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans". *Science* 348.6235, pp. 648–660.

Asangani, I. A., V. L. Dommeti, X. Wang, R. Malik, M. Cieslik, R. Yang, J. Escara-Wilke, K. Wilder-Romans, S. Dhanireddy, et al. (2014). "Therapeutic targeting of

BET bromodomain proteins in castration-resistant prostate cancer". *Nature* 510.7504, pp. 278–282.

Augustin, H. G. and G. Y. Koh (2017). "Organotypic vasculature: From descriptive heterogeneity to functional pathophysiology". *Science* 357.6353, Article eaal2379.

Banerji, J., S. Rusconi, and W. Schaffner (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." *Cell* 27.2, pp. 299–308.

Barski, A., S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao (2007). "High-Resolution Profiling of Histone Methylations in the Human Genome". *Cell* 129.4, pp. 823–837.

Basuyaux, J. P., E. Ferreira, D. Stéhelin, and G. Butticè (1997). "The Ets transcription factors interact with each other and with the c-Fos/c-Jun complex via distinct protein domains in a DNA-dependent and -independent manner." *The Journal of Biological Chemistry* 272.42, pp. 26188–95.

Battle, S. L., N. D. Jayavelu, R. N. Azad, J. Hesson, F. N. Ahmed, E. G. Overbey, J. A. Zoller, J. Mathieu, H. Ruohola-Baker, et al. (2019). "Stem Cell Reports Resource Enhancer Chromatin and 3D Genome Architecture Changes from Naive to Primed Human Embryonic Stem Cell States". *Stem Cell Reports* 12, pp. 1129–1144.

Bedford, D. C., L. H. Kasper, T. Fukuyama, and P. K. Brindle (2010). "Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases." *Epigenetics* 5.1, pp. 9–15.

Benner, C., S. Konovalov, C. Mackintosh, K. R. Hutt, R. Stunnenberg, and I. Garcia-Bassets (2013). "Decoding a signature-based model of transcription cofactor recruitment dictated by cardinal cis-regulatory elements in proximal promoter regions." *PLoS Genetics* 9.11, Article e1003906.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, et al. (2006). "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells". *Cell* 125.2, pp. 315–326.

Birdsey, G. M., N. H. Dryden, A. V. Shah, R. Hannah, M. D. Hall, D. O. Haskard, M. Parsons, J. C. Mason, M. Zvelebil, et al. (2012). "The transcription factor Erg regulates expression of histone deacetylase 6 and multiple pathways involved in endothelial cell migration and angiogenesis". *Blood* 119.3, pp. 894–903.

Birdsey, G. M., A. V. Shah, N. Dufton, L. E. Reynolds, L. Osuna Almagro, Y. Yang, I. M. Aspalter, S. T. Khan, J. C. Mason, et al. (2015). "The Endothelial Transcription Factor ERG Promotes Vascular Stability and Growth through Wnt/β-Catenin Signaling". *Developmental Cell* 32.1, pp. 82–96.

Birdsey, G. M., N. H. Dryden, V. Amsellem, F. Gebhardt, K. Sahnan, D. O. Haskard, E. Dejana, J. C. Mason, and A. M. Randi (2008). "Transcription factor Erg regulates angiogenesis and endothelial apoptosis through VE-cadherin." *Blood* 111.7, pp. 3498–3506.

Blee, A. M., S. Liu, L. Wang, and H. Huang (2016). "BET bromodomain-mediated interaction between ERG and BRD4 promotes prostate cancer cell invasion". *Oncotarget* 7.25, pp. 38319–38332.

Boeva, V. (2016). "Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells." *Frontiers in Genetics* 7.24, pp. 1–15.

Boyer, L. A., T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, et al. (2005). "Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells". *Cell* 122.6, pp. 947–956.

Brown, J. D., Z. B. Feldman, S. P. Doherty, J. M. Reyes, P. B. Rahl, C. Y. Lin, Q. Sheng, Q. Duan, A. J. Federation, et al. (2018). "BET bromodomain proteins regulate enhancer function during adipogenesis". *Proceedings of the National Academy of Sciences* 115.9, pp. 2144–2149.

Brown, J. D., C. Y. Lin, Q. Duan, G. Griffin, A. J. Federation, R. M. Paranal, S. Bair, G. Newton, A. H. Lichtman, et al. (2014). "NF-κB directs dynamic super enhancer formation in inflammation and atherogenesis." *Molecular Cell* 56.2, pp. 219–31.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position". *Nature Methods* 10.12, pp. 1213–1218.

Cai, C., H. Wang, H. H. He, S. Chen, L. He, F. Ma, L. Mucci, Q. Wang, C. Fiore, et al. (2013). "ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer". *The Journal of Clinical Investigation* 123.3, pp. 1109–1122.

Camuzeaux, B., C. Spriet, L. Héliot, J. Coll, and M. Duterque-Coquillaud (2005). "Imaging Erg and Jun transcription factor interaction in living cells using fluorescence resonance energy transfer analyses". *Biochemical and Biophysical Research Communications* 332.4, pp. 1107–1114.

Canver, M. C., D. E. Bauer, A. Dass, Y. Y. Yien, J. Chung, T. Masuda, T. Maeda, B. H. Paw, and S. H. Orkin (2014). "Characterization of Genomic Deletion Efficiency Mediated by Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/Cas9 Nuclease System in Mammalian Cells". *Journal of Biological Chemistry* 289.31, pp. 21312–21324.

Carlson, T. R., Y. Feng, P. C. Maisonpierre, M. Mrksich, and A. O. Morla (2001). "Direct Cell Adhesion to the Angiopoietins Mediated by Integrins". *Journal of Biological Chemistry* 276.28, pp. 26516–26525.

Carmeliet, P. (2003). "Angiogenesis in health and disease". *Nature Medicine* 9.6, pp. 653–660.

Carrère, S., A. Verger, A. Flourens, D. Stehelin, and M. Duterque-Coquillaud (1998). "Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes via two distinct domains". *Oncogene* 16.25, pp. 3261–3268.

Cernilogar, F. M., S. Hasenöder, Z. Wang, K. Scheibner, I. Burtscher, M. Sterr, P. Smialowski, S. Groh, I. M. Evenroed, et al. (2019). "Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2". *Nucleic Acids Research* 47.17, pp. 9069–9086.

Chakrabarti, A. M., T. Henser-Brownhill, J. Monserrat, A. R. Poetsch, N. M. Luscombe, and P. Scaffidi (2019). "Target-Specific Precision of CRISPR-Mediated Genome Editing". *Molecular Cell* 73.4, pp. 699–713.

Chakravarty, D., A. Sboner, S. S. Nair, E. Giannopoulou, R. Li, S. Hennig, J. M. Mosquera, J. Pauwels, K. Park, et al. (2014). "The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer". *Nature Communications* 5.1, Article 5383.

Chang, J., W. Xu, X. Du, and J. Hou (2018). "MALAT1 silencing suppresses prostate cancer progression by upregulating miR-1 and downregulating KRAS." *OncoTargets and Therapy* 11, pp. 3461–3473.

Chapuy, B., M. R. McKeown, C. Y. Lin, S. Monti, M. G. Roemer, J. Qi, P. B. Rahl, H. H. Sun, K. T. Yeda, et al. (2013). "Discovery and Characterization of Super-Enhancer-Associated Dependencies in Diffuse Large B Cell Lymphoma". *Cancer Cell* 24.6, pp. 777–790.

Chen, J.-x., M. L. Lawrence, G. Cunningham, B. W. Christman, and B. Meyrick (2004). "HSP90 and Akt modulate Ang-1-induced angiogenesis via NO in coronary artery endothelium". *Journal of Applied Physiology* 96.2, pp. 612–620.

Chen, L. and Z. S. Qin (2016). "traseR: an R package for performing trait-associated SNP enrichment analysis in genomic intervals: Table 1." *Bioinformatics* 32.8, pp. 1214–1216.

Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, et al. (2008). "Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells". *Cell* 133.6, pp. 1106–1117.

Chen, Y., P. Chi, S. Rockowitz, P. J. Iaquinta, T. Shamu, S. Shukla, D. Gao, I. Sirota, B. S. Carver, et al. (2013). "ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss." *Nature Medicine* 19.8, pp. 1023–9.

Chistiakov, D. A., A. N. Orekhov, and Y. V. Bobryshev (2017). "Effects of shear stress on endothelial cells: go with the flow". *Acta Physiologica* 219.2, pp. 382–408.

Chng, K. R., C. W. Chang, S. K. Tan, C. Yang, S. Z. Hong, N. Y. W. Sng, and E. Cheung (2012). "A transcriptional repressor co-regulatory network governing androgen response in prostate cancers." *The EMBO Journal* 31.12, pp. 2810–2823.

Chronis, C., P. Fiziev, B. Papp, S. Sabri, J. Ernst, and K. P. Correspondence (2017). "Cooperative Binding of Transcription Factors Orchestrates Reprogramming". *Cell* 168, pp. 442–459.

Cirillo, L. A., C. E. McPherson, P. Bossard, K. Stevens, S. Cherian, E. Y. Shim, K. L. Clark, S. K. Burley, and K. S. Zaret (1998). "Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome." *The EMBO Journal* 17.1, pp. 244–254.

Cirillo, L. A., F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, and K. S. Zaret (2002). "Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4." *Molecular Cell* 9.2, pp. 279–289.

Collett, J. A., P. Mehrotra, A. Crone, W. C. Shelley, M. C. Yoder, and D. P. Basile (2017). "Endothelial colony-forming cells ameliorate endothelial dysfunction via secreted factors following ischemia-reperfusion injury". *American Journal of Physiology-Renal Physiology* 312.5, pp. 897–907.

Consortium, T. 1. G. P. (2015). "A global reference for human genetic variation". *Nature* 526.7571, pp. 68–74.

Corada, M., F. Orsenigo, M. F. Morini, M. E. Pitulescu, G. Bhat, D. Nyqvist, F. Breviario, V. Conti, A. Briot, et al. (2013). "Sox17 is indispensable for acquisition and maintenance of arterial identity". *Nature Communications* 4.1, Article 2609.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." *Proceedings of the National Academy of Sciences* 107.50, pp. 21931–6.

Dai, X., W. Gan, X. Li, S. Wang, W. Zhang, L. Huang, S. Liu, Q. Zhong, J. Guo, et al. (2017). "Prostate cancer–associated SPOP mutations confer resistance to BET inhibitors through stabilization of BRD4". *Nature Medicine* 23.9, pp. 1063–1071.

Davies, P. F., M. Civelek, Y. Fang, and I. Fleming (2013). "The atherosusceptible endothelium: endothelial phenotypes in complex haemodynamic shear stress regions in vivo." *Cardiovascular Research* 99.2, pp. 315–327.

De Val, S. and B. L. Black (2009). "Transcriptional Control of Endothelial Cell Development". *Developmental Cell* 16.2, pp. 180–195.

De Val, S., N. C. Chi, S. M. Meadows, S. Minovitsky, J. P. Anderson, I. S. Harris, M. L. Ehlers, P. Agarwal, A. Visel, et al. (2008). "Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors." *Cell* 135.6, pp. 1053–64.

DeMare, L. E., J. Leng, J. Cotney, S. K. Reilly, J. Yin, R. Sarro, and J. P. Noonan (2013). "The genomic landscape of cohesin-associated chromatin interactions." *Genome Research* 23.8, pp. 1224–1234.

Denayer, S., C. Helsen, L. Thorrez, A. Haelens, and F. Claessens (2010). "The rules of DNA recognition by the androgen receptor." *Molecular Endocrinology* 24.5, pp. 898–913.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions". *Nature* 485.7398, pp. 376–380.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner". *Bioinformatics* 29.1, pp. 15–21.

Doudna, J. A. and E. Charpentier (2014). "The new frontier of genome engineering with CRISPR-Cas9". *Science* 346.6213, pp. 1258096–1258096.

Du, Z., T. Fei, R. G. W. Verhaak, Z. Su, Y. Zhang, M. Brown, Y. Chen, and X. S. Liu (2013). "Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer". *Nature Structural & Molecular Biology* 20.7, pp. 908–913.

Dubois, C., X. Liu, P. Claus, G. Marsboom, P. Pokreisz, S. Vandenwijngaert, H. Dépelteau, W. Streb, L. Chaothawee, et al. (2010). "Differential Effects of Progenitor Cell

Populations on Left Ventricular Remodeling and Myocardial Neovascularization After Myocardial Infarction". *Journal of the American College of Cardiology* 55.20, pp. 2232–2243.

Dufton, N. P., C. R. Peghaire, L. Osuna-Almagro, C. Raimondi, V. Kalna, A. Chuahan, G. Webb, Y. Yang, G. M. Birdsey, et al. (2017). "Dynamic regulation of canonical TGFβ signalling by endothelial transcription factor ERG protects from liver fibrogenesis". *Nature Communications* 8.1, Article 895.

Eichmann, A., C. Corbel, V. Nataf, P. Vaigot, C. Bréant, and N. M. L. Douarin (1997). "Ligand-dependent development of the endothelial and hemopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor receptor 2". *Proceedings of the National Academy of Sciences* 94.10, pp. 5141–5146.

Elcheva, I., V. Brok-Volchanskaya, A. Kumar, P. Liu, J.-H. Lee, L. Tong, M. Vodyanik, S. Swanson, R. Stewart, et al. (2014). "Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators". *Nature Communications* 5, pp. 4372–4383.

ENCODE Project Consortium (2012). "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489.7414, pp. 57–74.

Ernst, J., A. Melnikov, X. Zhang, L. Wang, P. Rogov, T. S. Mikkelsen, and M. Kellis (2016). "Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions". *Nature Biotechnology* 34.11, pp. 1180–1190.

Fabregat, A., K. Sidiropoulos, P. Garapati, M. Gillespie, K. Hausmann, R. Haw, B. Jassal, S. Jupe, F. Korninger, et al. (2016). "The Reactome pathway Knowledgebase." *Nucleic Acids Research* 44.D1, pp. 481–487.

Fish, J. E., M. C. Gutierrez, L. T. Dang, N. Khyzha, Z. Chen, S. Veitch, H. S. Cheng, M. Khor, L. Antounians, et al. (2017). "Dynamic regulation of VEGF-inducible genes by an ERK-ERG-p300 transcriptional network". *Development* 144.13, pp. 2428–2444.

FitzGerald, P. C., A. Shlyakhtenko, A. A. Mir, and C. Vinson (2004). "Clustering of DNA sequences in human promoters." *Genome Research* 14.8, pp. 1562–1574.

Fukasawa, R., S. Iida, T. Tsutsui, Y. Hirose, and Y. Ohkuma (2015). "Mediator complex cooperatively regulates transcription of retinoic acid target genes with Polycomb Repressive Complex 2 during neuronal differentiation". *Journal of Biochemistry* 158.5, pp. 373–384.

Furuta, C., H. Ema, S.-I. Takayanagi, T. Ogaeri, D. Okamura, Y. Matsui, and H. Nakauchi (2006). "Discordant developmental waves of angioblasts and hemangioblasts in the early gastrulating mouse embryo." *Development* 133.14, pp. 2771–2779.

Gerhardt, H., M. Golding, M. Fruttiger, C. Ruhrberg, A. Lundkvist, A. Abramsson, M. Jeltsch, C. Mitchell, K. Alitalo, et al. (2003). "VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia." *The Journal of Cell Biology* 161.6, pp. 1163–1177.

Gerstein, M. B., A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, et al. (2012). "Architecture of the human regulatory network derived from ENCODE data". *Nature* 489.7414, pp. 91–100.

Giannotta, M., M. Trani, and E. Dejana (2013). "VE-cadherin and endothelial adherens junctions: active guardians of vascular integrity." *Developmental Cell* 26.5, pp. 441–454.

Gilbert, L. A., M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, et al. (2014). "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation". *Cell* 159.3, pp. 647–661.

Gilbert, L. A., M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, et al. (2013). "CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes". *Cell* 154.2, pp. 442–451.

Gillette, T. G. and J. A. Hill (2015). "Readers, writers, and erasers: chromatin as the whiteboard of heart disease." *Circulation Research* 116.7, pp. 1245–1253.

Ginsberg, M., D. James, B.-S. Ding, D. Nolan, F. Geng, J. Butler, W. Schachterle, V. Pulijaal, S. Mathew, et al. (2012). "Efficient Direct Reprogramming of Mature Amniotic Cells into Endothelial Cells by ETS Factors and TGFβ Suppression". *Cell* 151.3, pp. 559–575.

Goodman, R. H. and S. Smolik (2000). "CBP/p300 in cell growth, transformation, and development." *Genes & Development* 14.13, pp. 1553–1577.

Gorski, J. J., K. I. Savage, J. M. Mulligan, S. S. McDade, J. K. Blayney, Z. Ge, and D. P. Harkin (2011). "Profiling of the BRCA1 transcriptome through microarray and ChIP-chip analysis". *Nucleic Acids Research* 39.22, pp. 9536–9548.

Groner, A. C., S. Meylan, A. Ciuffi, N. Zangger, G. Ambrosini, N. Dénervaud, P. Bucher, and D. Trono (2010). "KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading". *PLoS Genetics* 6.3. Ed. by H. D. Madhani, Article e1000869.

Grossman, S. R., X. Zhang, L. Wang, J. Engreitz, A. Melnikov, P. Rogov, R. Tewhey, A. Isakova, B. Deplancke, et al. (2017). "Systematic dissection of genomic features determining transcription factor binding and enhancer function". *Proceedings of the National Academy of Sciences* 114.7, pp. 1291–1300.

Grossman, S. R., J. Engreitz, J. P. Ray, T. H. Nguyen, N. Hacohen, and E. S. Lander (2018). "Positional specificity of different transcription factor classes within enhancers." *Proceedings of the National Academy of Sciences* 115.30, pp. 7222–7230.

GTEx Consortium, T. G. (2013). "The Genotype-Tissue Expression (GTEx) project." *Nature Genetics* 45.6, pp. 580–585.

Gupta, R. M., J. Hadaya, A. Trehan, S. M. Zekavat, C. Roselli, D. Klarin, C. A. Emdin, C. R. Hilvering, V. Bianchi, et al. (2017). "A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression". *Cell* 170.3, pp. 522–533.

Hägglöf, C., P. Hammarsten, K. Strömvall, L. Egevad, A. Josefsson, P. Stattin, T. Granfors, and A. Bergh (2014). "TMPRSS2-ERG expression predicts prostate cancer survival and associates with stromal biomarkers." *PLoS ONE* 9.2, Article e86824.

Hay, D., J. R. Hughes, C. Babbs, J. O. J. Davies, B. J. Graham, L. L. P. Hanssen, M. T. Kassouf, A. M. Oudelaar, J. A. Sharpe, et al. (2016). "Genetic dissection of the α-globin super-enhancer in vivo". *Nature Genetics* 48.8, pp. 895–903.

He, A., S. W. Kong, Q. Ma, and W. T. Pu (2011). "Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart". *Proceedings of the National Academy of Sciences* 108.14, pp. 5632–5637.

Heidemann, J., H. Ogawa, M. B. Dwinell, P. Rafiee, C. Maaser, H. R. Gockel, M. F. Otterson, D. M. Ota, N. Lügering, et al. (2003). "Angiogenic Effects of Interleukin 8 (CXCL8) in Human Intestinal Microvascular Endothelial Cells Are Mediated by CXCR2". *Journal of Biological Chemistry* 278.10, pp. 8508–8515.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome". *Nature Genetics* 39.3, pp. 311–318.

Heinz, S., C. E. Romanoski, C. Benner, K. A. Allison, M. U. Kaikkonen, L. D. Orozco, and C. K. Glass (2013). "Effect of natural genetic variation on enhancer selection and function". *Nature* 503.7477, pp. 487–492.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." *Molecular Cell* 38.4, pp. 576–589.

Heinz, S., C. E. Romanoski, C. Benner, and C. K. Glass (2015). "The selection and function of cell type-specific enhancers." *Nature reviews. Molecular Cell Biology* 16.3, pp. 144–154.

Henriques, T., B. S. Scruggs, M. O. Inouye, G. W. Muse, L. H. Williams, A. B. Burkholder, C. A. Lavender, D. C. Fargo, and K. Adelman (2018). "Widespread transcriptional pausing and elongation control at enhancers." *Genes & Development* 32.1, pp. 26–41.

Hirase, T., J. M. Staddon, M. Saitou, Y. Ando-Akatsuka, M. Itoh, M. Furuse, K. Fujimoto, S. Tsukita, and L. L. Rubin (1997). "Occludin as a possible determinant of tight junction permeability in endothelial cells." *Journal of Cell Science* 110.1, pp. 1603–1613.

Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, R. A. Young, N. Ahmadiyeh, et al. (2013). "Super-Enhancers in the Control of Cell Identity and Disease". *Cell* 155.4, pp. 934–947.

Ho, L. and G. R. Crabtree (2010). "Chromatin remodelling during development". *Nature* 463.7280, Article 474.

Hogan, N. T., M. B. Whalen, L. K. Stolze, N. K. Hadeli, M. T. Lam, J. R. Springstead, C. K. Glass, and C. E. Romanoski (2017). "Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells". *eLife* 6, pp. 1–28.

Hollenhorst, P. C., L. P. McIntosh, and B. J. Graves (2011). "Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors". *Annual Review of Biochemistry* 80.1, pp. 437–471.

Huang, J., K. Li, W. Cai, X. Liu, Y. Zhang, S. H. Orkin, J. Xu, and G.-C. Yuan (2018). "Dissecting super-enhancer hierarchy based on chromatin interactions". *Nature Communications* 9.1, Article 943.

Huang, J., X. Liu, D. Li, Z. Shao, H. Cao, Y. Zhang, E. Trompouki, T. V. Bowman, L. I. Zon, et al. (2016). "Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis." *Developmental C45ell* 36.1, pp. 9–23.

Huang, Y., W. Li, X. Yao, Q.-j. Lin, J.-w. Yin, Y. Liang, M. Heiner, B. Tian, J. Hui, and G. Wang (2012). "Mediator Complex Regulates Alternative mRNA Processing via the MED23 Subunit". *Molecular Cell* 45.4, pp. 459–469.

Huang, Z.-Q., J. Li, L. M. Sachs, P. A. Cole, and J. Wong (2003). "A role for cofactor-cofactor and cofactor-histone interactions in targeting p300, SWI/SNF and Mediator for transcription". *The EMBO Journal* 22.9, pp. 2146–2155.

Hull, R. P., P. K. Srivastava, Z. D'Souza, S. S. Atanur, F. Mechta-Grigoriou, L. Game, E. Petretto, H. T. Cook, T. J. Aitman, and J. Behmoaras (2013). "Combined ChIP-Seq and transcriptome analysis identifies AP-1/JunD as a primary regulator of oxidative stress and IL-1β synthesis in macrophages." *BMC Genomics* 14, pp. 92–108.

Iglesias-Gato, D., Y.-C. Chuan, N. Jiang, C. Svensson, J. Bao, I. Paul, L. Egevad, B. M. Kessler, P. Wikstrom, et al. (2015). "OTUB1 de-ubiquitinating enzyme promotes prostate cancer cell invasion in vitro and tumorigenesis in vivo". *Molecular Cancer* 14.1, pp. 8–22.

Iwafuchi-Doi, M. and K. S. Zaret (2014). "Pioneer transcription factors in cell reprogramming". *Genes & Development* 28.24, pp. 2679–2692.

Iyengar, S. and P. J. Farnham (2011). "KAP1 protein: an enigmatic master regulator of the genome." *The Journal of Biological Chemistry* 286.30, pp. 26267–26276.

Jang, M. K., K. Mochizuki, M. Zhou, H.-S. Jeong, J. N. Brady, and K. Ozato (2005). "The Bromodomain Protein Brd4 Is a Positive Regulatory Component of P-TEFb and Stimulates RNA Polymerase II-Dependent Transcription". *Molecular Cell* 19.4, pp. 523–534.

Javierre, B. M., O. S. Burren, S. P. Wilder, R. Kreuzhuber, S. M. Hill, S. Sewitz, J. Cairns, S. W. Wingett, C. Várnai, et al. (2016). "Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters." *Cell* 167.5, pp. 1369–1384.

Jayaraman, G., R. Srinivas, C. Duggan, E. Ferreira, S. Swaminathan, K. Somasundaram, J. Williams, C. Hauser, M. Kurkinen, et al. (1999). "p300/cAMP-responsive element-binding protein interactions with ets-1 and ets-2 in the transcriptional activation of the human stromelysin promoter." *The Journal of Biological Chemistry* 274.24, pp. 17342–17352.

Jensen, K. T., L. Fløe, T. S. Petersen, J. Huang, F. Xu, L. Bolund, Y. Luo, and L. Lin (2017). "Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency". *FEBS Letters* 591.13, pp. 1892–1901.

Jones, G. T., G. Tromp, H. Kuivaniemi, S. Gretarsdottir, A. F. Baas, B. Giusti, E. Strauss, F. N. G. Van't Hof, T. R. Webb, et al. (2017). "Meta-Analysis of Genome-Wide Association Studies for Abdominal Aortic Aneurysm Identifies Four New Disease-Specific Risk Loci." *Circulation Research* 120.2, pp. 341–353.

Kadoch, C., R. T. Williams, J. P. Calarco, E. L. Miller, C. M. Weber, S. M. G. Braun, J. L. Pulice, E. J. Chory, and G. R. Crabtree (2017). "Dynamics of BAF–Polycomb complex opposition on heterochromatin in normal and oncogenic states". *Nature Genetics* 49.2, pp. 213–222.

Kagey, M. H., J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture". *Nature* 467.7314, pp. 430–435.

Kaikkonen, M. U., H. Niskanen, C. E. Romanoski, E. Kansanen, A. M. Kivela, J. Laitalainen, S. Heinz, C. Benner, C. K. Glass, and S. Yla-Herttuala (2014). "Control of VEGF-A transcriptional programs by pausing and genomic compartmentalization". *Nucleic Acids Research* 42.20, pp. 12570–12584.

Kaikkonen, M. U., N. J. Spann, S. Heinz, C. E. Romanoski, K. A. Allison, J. D. Stender, H. B. Chun, D. F. Tough, R. K. Prinjha, et al. (2013). "Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription". *Molecular Cell* 51.3, pp. 310–325.

Kalna, V., Y. Yang, C. R. Peghaire, K. Frudd, R. Hannah, A. V. Shah, L. Osuna Almagro, J. J. Boyle, B. Göttgens, et al. (2019). "The Transcription Factor ERG Regulates Super-Enhancers Associated With an Endothelial-Specific Gene Expression Program". *Circulation Research* 124.9, pp. 1337–1349.

Kanda, S., Y. Miyata, Y. Mochizuki, T. Matsuyama, and H. Kanetake (2005). "Angiopoietin 1 Is Mitogenic for Cultured Endothelial Cells". *Cancer Research* 65.15, pp. 6820–6827.

Kanki, Y., T. Kohro, S. Jiang, S. Tsutsumi, I. Mimura, J.-i. Suehiro, Y. Wada, Y. Ohta, S. Ihara, et al. (2011). "Epigenetically coordinated GATA2 binding is necessary for endothelium-specific *endomucin* expression". *The EMBO Journal* 30.13, pp. 2582–2595.

Kanki, Y., R. Nakaki, T. Shimamura, T. Matsunaga, K. Yamamizu, S. Katayama, J.-i. Suehiro, T. Osawa, H. Aburatani, et al. (2017). "Dynamically and epigenetically coordinated GATA/ETS/SOX transcription factor expression is indispensable for endothelial cell differentiation". *Nucleic Acids Research* 45.8, pp. 4344–4358.

Karampini, E., M. Schillemans, M. Hofman, F. van Alphen, M. de Boer, T. W. Kuijpers, M. van den Biggelaar, J. Voorberg, and R. Bierings (2019). "Defective AP-3-dependent VAMP8 trafficking impairs Weibel-Palade body exocytosis in Hermansky-Pudlak Syndrome type 2 blood outgrowth endothelial cells." *Haematologica* 104.10, pp. 2091–2099.

Kedage, V., B. G. Strittmatter, P. B. Dausinas, and P. C. Hollenhorst (2017). "Phosphorylation of the oncogenic transcription factor ERG in prostate cells dissociates polycomb repressive complex 2, allowing target gene activation". *Journal of Biological Chemistry* 292.42, pp. 17225–17235.

Khan, A., A. Mathelier, and X. Zhang (2018). "Super-enhancers are transcriptionally more active and cell type-specific than stretch enhancers". *Epigenetics* 13.9, pp. 910–922.

Kim, B.-K., J.-Y. Im, G. Han, W.-J. Lee, K.-J. Won, K.-S. Chung, K. Lee, H. S. Ban, K. Song, and M. Won (2014). "p300 cooperates with c-Jun and PARP-1 at the p300 binding site to activate RhoB transcription in NSC126188-mediated apoptosis". *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839.5, pp. 364–373.

Kim, Y. J., K. R. Cecchini, and T. H. Kim (2011). "Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus." *Proceedings of the National Academy of Sciences* 108.18, pp. 7391–7396.

Korkmaz, G., R. Lopes, A. P. Ugalde, E. Nevedomskaya, R. Han, K. Myacheva, W. Zwart, R. Elkon, and R. Agami (2016). "Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9". *Nature Biotechnology* 34.2, pp. 192–198.

Krause, M. D., R.-T. Huang, D. Wu, T.-P. Shentu, D. L. Harrison, M. B. Whalen, L. K. Stolze, A. Di Rienzo, I. P. Moskowitz, et al. (2018). "Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics". *Proceedings of the National Academy of Sciences* 115.48, pp. 11349–11358.

Krock, B. L., N. Skuli, and M. C. Simon (2011). "Hypoxia-induced angiogenesis: good and evil." *Genes & cancer* 2.12, pp. 1117–1133.

Kron, K. J., A. Murison, S. Zhou, V. Huang, T. N. Yamaguchi, Y.-J. Shiah, M. Fraser, T. van der Kwast, P. C. Boutros, et al. (2017). "TMPRSS2–ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer". *Nature Genetics* 49.9, pp. 1336–1345.

Kruse, E. A., S. J. Loughran, T. M. Baldwin, E. C. Josefsson, S. Ellis, D. K. Watson, P. Nurden, D. Metcalf, D. J. Hilton, et al. (2009). "Dual requirement for the ETS transcription factors Fli-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage". *Proceedings of the National Academy of Sciences* 106.33, pp. 13814–13819.

Kume, T., H. Jiang, J. M. Topczewska, and B. L. Hogan (2001). "The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis." *Genes & Development* 15.18, pp. 2470–2482.

Kwak, H. J., J. N. So, S. J. Lee, I. Kim, and G. Y. Koh (1999). "Angiopoietin-1 is an apoptosis survival factor for endothelial cells." *FEBS Letters* 448.2-3, pp. 249–253.

Lalonde, S., V.-A. Codina-Fauteux, S. M. de Bellefon, F. Leblanc, M. Beaudoin, M.-M. Simon, R. Dali, T. Kwan, K. S. Lo, et al. (2019). "Integrative analysis of vascular endothelial cell genomic features identifies AIDA as a coronary artery disease candidate gene". *Genome Biology* 20.1, Article 133.

Lang, I., M. A. Pabst, U. Hiden, A. Blaschitz, G. Dohr, T. Hahn, and G. Desoye (2003). "Heterogeneity of microvascular endothelial cells isolated from human term placenta and macrovascular umbilical vein endothelial cells". *European Journal of Cell Biology* 82.4, pp. 163–173.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2". *Nature Methods* 9.4, pp. 357–359.

Lara-Astiaso, D., A. Weiner, E. Lorenzo-Vivas, I. Zaretsky, D. A. Jaitin, E. David, H. Keren-Shaul, A. Mildner, D. Winter, et al. (2014). "Chromatin state dynamics during blood formation". *Science* 345.6199, pp. 943–949.

Lathen, C., Y. Zhang, J. Chow, M. Singh, G. Lin, V. Nigam, Y. A. Ashraf, J. X. Yuan, I. M. Robbins, and P. A. Thistlethwaite (2014). "ERG-APLNR axis controls pulmonary venule endothelial proliferation in pulmonary veno-occlusive disease." *Circulation* 130.14, pp. 1179–1191.

Le Bras, A., C. Samson, M. Trentini, B. Caetano, E. Lelievre, V. Mattot, F. Beermann, and F. Soncin (2010). "VE-statin/egfl7 Expression in Endothelial Cells Is Regulated by a Distal Enhancer and a Proximal Promoter under the Direct Control of Erg and GATA-2". *PLoS ONE* 5.8, Article e12156.

Lee, D., C. Park, H. Lee, J. J. Lugus, S. H. Kim, E. Arentson, Y. S. Chung, G. Gomez, M. Kyba, et al. (2008). "ER71 Acts Downstream of BMP, Notch, and Wnt Signaling in Blood and Vessel Progenitor Specification". *Cell Stem Cell* 2.5, pp. 497–507.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup (2009). "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* 25.16, pp. 2078–2079.

Liao, Y., G. K. Smyth, and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". *Bioinformatics* 30.7, pp. 923–930.

Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdottir, P. Tamayo, and J. P. Mesirov (2011). "Molecular signatures database (MSigDB) 3.0". *Bioinformatics* 27.12, pp. 1739–1740.

Lim, K.-C., T. Hosoya, W. Brandt, C.-J. Ku, S. Hosoya-Ohmura, S. A. Camper, M. Yamamoto, and J. D. Engel (2012). "Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning". *Journal of Clinical Investigation* 122.10, pp. 3705–3717.

Lizio, M., J. Harshbarger, H. Shimoji, J. Severin, T. Kasukawa, S. Sahin, I. Abugessaisa, S. Fukuda, F. Hori, et al. (2015). "Gateways to the FANTOM5 promoter level mammalian expression atlas". *Genome Biology* 16.1, Article 22.

Long, H. K., S. L. Prescott, and J. Wysocka (2016). "Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution". *Cell* 167.5, pp. 1170–1187.

Looney, A. P., R. Han, L. Stawski, G. Marden, M. Iwamoto, and M. Trojanowska (2017). "Synergistic Role of Endothelial ERG and FLI1 in Mediating Pulmonary Vascular Homeostasis." *American Journal of Respiratory Cell and Molecular Biology* 57.1, pp. 121–131.

Lopes, R., R. Agami, and G. Korkmaz (2017). "GRO-seq, A Tool for Identification of Transcripts Regulating Gene Expression." *Methods in Molecular Biology* 1543, pp. 45–55.

Lother, A., S. Bergemann, L. Deng, M. Moser, C. Bode, and L. Hein (2018). "Cardiac Endothelial Cell Transcriptome." *Arteriosclerosis, Thrombosis, and Vascular Biology* 38.3, pp. 566–574.

Love, M. I., W. Huber, and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biology* 15.12, Article 550.

Lovén, J., H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, R. A. Young, et al. (2013). "Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers". *Cell* 153.2, pp. 320–334.

Lu, L., N. Fu, X. Luo, X. Li, and X. Li (2015). "Overexpression of cofilin 1 in prostate cancer and the corresponding clinical implications". *Oncology Letters* 9.6, pp. 2757–2761.

Lupien, M., J. Eeckhoute, C. A. Meyer, Q. Wang, Y. Zhang, W. Li, J. S. Carroll, X. S. Liu, and M. Brown (2008). "FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription." *Cell* 132.6, pp. 958–970.

MacArthur, J., E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon, A. Milano, et al. (2017). "The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)." *Nucleic Acids Research* 45, pp. 896–901.

Mackereth, C. D., M. Schärpf, L. N. Gentile, S. E. MacIntosh, C. M. Slupsky, and L. P. McIntosh (2004). "Diversity in Structure and Function of the Ets Family PNT Domains". *Journal of Molecular Biology* 342.4, pp. 1249–1264.

Madani Tonekaboni, S. A., P. Mazrooei, V. Kofia, B. Haibe-Kains, and M. Lupien (2019). "Identifying clusters of *cis*-regulatory elements underpinning TAD structures and lineage-specific regulatory networks". *Genome Research* 29.10, pp. 1733–1743.

Madison, B. J., K. A. Clark, N. Bhachech, P. C. Hollenhorst, B. J. Graves, and S. L. Currie (2018). "Electrostatic repulsion causes anticooperative DNA binding between tumor suppressor ETS transcription factors and JUN−FOS at composite DNA sites". *The Journal of Biological Chemistry* 293.48, pp. 18624–18635.

Mailman, M. D., M. Feolo, Y. Jin, M. Kimura, K. Tryka, R. Bagoutdinov, L. Hao, A. Kiang, J. Paschall, et al. (2007). "The NCBI dbGaP database of genotypes and phenotypes". *Nature Genetics* 39.10, pp. 1181–1186.

Mangeot, P. E., V. Risson, F. Fusil, A. Marnef, E. Laurent, J. Blin, V. Mournetas, E. Massouridès, T. J. M. Sohier, et al. (2019). "Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins". *Nature Communications* 10.1, p. 45.

Mani, R.-S., M. K. Iyer, Q. Cao, J. C. Brenner, L. Wang, A. Ghosh, X. Cao, R. J. Lonigro, S. A. Tomlins, et al. (2011). "TMPRSS2-ERG-Mediated Feed-Forward Regulation of Wild-Type ERG in Human Prostate Cancers". *Cancer Research* 71.16, pp. 5387–5392.

Mani, R.-S., S. A. Tomlins, K. Callahan, A. Ghosh, M. K. Nyati, S. Varambally, N. Palanisamy, and A. M. Chinnaiyan (2009). "Induced Chromosomal Proximity and Gene Fusions in Prostate Cancer". *Science* 326.5957, pp. 1230–1230.

Mansour, M. R., B. J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A. D. Durbin, J. Etchin, L. Lawton, S. E. Sallan, et al. (2014). "An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element". *Science* 346.6215, pp. 1373–1377.

Marcu, R., Y. J. Choi, J. Xue, C. L. Fortin, Y. Wang, R. J. Nagao, J. Xu, J. W. MacDonald, T. K. Bammler, et al. (2018). "Human Organ-Specific Endothelial Cell Heterogeneity". *iScience* 4, pp. 20–35.

Matheny, H. E., T. L. Deem, and J. M. Cook-Mills (2000). "Lymphocyte Migration Through Monolayers of Endothelial Cell Lines Involves VCAM-1 Signaling Via Endothelial Cell NADPH Oxidase". *The Journal of Immunology* 164.12, pp. 6550–6559.

Mayran, A., K. Sochodolsky, K. Khetchoumian, J. Harris, Y. Gauthier, A. Bemmo, A. Balsalobre, and J. Drouin (2019). "Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening". *Nature Communications* 10.1, Article 3807.

McLaughlin, F., V. J. Ludbrook, J. Cox, I. von Carlowitz, S. Brown, and A. M. Randi (2001). "Combined genomic and antisense analysis reveals that the transcription factor Erg is implicated in endothelial cell differentiation." *Blood* 98.12, pp. 3332–3339.

McLaughlin, F., V. J. Ludbrook, I. Kola, C. J. Campbell, and A. M. Randi (1999). "Characterisation of the tumour necrosis factor (TNF)-(alpha) response elements in the human ICAM-2 promoter." *Journal of Cell Science* 112, pp. 4695–703.

McLean, C. Y., D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, and G. Bejerano (2010). "GREAT improves functional interpretation of cis-regulatory regions". *Nature Biotechnology* 28.5, pp. 495–501.

Meadows, S. M., C. T. Myers, and P. A. Krieg (2011). "Regulation of endothelial cell development by ETS transcription factors". *Seminars in Cell & Developmental Biology* 22.9, Article 976.

Meadows, S. M., M. C. Salanga, and P. A. Krieg (2009). "Kruppel-like factor 2 cooperates with the ETS family protein ERG to activate Flk1 expression during vascular development." *Development* 136.7, pp. 1115–1125.

Medina, R. J., C. L. Barber, F. Sabatier, F. Dignat-George, J. M. Melero-Martin, K. Khosrotehrani, O. Ohneda, A. M. Randi, J. K. Chan, et al. (2017). "Endothelial Progenitors: A Consensus Statement on Nomenclature". *STEM CELLS Translational Medicine* 6.5, pp. 1316–1320.

Mifsud, B., F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder, L. Ferreira, S. W. Wingett, S. Andrews, W. Grey, et al. (2015). "Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C". *Nature Genetics* 47.6, pp. 598–606.

Miguel-Escalada, I., S. Bonàs-Guarch, I. Cebola, J. Ponsa-Cobas, J. Mendieta-Esteban, G. Atla, B. M. Javierre, D. M. Y. Rolando, I. Farabella, et al. (2019). "Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes". *Nature Genetics* 51.7, pp. 1137–1148.

Miraldi, E. R., M. Pokrovskii, A. Watters, D. M. Castro, N. De Veaux, J. A. Hall, J.-Y. Lee, M. Ciofani, A. Madar, et al. (2019). "Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells." *Genome Research* 29.3, pp. 449–463.

Mohamed, A. A., S.-H. Tan, N. Mikhalkevich, S. Ponniah, V. Vasioukhin, C. J. Bieberich, I. A. Sesterhenn, A. Dobi, S. Srivastava, and T. L. Sreenath (2010). "Ets Family

Protein, Erg Expression in Developing and Adult Mouse Tissues by a Highly Specific Monoclonal Antibody". *Journal of Cancer* 1, pp. 197–208.

Moorthy, S. D., S. Davidson, V. M. Shchuka, G. Singh, N. Malek-Gilani, L. Langroudi, A. Martchenko, V. So, N. N. Macpherson, and J. A. Mitchell (2017). "Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes." *Genome Research* 27.2, pp. 246–258.

Morgunova, E. and J. Taipale (2017). "Structural perspective of cooperative transcription factor binding". *Current Opinion in Structural Biology* 47, pp. 1–8.

Morita, R., M. Suzuki, H. Kasahara, N. Shimizu, T. Shichita, T. Sekiya, A. Kimura, K.-i. Sasaki, H. Yasukawa, and A. Yoshimura (2015). "ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells". *Proceedings of the National Academy of Sciences* 112.1, pp. 160–165.

Mounir, Z., J. M. Korn, T. Westerling, F. Lin, C. A. Kirby, M. Schirle, G. McAllister, G. Hoffman, N. Ramadan, et al. (2016). "ERG signaling in prostate cancer is driven through PRMT5-dependent methylation of the Androgen Receptor". *eLife* 5, Article e13964.

Muerdter, F., Ł. M. Boryń, and C. D. Arnold (2015). "STARR-seq - principles and applications." *Genomics* 106.3, pp. 145–150.

Mumbach, M. R., A. T. Satpathy, E. A. Boyle, C. Dai, B. G. Gowen, S. W. Cho, M. L. Nguyen, A. J. Rubin, J. M. Granja, et al. (2017). "Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements". *Nature Genetics* 49.11, pp. 1602–1612.

Musunuru, K., A. Strong, M. Frank-Kamenetsky, N. E. Lee, T. Ahfeldt, K. V. Sachs, X. Li, H. Li, N. Kuperwasser, et al. (2010). "From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus". *Nature* 466.7307, pp. 714–719.

Nagai, N., H. Ohguchi, R. Nakaki, Y. Matsumura, Y. Kanki, J. Sakai, H. Aburatani, and T. Minami (2018). "Downregulation of ERG and FLI1 expression in endothelial cells triggers endothelial-to-mesenchymal transition". *PLOS Genetics* 14.11. Ed. by E. M. Zeisberg, Article e1007826.

Ng, A. P., S. J. Loughran, D. Metcalf, C. D. Hyland, C. A. de Graaf, Y. Hu, G. K. Smyth, D. J. Hilton, B. T. Kile, and W. S. Alexander (2011). "Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice". *Blood* 118.9, pp. 2454–2461.

Nikolova-Krstevski, V., L. Yuan, A. Le Bras, P. Vijayaraj, M. Kondo, I. Gebauer, M. Bhasin, C. V. Carman, and P. Oettgen (2009). "ERG is required for the differentiation of embryonic stem cells along the endothelial lineage." *BMC Developmental Biology* 9.1, Article 72.

Niskanen, H., I. Tuszynska, R. Zaborowski, M. Heinäniemi, S. Ylä-Herttuala, B. Wilczynski, and M. U. Kaikkonen (2018). "Endothelial cell differentiation is encompassed by changes in long range interactions between inactive chromatin regions." *Nucleic Acids Research* 46.4, pp. 1724–1740.

Nolan, D. J., M. Ginsberg, E. Israely, B. Palikuqi, M. G. Poulos, D. James, B.-S. Ding, W. Schachterle, Y. Liu, et al. (2013). "Molecular Signatures of Tissue-Specific Mi-

crovascular Endothelial Cell Heterogeneity in Organ Maintenance and Regeneration". *Developmental Cell* 26.2, pp. 204–219.

Nolis, I. K., D. J. McKay, E. Mantouvalou, S. Lomvardas, M. Merika, and D. Thanos (2009). "Transcription factors mediate long-range enhancer-promoter interactions". *Proceedings of the National Academy of Sciences* 106.48, pp. 20222–20227.

Northcott, P. A., C. Lee, T. Zichner, A. M. Stütz, S. Erkek, D. Kawauchi, D. J. H. Shih, V. Hovestadt, M. Zapatka, et al. (2014). "Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma". *Nature* 511.7510, pp. 428–434.

Okita, K., T. Ichisaka, and S. Yamanaka (2007). "Generation of germline-competent induced pluripotent stem cells". *Nature* 448.7151, pp. 313–317.

Papanicolaou, K. N., Y. Izumiya, and K. Walsh (2008). "Forkhead transcription factors and cardiovascular biology." *Circulation Research* 102.1, pp. 16–31.

Papapetropoulos, A., D. Fulton, K. Mahboubi, R. G. Kalb, D. S. O'Connor, F. Li, D. C. Altieri, and W. C. Sessa (2000). "Angiopoietin-1 Inhibits Endothelial Cell Apoptosis via the Akt/Survivin Pathway". *Journal of Biological Chemistry* 275.13, pp. 9102–9105.

Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." *Nature Reviews Genetics* 10.10, pp. 669–680.

Parker, S. C. J., M. L. Stitzel, D. L. Taylor, J. M. Orozco, M. R. Erdos, J. A. Akiyama, K. L. van Bueren, P. S. Chines, N. Narisu, et al. (2013). "Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants." *Proceedings of the National Academy of Sciences* 110.44, pp. 17921–17926.

Paschalaki, K. E. and A. M. Randi (2018). "Recent Advances in Endothelial Colony Forming Cells Toward Their Use in Clinical Translation". *Frontiers in Medicine* 5, Article 295.

Pasquali, L., K. J. Gaulton, S. A. Rodríguez-Seguí, L. Mularoni, I. Miguel-Escalada, I. Akerman, J. J. Tena, I. Morán, C. Gómez-Marín, et al. (2014). "Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants." *Nature Genetics* 46.2, pp. 136–143.

Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford (2017). "Salmon provides fast and bias-aware quantification of transcript expression". *Nature Methods* 14.4, pp. 417–419.

Patsch, C., L. Challet-Meylan, E. C. Thoma, E. Urich, T. Heckel, J. F. O'Sullivan, S. J. Grainger, F. G. Kapp, L. Sun, et al. (2015). "Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells." *Nature Cell Biology* 17.8, pp. 994–1003.

Pedrosa, A.-R., A. Trindade, A.-C. Fernandes, C. Carvalho, J. Gigante, A. T. Tavares, R. Diéguez-Hurtado, H. Yagita, R. H. Adams, and A. Duarte (2015). "Endothelial Jagged1 Antagonizes Dll4 Regulation of Endothelial Branching and Promotes Vascular Maturation Downstream of Dll4/Notch1". *Arteriosclerosis, Thrombosis, and Vascular Biology* 35.5, pp. 1134–1146.

Peghaire, C., N. P. Dufton, M. Lang, I. I. Salles-Crawley, J. Ahnström, V. Kalna, C. Raimondi, C. Pericleous, L. Inuabasi, et al. (2019). "The transcription factor ERG

regulates a low shear stress-induced anti-thrombotic pathway in the microvasculature". *Nature Communications* 10.1, Article 5014.

Pigazzi, M., R. Masetti, F. Martinolli, E. Manara, A. Beghin, R. Rondelli, F. Locatelli, F. Fagioli, A. Pession, and G. Basso (2012). "Presence of high-ERG expression is an independent unfavorable prognostic marker in MLL-rearranged childhood myeloid leukemia". *Blood* 119.4, pp. 1086–1087.

Pimanda, J. E., W. Y. I. Chan, I. J. Donaldson, M. Bowen, A. R. Green, and B. Göttgens (2006). "Endoglin expression in the endothelium is regulated by Fli-1, Erg, and Elf-1 acting on the promoter and a -8-kb enhancer." *Blood* 107.12, pp. 4737–4745.

Pinz, S., S. Unser, and A. Rascle (2016). "Signal transducer and activator of transcription STAT5 is recruited to c-Myc super-enhancer". *BMC Molecular Biology* 17.1, Article 10.

Potente, M., C. Urbich, K.-i. Sasaki, W. K. Hofmann, C. Heeschen, A. Aicher, R. Kollipara, R. A. DePinho, A. M. Zeiher, and S. Dimmeler (2005). "Involvement of Foxo transcription factors in angiogenesis and postnatal neovascularization". *Journal of Clinical Investigation* 115.9, pp. 2382–2392.

Pruitt, K. D., T. Tatusova, and D. R. Maglott (2007). "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." *Nucleic Acids Research* 35, pp. 61–65.

Qi, L. S., M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim (2013). "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell* 152.5, pp. 1173–1183.

Qian, X., C. Li, B. Pang, M. Xue, J. Wang, and J. Zhou (2012). "Spondin-2 (SPON2), a More Prostate-Cancer-Specific Diagnostic Biomarker". *PLoS ONE* 7.5. Ed. by C. L. Addison, Article e37225.

Quevedo, M., L. Meert, M. R. Dekker, D. H. W. Dekkers, J. H. Brandsma, D. L. C. van den Berg, Z. Ozgür, W. F. J. van IJcken, J. Demmers, et al. (2019). "Mediator complex interaction partners organize the transcriptional network that defines neural stem cells". *Nature Communications* 10.1, Article 2669.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* 26.6, pp. 841–842.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn, and J. Wysocka (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." *Nature* 470.7333, pp. 279–283.

Rafii, S., J. M. Butler, and B.-S. Ding (2016). "Angiocrine functions of organ-specific endothelial cells." *Nature* 529.7586, pp. 316–325.

Rajendran, P., T. Rengarajan, J. Thangavel, Y. Nishigaki, D. Sakthisekaran, G. Sethi, and I. Nishigaki (2013). "The vascular endothelium and human diseases." *International Journal of Biological Sciences* 9.10, pp. 1057–1069.

Rambout, X., C. Detiffe, J. Bruyr, E. Mariavelle, M. Cherkaoui, S. Broh?e, P. Demoiti?, M. Lebrun, R. Soin, et al. (2016). "The transcription factor ERG recruits CCR4?NOT to control mRNA decay and mitotic progression". *Nature Structural & Molecular Biology* 23.7, pp. 663–672.

Ramírez, F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, and T. Manke (2016). "deepTools2: A Next Generation Web Server for Deep-Sequencing Data Analysis". *Nucleic Acids Research* 44, pp. 160–165.

Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, et al. (2014). "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping". *Cell* 159.7, pp. 1665–1680.

Regan, M. C., P. S. Horanyi, E. E. Pryor, J. L. Sarver, D. S. Cafiso, and J. H. Bushweller (2013). "Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited". *Proceedings of the National Academy of Sciences* 110.33, pp. 13374–13379.

Rickman, D. S., T. D. Soong, B. Moss, J. M. Mosquera, J. Dlabal, S. Terry, T. Y. MacDonald, J. Tripodi, K. Bunting, et al. (2012). "Oncogene-mediated alterations in chromatin conformation". *Proceedings of the National Academy of Sciences* 109.23, pp. 9083–9088.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies". *Nucleic Acids Research* 43, pp. 1–25.

Sabari, B. R., A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, et al. (2018). "Coactivator condensation at super-enhancers links phase separation and gene control." *Science* 361.6400, Article eaar3958.

Sabbagh, M. F., J. S. Heng, C. Luo, R. G. Castanon, J. R. Nery, A. Rattner, L. A. Goff, J. R. Ecker, and J. Nathans (2018). "Transcriptional and epigenomic landscapes of CNS and non-CNS vascular endothelial cells". *eLife* 7, Article e36187.

Sahu, B., M. Laakso, K. Ovaska, T. Mirtti, J. Lundin, A. Rannikko, A. Sankila, J.-P. Turunen, M. Lundin, et al. (2011). "Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer". *The EMBO Journal* 30.19, pp. 3962–3976.

Saint-André, V., A. J. Federation, C. Y. Lin, B. J. Abraham, J. Reddy, T. I. Lee, J. E. Bradner, and R. A. Young (2016). "Models of human core transcriptional regulatory circuitries." *Genome Research* 26.3, pp. 385–396.

Sakamoto, Y., K. Hara, M. Kanai-Azuma, T. Matsui, Y. Miura, N. Tsunekawa, M. Kurohmaru, Y. Saijoh, P. Koopman, and Y. Kanai (2007). "Redundant roles of Sox17 and Sox18 in early cardiovascular development of mouse embryos". *Biochemical and Biophysical Research Communications* 360.3, pp. 539–544.

Sandoval, G. J., J. L. Pulice, H. Pakula, M. Schenone, D. Y. Takeda, M. Pop, G. Boulay, K. E. Williamson, M. J. McBride, et al. (2018). "Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis". *Molecular Cell* 71.4, pp. 554–566.

Sanjana, N. E., J. Wright, K. Zheng, O. Shalem, P. Fontanillas, J. Joung, C. Cheng, A. Regev, and F. Zhang (2016). "High-resolution interrogation of functional elements in the noncoding genome." *Science* 353.6307, pp. 1545–1549.

Santagata, S., F. Demichelis, A. Riva, S. Varambally, M. D. Hofer, J. L. Kutok, R. Kim, J. Tang, J. E. Montie, et al. (2004). "JAGGED1 Expression Is Associated with Prostate Cancer Metastasis and Recurrence". *Cancer Research* 64.19, pp. 6854–6857.

Sanyal, A., B. R. Lajoie, G. Jain, and J. Dekker (2012). "The long-range interaction landscape of gene promoters". *Nature* 489.7414, pp. 109–113.

Schlereth, K., D. Weichenhan, T. Bauer, T. Heumann, E. Giannakouri, D. Lipka, S. Jaeger, M. Schlesner, P. Aloy, et al. (2018). "The transcriptomic and epigenetic map of vascular quiescence in the continuous lung endothelium". *eLife* 7, Article e34423.

Scott, D. A. and F. Zhang (2017). "Implications of human genetic variation in CRISPR-based therapeutic genome editing". *Nature Medicine* 23.9, pp. 1095–1101.

Selvaraj, N., J. A. Budka, M. W. Ferris, J. P. Plotnik, and P. C. Hollenhorst (2015). "Extracellular Signal-Regulated Kinase Signaling Regulates the Opposing Roles of JUN Family Transcription Factors at ETS/AP-1 Sites and in Cell Migration". *Molecular and Cellular Biology* 35.1, pp. 88–100.

Shah, A. V., G. M. Birdsey, C. Peghaire, M. E. Pitulescu, N. P. Dufton, Y. Yang, I. Weinberg, L. Osuna Almagro, L. Payne, et al. (2017). "The endothelial transcription factor ERG mediates Angiopoietin-1-dependent control of Notch signalling and vascular stability". *Nature Communications* 8, Article 16002.

Shah, A. V., G. M. Birdsey, and A. M. Randi (2016). "Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG". *Vascular Pharmacology* 86, pp. 3–13.

Shalem, O., N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, et al. (2014). "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells". *Science* 343.6166, pp. 84–87.

Shi, X., J. Richard, K. M. Zirbes, W. Gong, G. Lin, M. Kyba, J. A. Thomson, N. Koyano-Nakagawa, and D. J. Garry (2014). "Cooperative interaction of Etv2 and Gata2 regulates the development of endothelial and hematopoietic lineages". *Developmental Biology* 389.2, pp. 208–218.

Shin, H. Y., M. Willi, K. H. Yoo, X. Zeng, C. Wang, G. Metser, and L. Hennighausen (2016). "Hierarchy within the mammary STAT5-driven Wap super-enhancer". *Nature Genetics* 48.8, pp. 904–911.

Shrinivas, K., B. R. Sabari, E. L. Coffey, I. A. Klein, A. Boija, A. V. Zamudio, J. Schuijers, N. M. Hannett, P. A. Sharp, et al. (2019). "Enhancer Features that Drive Formation of Transcriptional Condensates". *Molecular Cell* 75.3, pp. 549–561.

Siersbæk, R., A. Rabiee, R. Nielsen, S. Sidoli, S. Traynor, A. Loft, L. L. C. Poulsen, A. Rogowska-Wrzesinska, O. N. Jensen, and S. Mandrup (2014). "Transcription Factor Cooperativity in Early Adipogenic Hotspots and Super-Enhancers". *Cell Reports* 7.5, pp. 1443–1455.

Song, L. and G. E. Crawford (2010). "DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells." *Cold Spring Harbor Protocols* 2, Protocol 5384.

Sorensen, P. H., S. L. Lessnick, D. Lopez-Terrada, X. F. Liu, T. J. Triche, and C. T. Denny (1994). "A second Ewing's sarcoma translocation, t(21;22), fuses the EWS

gene to another ETS–family transcription factor, ERG". *Nature Genetics* 6.2, pp. 146–151.

Al-Soudi, A., M. Kaaij, and S. Tas (2017). "Endothelial cells: From innocent bystanders to active participants in immune responses". *Autoimmunity Reviews* 16.9, pp. 951–962.

Sperone, A., N. H. Dryden, G. M. Birdsey, L. Madden, M. Johns, P. C. Evans, J. C. Mason, D. O. Haskard, J. J. Boyle, et al. (2011). "The transcription factor Erg inhibits vascular inflammation by repressing NF-kappaB activation and proinflammatory gene expression in endothelial cells." *Arteriosclerosis, Thrombosis, and Vascular Biology* 31.1, pp. 142–150.

Spitz, F. and E. E. M. Furlong (2012). "Transcription factors: from enhancer binding to developmental control". *Nature Reviews Genetics* 13.9, pp. 613–626.

Sriram, G., J. Y. Tan, I. Islam, A. J. Rufaihah, and T. Cao (2015). "Efficient differentiation of human embryonic stem cells to arterial and venous endothelial cells under feeder- and serum-free conditions". *Stem Cell Research & Therapy* 6.1, Article 261.

Stunnenberg, H. G., M. Hirst, S. Abrignani, D. Adams, M. de Almeida, L. Altucci, V. Amin, I. Amit, S. E. Antonarakis, et al. (2016). "The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery". *Cell* 167.5, pp. 1145–1149.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proceedings of the National Academy of Sciences* 102.43, 15545–155tak50.

Taddei, A., C. Giampietro, A. Conti, F. Orsenigo, F. Breviario, V. Pirazzoli, M. Potente, C. Daly, S. Dimmeler, and E. Dejana (2008). "Endothelial adherens junctions control tight junctions by VE-cadherin-mediated upregulation of claudin-5". *Nature Cell Biology* 10.8, pp. 923–934.

Takahashi, H., T. J. Parmely, S. Sato, C. Tomomori-Sato, C. A. Banks, S. E. Kong, H. Szutorisz, S. K. Swanson, S. Martin-Brown, et al. (2011). "Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors". *Cell* 146.1, pp. 92–104.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." *Cell* 126.4, pp. 663–676.

Taoudi, S., T. Bee, A. Hilton, K. Knezevic, J. Scott, T. A. Willson, C. Collin, T. Thomas, A. K. Voss, et al. (2011). "ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification". *Genes & Development* 25.3, pp. 251–262.

Thoms, J. A. I., Y. Birger, S. Foster, K. Knezevic, Y. Kirschenbaum, V. Chandrakanthan, G. Jonquieres, D. Spensberger, J. W. Wong, et al. (2011). "ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer". *Blood* 117.26, pp. 7079–7089.

Thurman, R. E., E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, et al. (2012). "The accessible chromatin landscape of the human genome". *Nature* 489.7414, pp. 75–82.

Tomlins, S. A., D. R. Rhodes, S. Perner, S. M. Dhanasekaran, R. Mehra, X.-W. Sun, S. Varambally, X. Cao, J. Tchinda, et al. (2005). "Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer." *Science* 310.5748, pp. 644–648.

Vahedi, G., Y. Kanno, Y. Furumoto, K. Jiang, S. C. J. Parker, M. R. Erdos, S. R. Davis, R. Roychoudhuri, N. P. Restifo, et al. (2015). "Super-enhancers delineate disease-associated regulatory nodes in T cells". *Nature* 520.7548, pp. 558–562.

Vandel, J., O. Cassan, S. Lèbre, C.-H. Lecellier, and L. Bréhélin (2019). "Probing transcription factor combinatorics in different promoter classes and in enhancers". *BMC Genomics* 20.1, Article 103.

Verger, A., E. Buisine, S. Carrère, R. Wintjens, A. Flourens, J. Coll, D. Stéhelin, and M. Duterque-Coquillaud (2001). "Identification of amino acid residues in the ETS transcription factor Erg that mediate Erg-Jun/Fos-DNA ternary complex formation." *The Journal of Biological Chemistry* 276.20, pp. 17181–17189.

Vestweber, D. (2008). "VE-Cadherin: the major endothelial adhesion molecule controlling cellular junctions and blood vessel formation". *Arteriosclerosis, Thrombosis, and Vascular Biology* 28.2, pp. 223–232.

Vierbuchen, T., E. Ling, C. J. Cowley, C. H. Couch, X. Wang, D. A. Harmin, C. W. Roberts, and M. E. Greenberg (2017). "AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection". *Molecular Cell* 68.6, pp. 1067–1082.

Vijayaraj, P., A. Le Bras, N. Mitchell, M. Kondo, S. Juliao, M. Wasserman, D. Beeler, K. Spokes, W. C. Aird, et al. (2012). "Erg is a crucial regulator of endocardial-mesenchymal transformation during cardiac valve morphogenesis." *Development* 139.21, pp. 3973–3985.

Villar, D., C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, et al. (2015). "Enhancer Evolution across 20 Mammalian Species". *Cell* 160.3, pp. 554–566.

Visel, A., M. J. Blow, Z. Li, T. Zhang, J. A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, et al. (2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers". *Nature* 457.7231, pp. 854–858.

Vlaeminck-Guillem, V., S. Carrere, F. Dewitte, D. Stehelin, X. Desbiens, and M. Duterque-Coquillaud (2000). "The Ets family member Erg gene is expressed in mesodermal tissues and neural crests at fundamental steps during mouse embryogenesis." *Mechanisms of Development* 91.1-2, pp. 331–335.

Vogeli, K. M., S.-W. Jin, G. R. Martin, and D. Y. R. Stainier (2006). "A common progenitor for haematopoietic and endothelial lineages in the zebrafish gastrula". *Nature* 443.7109, pp. 337–339.

Wakiya, K., A. Begue, D. Stehelin, and M. Shibuya (1996). "A cAMP Response Element and an Ets Motif Are Involved in the Transcriptional Regulation of *flt* -1 Tyrosine

Kinase (Vascular Endothelial Growth Factor Receptor 1) Gene". *Journal of Biological Chemistry* 271.48, pp. 30823–30828.

Wang, D., P. S. Gill, T. Chabrashvili, M. L. Onozato, J. Raggio, M. Mendonca, K. Dennehy, M. Li, P. Modlinger, et al. (2007). "Isoform-specific regulation by N(G),N(G)-dimethylarginine dimethylaminohydrolase of rat serum asymmetric dimethylarginine and vascular endothelium-derived relaxing factor/NO." *Circulation Research* 101.6, pp. 627–635.

Wang, S., R. K. Kollipara, N. Srivastava, R. Li, P. Ravindranathan, E. Hernandez, E. Freeman, C. G. Humphries, P. Kapur, et al. (2014). "Ablation of the oncogenic transcription factor ERG by deubiquitinase inhibition in prostate cancer." *Proceedings of the National Academy of Sciences* 111.11, pp. 4251–4256.

Watanabe, K., M. Umićević Mirkov, C. A. de Leeuw, M. P. van den Heuvel, and D. Posthuma (2019). "Genetic mapping of cell type specificity for complex traits". *Nature Communications* 10.1, Article 3222.

Wei, B., A. Jolma, B. Sahu, L. M. Orre, F. Zhong, F. Zhu, T. Kivioja, I. K. Sur, J. Lehtiö, et al. (2017). "Strong binding activity of few transcription factors is a major determinant of open chromatin". *bioRxiv*, Article 204743.

Wernig, M., A. Meissner, R. Foreman, T. Brambrink, M. Ku, K. Hochedlinger, B. E. Bernstein, and R. Jaenisch (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state". *Nature* 448.7151, pp. 318–324.

West, J. A., C. P. Davis, H. Sunwoo, M. D. Simon, R. I. Sadreyev, P. I. Wang, M. Y. Tolstorukov, and R. E. Kingston (2014). "The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites". *Molecular Cell* 55.5, pp. 791–802.

Whyte, W. A., D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, et al. (2013). "Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes". *Cell* 153.2, pp. 307–319.

Wilson, N. K., S. D. Foster, X. Wang, K. Knezevic, J. Schütte, P. Kaimakis, P. M. Chilarska, S. Kinston, W. H. Ouwehand, et al. (2010). "Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators." *Cell Stem Cell* 7.4, pp. 532–544.

Wu, D., B. Sunkel, Z. Chen, X. Liu, Z. Ye, Q. Li, C. Grenade, J. Ke, C. Zhang, et al. (2014). "Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer." *Nucleic Acids Research* 42.6, pp. 3607–3622.

Wu, T., Y. F. Kamikawa, and M. E. Donohoe (2018). "Brd4's Bromodomains Mediate Histone H3 Acetylation and Chromatin Remodeling in Pluripotent Cells through P300 and Brg1". *Cell Reports* 25.7, pp. 1756–1771.

Yamamoto, K. R. (1985). "Steroid Receptor Regulated Transcription of Specific Genes and Gene Networks". *Annual Review of Genetics* 19.1, pp. 209–252.

Yang, Y., A. M. Blee, D. Wang, J. An, Y. Pan, Y. Yan, T. Ma, Y. He, J. Dugdale, et al. (2017). "Loss of FOXO1 Cooperates with TMPRSS2–ERG Overexpression to Pro-

mote Prostate Tumorigenesis and Cell Invasion". *Cancer Research* 77.23, pp. 6524–6537.

Yin, J.-w. and G. Wang (2014). "The Mediator complex: a master coordinator of transcription and cell lineage development." *Development* 141.5, pp. 977–987.

Yoshimoto, M., A. M. Joshua, I. W. Cunha, R. A. Coudry, F. P. Fonseca, O. Ludkovski, M. Zielenska, F. A. Soares, and J. A. Squire (2008). "Absence of TMPRSS2:ERG fusions and PTEN losses in prostate cancer is associated with a favorable outcome". *Modern Pathology* 21.12, pp. 1451–1460.

Yu, G., L.-G. Wang, and Q.-Y. He (2015). "ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization." *Bioinformatics* 31.14, pp. 2382–2383.

Yu, J., J. Yu, R.-S. Mani, Q. Cao, C. J. Brenner, X. Cao, X. Wang, L. Wu, J. Li, et al. (2010). "An Integrated Network of Androgen Receptor, Polycomb, and TMPRSS2-ERG Gene Fusions in Prostate Cancer Progression". *Cancer Cell* 17.5, pp. 443–454.

Yuan, L., V. Nikolova-Krstevski, Y. Zhan, M. Kondo, M. Bhasin, L. Varghese, K. Yano, C. V. Carman, W. C. Aird, and P. Oettgen (2009). "Antiinflammatory Effects of the ETS Factor ERG in Endothelial Cells Are Mediated Through Transcriptional Repression of the Interleukin-8 Gene". *Circulation Research* 104.9, pp. 1049–1057.

Yuan, L., A. Sacharidou, A. N. Stratman, A. Le Bras, P. J. Zwiers, K. Spokes, M. Bhasin, S.-c. Shih, J. A. Nagy, et al. (2011). "RhoJ is an endothelial cell-restricted Rho GTPase that mediates vascular morphogenesis and is regulated by the transcription factor ERG". *Blood* 118.4, pp. 1145–1153.

Yuan, L., A. Le Bras, A. Sacharidou, K. Itagaki, Y. Zhan, M. Kondo, C. V. Carman, G. E. Davis, W. C. Aird, and P. Oettgen (2012). "ETS-related gene (ERG) controls endothelial cell permeability via transcriptional regulation of the claudin 5 (CLDN5) gene." *The Journal of Biological Chemistry* 287.9, pp. 6582–6591.

Zammarchi, F., G. Boutsalis, and L. Cartegni (2013). "5 UTR Control of Native ERG and of Tmprss2:ERG Variants Activity in Prostate Cancer". *PLoS ONE* 8.3. Ed. by B. Tian, Article e49721.

Zaret, K. S. and J. S. Carroll (2011). "Pioneer transcription factors: establishing competence for gene expression." *Genes & Development* 25.21, pp. 2227–2241.

Zentner, G. E., P. J. Tesar, and P. C. Scacheri (2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions." *Genome Research* 21.8, pp. 1273–1283.

Zhang, B., D. S. Day, J. W. Ho, L. Song, J. Cao, D. Christodoulou, J. G. Seidman, G. E. Crawford, P. J. Park, and W. T. Pu (2013). "A dynamic H3K27ac signature identifies VEGFA-stimulated endothelial enhancers and requires EP300 activity". *Genome Research* 23.6, pp. 917–927.

Zhang, X., P. S. Choi, J. M. Francis, M. Imielinski, H. Watanabe, A. D. Cherniack, and M. Meyerson (2016). "Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers." *Nature Genetics* 48.2, pp. 176–182.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, et al. (2008). "Model-based analysis of ChIP-Seq (MACS)." *Genome Biology* 9.9, Article R137.

Zhang, Z., K. R. Chng, S. Lingadahalli, Z. Chen, M. H. Liu, H. H. Do, S. Cai, N. Rinaldi, H. M. Poh, et al. (2019). "An AR-ERG transcriptional signature defined by long-range chromatin interactomes in prostate cancer cells." *Genome Research* 29.2, pp. 223–235.

Zhao, J. C., K.-W. Fong, H.-J. Jin, Y. A. Yang, J. Kim, and J. Yu (2016). "FOXA1 acts upstream of GATA2 and AR in hormonal regulation of gene expression". *Oncogene* 35.33, pp. 4335–4344.

Zhou, P., F. Gu, L. Zhang, B. N. Akerberg, Q. Ma, K. Li, A. He, Z. Lin, S. M. Stevens, et al. (2017). "Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq". *eLife* 6, Article e22039.

Zhu, F., L. Farnung, E. Kaasinen, B. Sahu, Y. Yin, B. Wei, S. O. Dodonova, K. R. Nitta, E. Morgunova, et al. (2018). "The interaction landscape between transcription factors and the nucleosome". *Nature* 562.7725, pp. 76–81.

Zuin, J., J. R. Dixon, M. I. J. A. van der Reijden, Z. Ye, P. Kolovos, R. W. W. Brouwer, M. P. C. van de Corput, H. J. G. van de Werken, T. A. Knoch, et al. (2014). "Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells". *Proceedings of the National Academy of Sciences* 111.3, pp. 996–1001.

APPENDIX: PERMISSION TO REPUBLISH

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 29, 2019

This Agreement between Imperial College London -- Viktoria Kalna ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4680070883921
License date	Oct 01, 2019
Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	Reference Module in Biomedical Sciences
Licensed Content Author	Bruce M. Carlson
Licensed Content Date	Jan 1, 2014
Licensed Content Pages	1
Start Page	
End Page	
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier chapter?	No
Will you be translating?	No
Original figure numbers	Figure 1
Title of your thesis/dissertation	Transcriptional and epigenetic regulation of lineage identity in endothelial cells by the transcription factor ERG via super-enhancers
Expected completion date	Oct 2019
Estimated size (number of pages)	200
Requestor Location	Imperial College London Hammersmith Hospital Campus Du Cane Road London, Non-US/Non-Canadian W12 0NN United Kingdom Attn: Imperial College London
Publisher Tax ID	GB 494 6272 12
Billing Type	Invoice
Billing Address	Imperial College London Hammersmith Hospital Campus