The landscape of vascular accessible chromatin and prioritisation of functional risk alleles for cardiovascular disease

Kristina Boychova Markova

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Supervisors:
Professor Panos Deloukas
Doctor Andrew Smith

Clinical Pharmacology

William Harvey Research Institute

Faculty of Medicine and Dentistry

Queen Mary University of London

2021

ACKNOWLEDGMENTS

I am thankful to a number of people that have supported me during my PhD. First, I would like to express my appreciations to my supervisor Prof Deloukas for his guidance throughout these years. I am grateful for the opportunities provided by him to extend my knowledge in the field of cardiovascular genetics. Second, a huge thanks to my supervisor Dr Andrew Smith for his constant support and patience. Dr Smith was next to me in every single step during the experimental work throughout my PhD. Thirdly, a huge thanks to Ms Li for her support with tissue culture and many of the approaches described here, including Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq). Special thanks to Dr Bourgeois for his support with computational work, from software installation to complex scripting that required my analyses; Dr Marouli for her support and sharing of presentation templates. I am also grateful to all the members of the cardiovascular genomics group for their support throughout this process. Working with them was valuable experience, helping me to develop my skills and knowledge necessary for both my PhD and future career. Last, I would like to thank my family and friends for their support throughout these years.

STATEMENT OF ORIGINALITY

I, Kristina Markova, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below, and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: 19/12/2021

ABSTRACT

Cardiovascular disease is a leading cause of mortality and disability. A meta-analysis of genome-wide association studies (GWAS) of 181,522 coronary artery (CAD) cases and 984,058 controls identified 897 CAD risk variants at the 1% false discovery rate whereas a blood pressure (BP) GWAS meta-analysis in over 1 million individuals identified 901 BP risk alleles. To prioritise CAD and BP variants for functional studies I mapped the vascular landscape of accessible chromatin in human coronary artery endothelial cells (HCAECs) and smooth muscle cells (HCASMCs). Furthermore, I stimulated HCAECs with vascular endothelial growth factor A (VEGFA) to assess its effect on accessible chromatin. This study led to the identification of 86,811 ATAC-seq peaks in unstimulated HCAECs, 108,736 ATAC-seq peaks in VEGFA-stimulated HCAECs, and 209,743 ATAC-seq peaks in HCASMCs. Ras and cellular senescence pathways were specific to HCAEC, whereas cAMP, vascular smooth muscle contraction, and prolactin pathways were specific to HCASMC. The most significantly enriched transcription factor binding motifs in HCAEC included CTCF, BORIS, ATF3, JUNB, and Fra1, whereas in HCASMC I found BATF, AFT3, FRA1, AP-1 and JunB. A total of 17,570 peaks were found to be restricted to HCAEC accessible chromatin, when compared to ENCODE DNaseI-seq data from 125 cell types. They showed enrichment of TGF-beta and Wnt signalling pathways. There are 39,935 ATAC-seq peaks in the unique VEGFA-stimulated HCAEC, which showed enrichment for phospholipase D (PLD) signalling. Intersecting the ATAC-seq data with 21,461 CAD risk variants (all proxies of 897 index variants at r2 >0.8) as well as 14,168 BP variants (all proxies of 901 index variants at r2 >0.8) identified 499 CAD and 517 BP risk variants, respectively, located in vascular ATAC-seq peaks. These variants were further prioritised using tools such as RegulomeDB, Haploreg, and GTEx. In addition, further prioritization of functional CAD SNPs was performed based on the FGWAS approach. Examination of accessible chromatin in vascular cells can be used to prioritise functional variants, facilitating further focused studies on these loci and the eventual identification of novel therapeutics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

12

ABBREVIATIONS

Acute coronary syndrome	ACS
Adenosine Kinase	ADK
Allele-specific hypersensitivity	ASH
Angiotensin II protein J	APJ
Antigen-presenting cells	APCs
Assay for Transposase-Accessible Chromatin using sequencing	ATAC-seq
ATAC-Resuspension Buffer	RSB
Binary version of a SAM file	BAM
Blood pressure	BP
Bone Morphogenetic Protein 6	BMP6
Brother of the regulator of imprinting sites	BORIS
Cardiovascular disease	CVD
Cathepsin K	CTSK
Cathepsin S	CTSS
CCCTC-Binding Factor	CTCF
Cell division control protein 42 homolog	CDC42
Coronary artery disease	CAD
Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis CARDIoGRAM	CARDIoGRAMplus C4D
Cyclo-oxygenase 1	COX-1
Diastolic	DBP
DNase hypersensitivity sites	DHSs
DNase I hypersensitive sites sequencing	DNase-seq
Downstream core element	DCE
Downstream promoter element	DPE
Early growth response	EGR
Electrophoretic Mobility Shift Assays	EMSAs
Endothelial cells	ECs
Endothelial NO synthase	eNOS
Enhancer RNAs	eRNA

Estonian Genome Centre University of Tartu	EGCUT
Estrogen related receptor alpha	ESRRA
Extracellular-signal-regulated kinase-1/2	ERK1/2
False discovery rate	FDR
FES Proto-Oncogene Tyrosine Kinase	FES
Fibroblast growth factor 2	FGF2
Formaldehyde-Assisted Isolation of Regulatory Elements	FAIRE-seq
Four And A Half LIM Domains 2	FHL2
Fraction of reads in peaks	FRiP
Genome-wide association studies	GWAS
Hairy/enhancer-of-split related with YRPW motif protein 2	HEY2
Haplotype Reference Consortium	HRC
Heparan sulphate proteoglycans	HSPGs
High density lipoprotein	HDL
Histone H3 lysine 27 acetylation	H3K27ac
Histone H3 lysine 36 trimethylation	H3K36me3
Histone H3 lysine 4 monomethylation	H3K4me1
Histone H3 lysine 4 trimethylation	H3K4me3
Human Carotid Artery Endothelial Cells	HCtAEC
Human Coronary Artery Endothelial Cells	HCAEC
Human Coronary Artery Smooth Muscle Cells	HCASMC
Human umbilical vein endothelial cells	HUVECs
Hypergeometric Optimization of Motif EnRichment	HOMER
Immortalised human umbilical vein endothelial cells	iHUVECs
Initiator element	Inr
Interferon-regulatory factor	IRF
International Consortium of Blood Pressure	ICBP
Ischemic heart disease	IHD
Linkage disequilibrium	LD
Locus control regions	LCR
Low density lipoprotein	LDL
Massively Parallel Reporter Assay	MPRA

Micrococcal nuclease digestion with deep sequencing	MNaseq-seq
Million Veterans Program	MVP
Mitogen-activated protein kinase	MAPK
Model-based Analysis for ChIP-Seq	MASC2
Motif ten element	MTE
Myeloid zinc finger 1	MZF1
Myocardial infarction	MI
Next-generation sequencing	NGS
Nitric oxide	NO
Nitric Oxide Synthase 3	NOS3
Nuclear receptor	NR4A
Nucleosome-depleted region	NDR
Paxillin	PXN
Phosphate buffered saline	PBS
Phosphoinositide 3-kinase	PI3K
Phospholipase Cγ	PLCγ
Phospholipase D	PLD
Placenta growth factor	PLGF
Platelet Derived Growth Factor Subunit B	PDGFB
Platelet-derived growth factor	PDGF
Platelet-derived growth factor receptors	PDGFRs
Polycomb group response elements	PREs
Polymerase chain reaction	PCR
Post-translational modification	PTMs
Preinitiation complex	PIC
Probability of association	PPA
Protein kinase C	PKC
Protein kinase D	PKD
Pulse pressure	PP
Quality control	QC
Ras GTPase-activating-like protein IQGAP1	IQGAP1
Receptor tyrosine kinases	RTKs

Rho guanine nucleotide exchange factor 26	ARHGEF26
RNA polymerase II	RNAPII
Runs of Homozygosity	ROH
Self-transcribing active regulatory region sequencing	STARR-seq
Sequence alignment map	SAM
Sequencing-by-synthesis	SBS
Single-nucleotide polymorphism	SNP
SMAD Family Member 1	SMAD1
SMAD Family Member 3	SMAD3
Small interfering RNA	siRNA
SP100 nuclear antigen	SP100
Sphingosine kinase-1	SK1
Sphingosine-1-phosphate	S1P
Spliced Transcripts Alignment to a Reference	STAR
Switching B cell complex subunit	SWAP70
Systolic	SBP
TFIIB-recognition element	BRE
The Encyclopedia of DNA Elements	ENCODE
The Monitoring Trends and Determinants in Cardiovascular Diseases	MONICA
Toll-like receptor `	TLR
Transcription Factor 7 Like 2	TCF7L2
Transcription factors	TFs
Transcription start sites	TSSs
Transforming Growth Factor Beta 1	TGFB1
Type 2 diabetes	T2D
UK Biobank	UKBB
University of California Santa Cruz	UCSC
Vascular endothelial growth factor	VEGF
Vascular smooth muscle cells	VSMCs

1. Introduction

1.1 From genome-wide association studies (GWAS) signals to functional characterisation of causal variants

Common human diseases and traits are influenced by both genetic and environmental factors. Despite the success in single-gene Mendelian disorders, the limited success of linkage studies in complex diseases has been attributed to their low power and resolution for variants of modest effect [1, 2]. GWAS represent a powerful tool for investigating the genetic architecture of complex traits and diseases [3, 4]. GWAS are important as they can also improve our understanding of the molecular mechanisms behind disease.

The underlying rationale for GWAS is the 'common disease, common variant' hypothesis, which postulates that common diseases are attributable in part to allelic variants present in more than 1–5% of the population [5, 6]. They have been facilitated by the development of commercial single-nucleotide polymorphism (SNP) chips or arrays that capture most, although not all, common variation in the genome.

Systematic studies of common genetic variants are facilitated by the fact that individuals who carry a particular SNP allele at one site often predictably carry specific alleles at other nearby variant sites. This correlation is known as linkage disequilibrium (LD), a particular combination of alleles along a chromosome is termed a haplotype.

More than 95% of common variants were discovered in the pilot phase of the 1000 Genomes Project, lower-frequency variants, particularly variants outside the coding exome, remain poorly characterised. However, recent efforts to map human genetic variation by exome sequencing [7] and whole genome sequencing [8-10] have identified the vast majority of common SNPs across the genome. Low-frequency variants exhibit increased levels of populations differentiation [11-13].

GWAS have improved our understanding of the genetic basis of many complex traits [14]. For several diseases, including breast cancer [15] [16, 17], prostate cancer [18-22], inflammatory bowel disease [23-27] and type 1 [28, 29] and type 2 diabetes [30-35], there has been rapid increase in the numbers of loci implicated in disease predisposition. Many hundreds of, mainly common, risk variants underlying the genetic architecture of continuous traits, including fat mass [36-38], adult height [39, 40], and blood circulating lipid levels [33, 41-43] have also been found.

GWAS findings have provided substantial insights into the genetic architecture of complex traits but, despite the progress and advancements, there are many challenges ahead. For most traits, known variants explain only a fraction of heritability, limiting the potential for early application to determine individual disease risk. Many explanations for missing heritability have been suggested, including larger number of variants of smaller effect yet to be found, rare variants that are poorly detected, structural variants that are poorly detected, lower power to detect gene-gene interactions, and inadequate accounting for shared environment among relatives. The power of GWAS to detect an association at a particular SNP is a function of sample size as well as the frequency and effect of the causal variant. Thus, studies have more power to detect associations with common variants with few of the reported associations involving rare variants. More extensive analyses using whole exome and genome sequence are needed to tackle rare variants, whereas follow-up studies and bigger meta-analyses of published studies are needed to catalogue all common variants associated to disease. In that context, increasing sample size for non-European descent individuals will be key to capture the full spectrum of allelic heterogeneity in complex traits including disease.

The biggest challenge of all is to identify the causal variants in the thousands of already discovered GWAS signals and elucidate the molecular mechanisms through which these variants function leading to the observed phenotype. The interplay between genetic and environmental factors further complicates this.

Despite the progress of biomedical research, the causes of common human diseases remain largely unknown at the molecular level. Family history is one of the main risk factors for nearly all diseases, including cardiovascular disease, cancer, diabetes, autoimmunity, psychiatric illnesses, and many others. Identifying the causal genes and variants would be an important step to improve prevention, diagnosis, and treatment of diseases. Although, the vast majority (>80%) of associated variants fall outside coding regions, emphasising the importance of understanding regulatory variation.

1.2 The functional genome

The human genome sequence provides the underlying code for human biology. Our understanding of the non-coding RNAs, alternatively spliced transcripts and regulatory sequences although paramount in functionally characterising GWAS signals is far from complete. Systematic analyses are essential to identify genes and regulatory regions. These

analyses can provide insights into the organisation of genes and regulatory information across cellular contexts, species, and individuals. project provides information about all functional elements in the human genome [44-46]. Functional elements are defined as a discrete genome segment that encode a defined product or display a reproducible biochemical signature. Comparative genomic studies demonstrate that 3-8% of bases maybe functional [47-51]. In a pilot phase covering 1% of the genome, the Encyclopedia of DNA Elements (ENCODE) project annotated 60% of mammalian evolutionarily constrained bases, but also identified many additional putative functional elements without evidence of constraint, they provide little, if any insight into those bases that are biologically important [45].

The advent of more powerful DNA sequencing technologies now enables whole genome and more precise analyses with a broad repertoire of functional assays. Recently, advances in DNA sequencing technologies have enabled whole-genome analyses with a broad repertoire of functional assays. The majority (>80%) of the human genome is involved in at least one RNA- and/or chromatin associated event. The majority of the human genome lies close to regulatory elements: 99% lies within 1.7 kb of a least one of the biochemical events measured by the ENCODE and 955 of the genomes lie with 8 kb of a DNA-protein interaction.

RNA sequence production and processing can be correlated with both chromatin marks and transcription factor binding at promoters. This indicates that promoter functionality can explain variation in RNA gene expression. Many non-coding variants lie within annotated functional regions. The majority of variants associated with disease using GWAS lie outside protein- coding regions. These variants are enriched within non-coding functional elements, with a majority located in or near ENCODE-defined regions that are outside of protein-coding genes. In many cases, the disease phenotypes can be associated with a specific cell type or transcription factor [52].

1.3 Chromatin

There are many ways non-coding functional elements can be identified, one of these is through examining the chromatin structure. DNA structure can provide clues to its functionality. In eukaryotes genomic DNA is packed to form chromatin. The genome is made accessible for readout by the complex machinery involved in DNA repair, DNA replication and gene transcription. Nucleosomes are formed by wrapping ~146 bp of DNA around a histone octamer core [53]. Short DNA fragments (termed 'linker DNA') connect nucleosomes into nucleosomes

arrays, which undergo short-range interactions with neighboring nucleosomes to form chromatin fibres. Subsequent fibre–fibre interactions contribute to the high degree chromosome compaction [54]. The beads-on-a-string organisation of individual nucleosomes can be termed the 'primary structure' of chromatin, which in turn defines 'secondary' and 'tertiary' higher-order chromatin structures (**Figure 1**) [55].
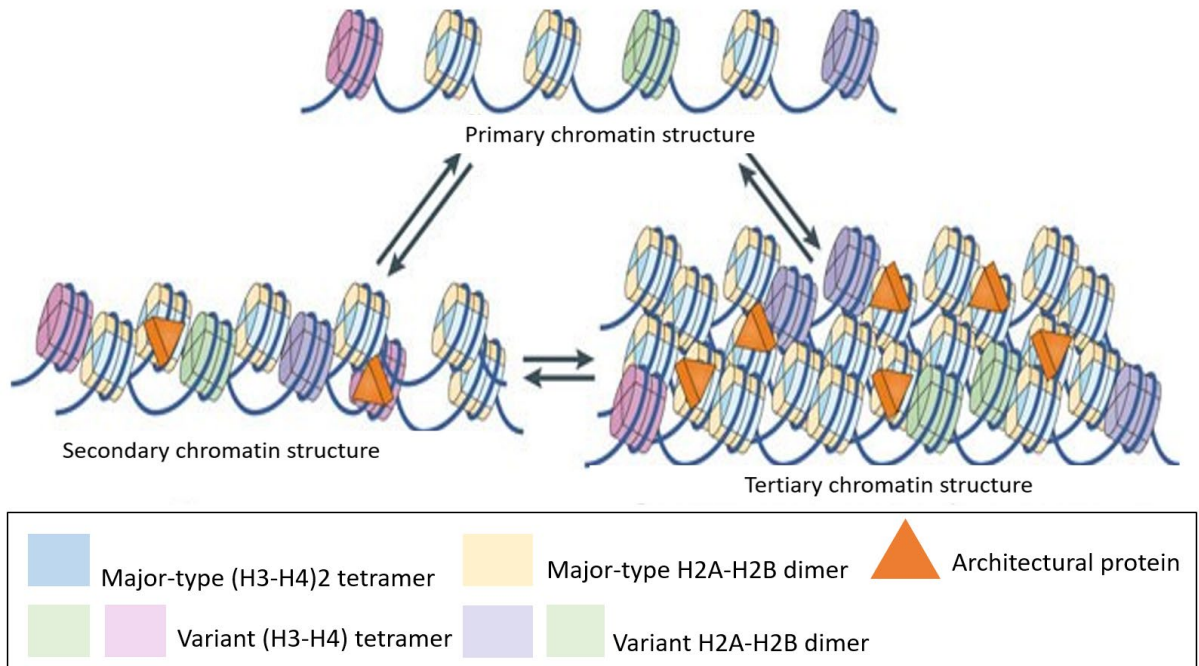
Figure 1. Primary, secondary, and tertiary chromatin structure (adapted from Luger *et al.,* 2012).

There is a large collection of high-resolution nucleosome crystal structures from different species, showing nucleosomes, histone variants and post-translational modification (PTMs) in complex with nuclear proteins [56]. Single-molecule approaches have revealed the dynamic properties of nucleosomes that were not apparent from the crystal structures [57]. It is now clear that the various crystal structures represent one possible state of the nucleosome, and that the incorporation of PTMs and histone variants has the potential to shift the equilibrium between different structural states. This affects the chromatin fibre compaction and nucleosome interaction with non-histone proteins. Numerous *in vitro* studies have addressed the effects of histone variants and PTMs on chromatin condensation. Several experimental and computational models for higher-order structure have been proposed. High-resolution sequencing approaches have been used to map nucleosome position over the entire genomes to near base-pair resolution [58]. Progress has been made in predicting nucleosome position from DNA sequence alone [59].

1.4 Biophysical determinants of accessibility

Physical access to DNA is determined by nucleosome organisation and occupancy, and DNA-associated macromolecules, including transcription factors (TFs) and architectural proteins. TF binding is correlated with accessibility due to a variety of reasons. First, the footprint of DNA is typically smaller than the ~146 bp occupied by a nucleosome. Second, molecular interactions are stochastic, thus providing more frequent access to unbound DNA. Finally, sequence dependent binding of proteins to DNA provides a recruitment substrate for nonspecific chromatin remodelers, which further open proximal chromatin. Both linear and topological organisations of histones have an important role in chromatin accessibility [60-62].

1.5 Nucleosome occupancy, density, and turnover

Chromatin accessibility is influenced by the density of associated proteins along the DNA and residence time of these factors. Regions of chromatin accessibility are distributed across a broad range of nucleosome turnover rates and occupancy levels. For instance, CCCTC-Binding Factor (CTCF)-bound insulators lie in regions associated with stable nucleosomes, and transcription start sites (TSSs) lie in regions associated with unstable nucleosomes. Nucleosome positioning and stability play an important role in regulating DNA access by TFs (**Figure 2**) [63-66].
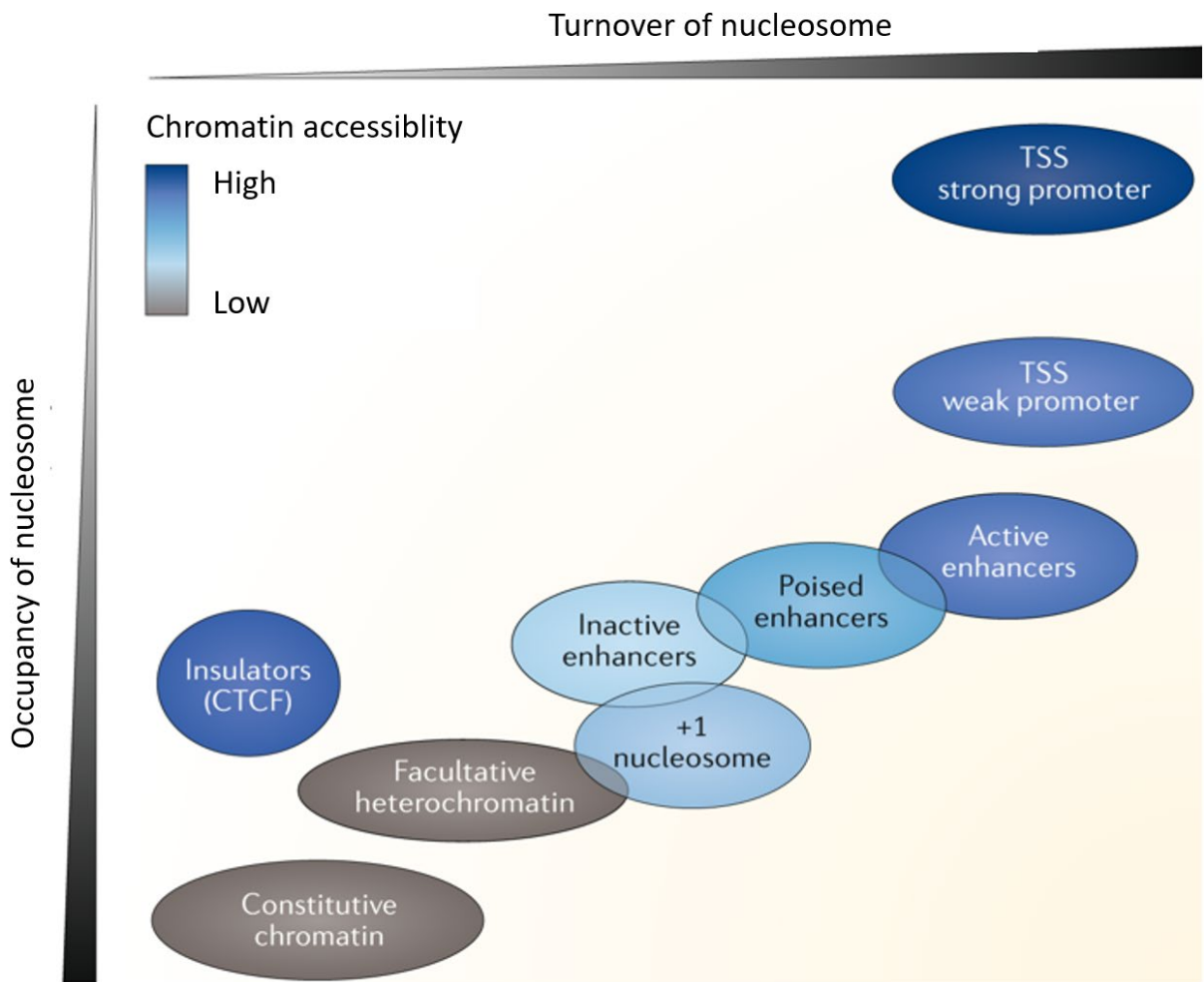
Figure 2. Nucleosome occupancy. Chromatin accessibility is associated with region of high nucleosome turnover, but also with dense nucleosome occupancy, and transcription start site (adapted from Klemm *et al.,* 2019).

Nucleosome occupancy is dynamically regulated, with higher turnover rates in active promoters and enhancers compared to inactive and heterochromatic regions [62, 67, 68]. For many promoters, the nucleosome-depleted region (NDR) [69] is defined by two-well positioned nucleosomes upstream and downstream of the TSS. Nucleosome exclusion within the NDR is maintained by a variety of factors, including promoter-proximally paused RNA polymerase and BRG1/BRM-associated factor [70]. Depletion of nucleosomes that occlude access to regulatory DNA is commonly observed. For example, CTCF-bound insulators, RNA-polymerase-bound TSSs and glucocorticoid-receptor-bound enhancers demonstrate that high-occupancy TFs coupled with cofactor-dependent and ATP-dependent chromatin remodeler recruitment [63, 71]. These structural motifs demonstrate statistical nucleosome positioning [72-74] upstream and downstream of strong DNA-protein interactions.

Nucleosome turnover rates are modulated by a variety of factors, including occupancy by linker histones and TFs, the local chromatin context (comprising histone density and ATP-dependent chromatin remodeler activity), histone chaperone availability, histone variant composition, and DNA sequence [75-80]. In eukaryotes, nucleosome positional activity may be due to multicellular systems [81] that require cell-type-specific nucleosome positioning [67, 71] and patterns among cells that share DNA sequence. Non-canonical histone variants enriched in regulatory region in the genome, including TSSs and enhancers, may influence the nucleosome turnover rates. Nucleosomes in these regions are less stable and contain post-translational modifications associated with rapid nucleosome turnover, such as histone acetylation and H3K4 methylation. High rates of non-canonical histone variant turnover at promoters and enhancers may initiate chromatin remodeling in human cells [64].

1.6   Accessibility and linker histones

Linker histones and other architectural proteins provide another layer of chromatin regulation playing an important t role in heterochromatin formation and nucleosome positioning [82-84]. Linker histones, including variants of histone H1, change the DNA nucleosome exit angle, neutralising the charge of linker DNA, thus folding the chromatin into a more compact and less accessible. Recently, HMGD1 has been shown in *D. melanogaster* to displace H1 in open chromatin, including active genes and distal regulatory elements. H1 depletion has been associated with transcriptional changes, which may reflect the effect of linker-histone-mediated regulation on transcriptional initiation or elongation [83, 85]. These findings indicate that H1 plays an important role in structural maintenance of closed chromatin.

1.7   Accessibility, nucleosomes, and the 3D genome

Core histone tail acetylation contributes to the decomposition of chromatin fibres [86, 87], whereas long linker DNA length and regular nucleosome spacing are correlated with heterochromatic regions. *In vitro* reconstitution experiments demonstrated that changes in linker DNA length contribute to chromatin fibre reorganisation both in term of nucleosomal packing geometry and response to supercoiling stress [79, 88].

Recent *ex vivo* experiments demonstrated that differences in intra fibre nucleosomal packing across the genome, with more contacts between next-nearest-neighbour nucleosomes in

heterochromatin than in euchromatin [89]. This difference may be due to locally ordered stretches of fibre, such as putative tetranucleosome units [90-92] or clusters of nucleosomes observed by super-resolution microscopy, which are smaller in euchromatin and larger in heterochromatin [93]. Using electron microscopy, directly visualised chromatin and showed that higher-order chromatin organisation does not fold into 30 nm diameter chromatin fibres. In living cells accessibility is determined by chromatin concentration rather than its geometry [94]. However, further work is required to resolve whether conformational features are important for regulating chromatin accessibility.

1.8 Measuring chromatin accessibility

Chromatin accessibility is measured by quantifying the susceptibility of chromatin to either cleavage of its constituent DNA or enzymatic methylation (**Figure 3**) [95]. A study by Hewish *et al.*, [96] used DNA endonucleases to fragment chromatin, showing that nucleosomes confer periodic hypersensitivity across the genome. This periodicity was probed with Southern blot hybridisation, showing a canonical 100–200 bp phasing pattern among DNase hypersensitivity sites (DHSs) that is conserved across genomic loci. This and subsequent work [97] provided direct evidence for stereotypical nucleosome phasing. Similar techniques were used to link chromatin remodeling with contemporaneous transcriptional activation of the heat shock locus in *Drosophila melanogaster* [98]. A variety of quantitative methods have been developed to measure site-specific chromatin accessibility using endonucleases and ligation-mediated polymerase chain reaction (PCR) [99, 100].
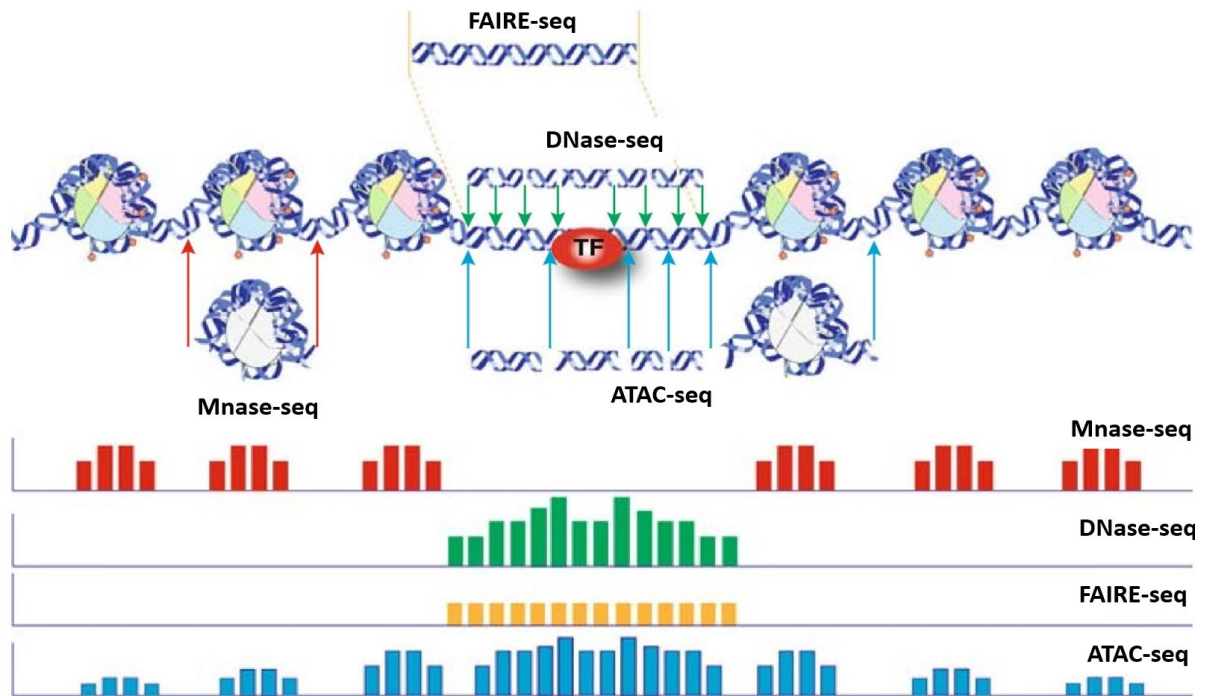
Figure 3. Chromatin accessibility assays. Chromatin accessibility assays performed with typical experimental conditions. Colored arrows represent representative DNA fragments with end locations within chromatin. Bar diagrams represent data signal obtained from each assay. The footprint created by TF is shown for ATC-seq and DNase-seq (adapted from Tsompana and Buck, 2014).

1.8.1 Micrococcal nuclease digestion with deep sequencing (MNaseq-seq)

MNase-seq has been used to study chromatin structure in a low-throughput manner [101]. Currently, MNase-seq digestion is used for genome-wide characterisation of average nucleosome occupancy and positioning in a qualitative and quantitative manner. Mononucleoosomes are extracted by extensive treatment of chromatin that is crosslinked with formaldehyde [102]. The nucleosomal population is submitted to single-end or paired-end next-generation sequencing (NGS) with different level of coverage depending of the goal of the experiment [103].

MNase-seq probes chromatin accessibility indirectly by unveiling areas of the genome occupied by nucleosomes. MNase-seq can provide information on TF occupancy. MNase-seq has been implemented in several organisms ranging from yeast to humans for mapping chromatin structure [75, 104, 105]. In addition, MNase-seq digestion has been successfully combined with Chip-seq for enrichment of regulatory factors or histone-tail modifications and variants. MNaseq-seq allows the mapping of both nucleosomes and non-histone proteins with high resolution.

Important considerations in the design of MNaseq-seq experiments include the level of digestion and the extent of chromatin crosslinking. In the absence of crosslinking, nucleosome organisation can change during regular chromatin preparation steps. Thus, the use of formaldehyde can accurately characterise chromatin structure [103]. MNase-seq has a high degree of AT-cleavage specificity in limiting enzyme concentrations and comparison between different experiments might vary for technical reasons. MNase-seq titration experiments support differential digestion susceptibility of certain nucleosome classes, with nucleosomes within promoter and nucleosome-free regions [106]. Thus, the combination of templates from different levels of MNase-seq digestion may alleviate biased sampling of mononucleosome populations [106].

Due to the effect of inter-nucleosomal linker length on the recovered signal the cause of differences in MNase-seq output across differential levels of enzymatic digestion is difficult to assess [107]. MNase-seq digestion simulation experiments demonstrated that nucleosome configurations with or near long linkers are sampled easier compared to nucleosomes with normal linkers. Comparison of *in vivo* experimental data of two distinct nucleosome configurations from different MNase-seq technical preparations highlights the importance of standardised collection of mononucleosomes for accurate and reproducible comparisons [107]. Particularly, extensive digestion of a standardised initial amount of crosslinked chromatin is considered ideal for comparisons of different MNase-seq experiments [103, 107].

Overall, MNaseq-seq is a method for probing chromatin structure and assessing TF occupancy in a range of cell types. However, it requires a large number of cells and careful enzymatic titrations accurate and reproducible evaluation of differential substrates.

1.8.2 DNase I hypersensitive sites sequencing (DNase-seq)

DNaseI probes accessible chromatin by cutting within nucleosome-free genomic regions characterised as DHSs. Active genes are susceptible to digestion with DNaseI. Chromatin structure is disrupted during gene activation and DHSs are the primary sites of active chromatin rendering access of trans-factors to regulatory elements [98, 108-110]. DHSs result during gene activation due to loss or temporal destabilisation of one or more nucleosomes from cis-

regulatory elements with the action of ATP-dependent nucleosome- and histone-remodelers [111, 112].

The identification of DHSs was originally based on Southern blotting with indirect end-labeling [109]. Low-throughput sequencing, real-time PCR strategies and later hybridisation to tiled microarrays have improved the efficacy and resolution of this method [113-115]. DNase-seq allows the identification of DHSs with unparalleled specificity, throughput, and sensitivity. Due to the drop in sequencing costs and quality of data, DNase-seq has become the 'golden standard' for probing chromatin accessibility. During DNase-seq, isolated nuclei are submitted to DNaseI digestion according to the Crawford or Stamatoyannopoulos protocol [116, 117]. In the Crawford protocol, DNaseI digested DNA is embedded into low-melt gel agarose plugs to prevent further shearing. Optimal digestions are selected by agarose pulsed field gel electrophoresis, with an optimal smear range from 1 MB to 20 to 100 KB and are blunt-end ligated to a biotinylated linker. After secondary enzymatic digestion with MmeI, ligation of a second biotinylated linker and library amplification, the digested population is assayed using NGS [75]. In the Stamatoyannopoulos protocol, DNA is digested with DNase I concentrations and assessed by q-PCR and/or agarose gel electrophoresis. Optimal digestions are purified with size selection of fragments smaller than 500 bp using sucrose gradients, and are submitted for high-throughput sequencing after library construction [117]. The main difference is that the Crawford protocol depends on the single enzymatic cleavage of chromatin, while the Stamatoyannopoulos protocol requires double cleavage events in close proximity to each other.

DNase-seq has been extensively used to probe cell-specific chromatin accessibility and its relation to gene expression in various cell lines. In addition, binding of sequence-specific regulatory factors within DHSs can affect the intensity of DNaseI cleavage and generate footprints (digital genomic footprinting (DGF) or DNaseI footprinting) that have been used to study TF occupancy at nucleotide resolution in a qualitative and quantitative manner [118]. DGF had been used to study cell-specific TF binding motifs in humans, which provides information about the role of TF binding in relation to chromatin structure, gene expression, and cellular differentiation [119, 120]. DGF has allowed the probing of functional allele-specific signatures with DHSs [120].

DNaseI can introduce cleavage bias, thus affecting its use as a reliable TF footprint detection assay. This cleavage bias attributed to protein protection of underlying nucleotides, even in the absence of TF binding [61, 121]. This observation is supported by the lack of correspondence between TF binding events detected with ChiP-seq compared to DGF [61]. Also, TFs with transient DNA binding times leave minimal to no detectable footprints, making the quality of footprinting highly factor dependent. These findings challenge DGF research of TF footprinting as a reliable assay of complex factor-chromatin interactions in a dynamic timescale [61, 121].

DNase-seq involves many sample preparations and enzyme titration steps. DNaseI concentrations need to be adjusted depending on the cell-type and cell number. Overall, DNase-seq is a reliable and robust method to identify active regulatory elements across the genome.

1.8.3   Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq)

FAIRE-seq is a quicker method for directly probing nucleosome-depleted areas of a genome [122, 123]. FAIRE-seq is based on the phenol-chloroform separation of nucleosome-bound and accessible areas of a genome in the interphase and aqueous phase respectively. FAIRE-seq involves crosslinking of chromatin with formaldehyde to capture *in vivo* protein-DNA interactions and subsequent shearing of chromatin with sonication. Nucleosome-depleted areas are released to the aqueous phase of the solution due to much higher crosslinking efficiency of histones to DNA, compared to other regulatory factors [124]. The fragments can be detected by quantitative PCR, tiling DNA microarrays [122], and single-end or paired-end NGS [125].

FAIRE-seq has been performed in a wide range of eukaryotic cells and tissues, demonstrating a negative relationship with nucleosome occupancy and an overlap with various cell type-specific marks of active chromatin [122, 126, 127]. FAIRE-seq has been used to identify active regulatory elements in a number of ENCODE human cell lines [52]. It has been used widely to detect accessible chromatin in normal and diseased cells [122, 125, 128], to associate specific chromatin states with known sequence variants of disease susceptibility [125] or

allele-specific signatures, and to decipher the effects of TF binding to chromatin structure [129].

Overall, FAIRE-seq can be applied to any type of cells or tissue and there is no requirement for initial preparation of cells and enzyme titrations [122]. FAIRE-seq has been shown to identify additional distal regulatory elements, although it remains unclear what these sites represent. In addition, FAIRE-seq overcomes the sequence-specific cleavage bias observed with MNase and DNaseI, and thus represents an ancillary approach for these assays [130-132].

FAIRE-seq depends on adequate fixation efficiency that can alter depending on cell permeability, composition, and a variety of other physiological factors. Also, FAIRE-seq has lower resolution in identifying open-chromatin at promoters of highly expressed genes compared to DNase-seq. One of the major limitations is that it has a lower signal-to-noise ratio compared to the other chromatin accessibility assays. This high background makes computational data interpretation difficult, with only stronger signals being informative.

1.8.4   Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)

ATAC-seq is a method for probing chromatin accessibility using hyperactive Tn5 transposase to interrogate active regulatory regions at a genome-wide level. DNA transposon transfers DNA sequence from one region of chromosome to another, which is assisted by DNA transposase. DNA transposon requires open chromatin and the transposase carries DNA sequence tags that are added to the nucleus. Open chromatin can be identified using labels of known sequences to construct a library for sequencing [95]. The most commonly used transposase is the Tn5 transposase, which acts as a probe for measuring chromatin accessibility at the genome-wide level through the "cut and paste" mechanism. DNA transposon can simultaneously fragment and tag the unprotected regions of DNA with sequencing adapters [133, 134].

ATAC-seq library preparations consist of three steps: nuclei isolation, transposition and amplification [135]. First, cells are suspended into intact, homogenous single cells, which are subsequently incubated in the lysis buffer to generate crude nuclei. Second, the re-suspended nuclei are incubated in the transposition reaction mix to yield DNA fragments. Finally,

transposed DNA is amplified to generate libraries for sequencing. The reaction of transposable enzyme to the chromatin of the sample is the key step of the ATAC-seq experiment [136].

ATAC-seq data quality control is performed to check if the library concentration reaches the sequencing criteria. Raw reads are collected through sequencing. Filtered reads are obtained by evaluating sequencing quality [137-139]. After removing adapter sequences and low quality reads, high-quality reads ~150 bp are processed for further analysis [140]. As with DNaseI-seq and FAIRE-seq, reads can be mapped to the human reference genome and accessible chromatin regions characterised [141-143]. Detailed analyses can be performed such as determining distribution of reads across the whole genome, determining the distribution of the peak length, distribution of peaks on functional elements of genes, and analysis of differential peaks among samples [144, 145]

1.9   Eukaryotic transcription

The expression of eukaryotic protein-coding genes can be regulated at several steps, including transcription initiation and elongation, mRNA processing, transport, translation, and stability. Most regulation, however, is believed to occur at the level of transcription initiation. Transcription of protein-coding genes is performed by RNA polymerase II. Transcribed genes typically contain two distinct families of cis acting transcriptional regulatory DNA element: a) a core promoter and proximal promoter elements and b) distal regulatory elements, including insulators, silencers, enhancers and locus control regions (LCR) (**Figure 4**) [146].
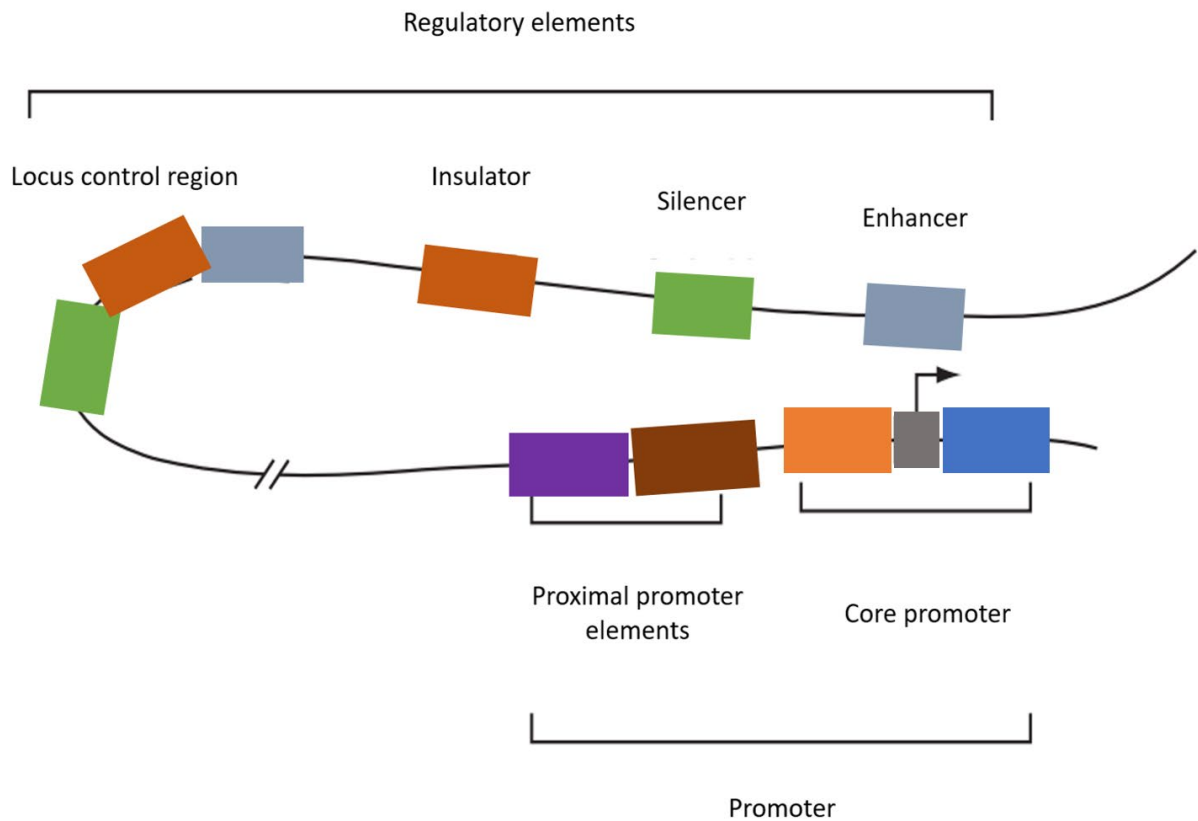
Figure 4. Gene regulatory region. The promoter is composed of a core promoter and proximal promoter elements, which can be located less than 1 kb upstream of the gene. Distal regulatory elements, which include enhancers, silencers, insulators, and LCR, can be located up to 1Mb from the promoter. Distal regulatory elements may contact the core promoter or proximal promoter elements through looping out the intervening DNA (adapted from Maston *et al.,* 2006).

These cis-acting transcriptional regulatory elements contain recognition sites for trans-acting DNA-binding transcription factors, which either repress or enhance transcription. Promoters can be complex, consisting of multiple transcriptional regulatory elements. The need for this complexity becomes clear when one considers that although the human genome contains ~20,000–25,000 genes, each of which may have a unique spatial/temporal expression pattern, it encodes only ~1850 DNA-binding transcription factors. The presence of multiple transcriptional regulatory elements within promoters confers combinatorial control of regulation, which exponentially increases the potential number of unique expression patterns. The challenge now is to understand how different permutations of the same regulatory elements alter gene expression. An understanding of how the combinatorial organisation of a promoter encodes regulatory information first requires an overview of the proteins that constitute the transcriptional machinery [146].

1.10   Transcriptional Regulatory Elements in the Human Genome

1.10.1   Core Promoter

Core promoters are regions at the start of a gene, which serves as the docking site for the basic transcriptional machinery and preinitiation complex (PIC) assembly [147]. Also, core promoters define the position of the TSS as well as the position of transcription. The first described core promoter region was the TATA box: the binding site for the TBP subunit of TFIID. In addition, the TATA box can be composed of numerous other elements, including a downstream promoter element (DPE), an initiator element (Inr), a TFIIB-recognition element (BRE), a downstream core element (DCE) and a motif ten element (MTE) (**Figure 5**) [148] .
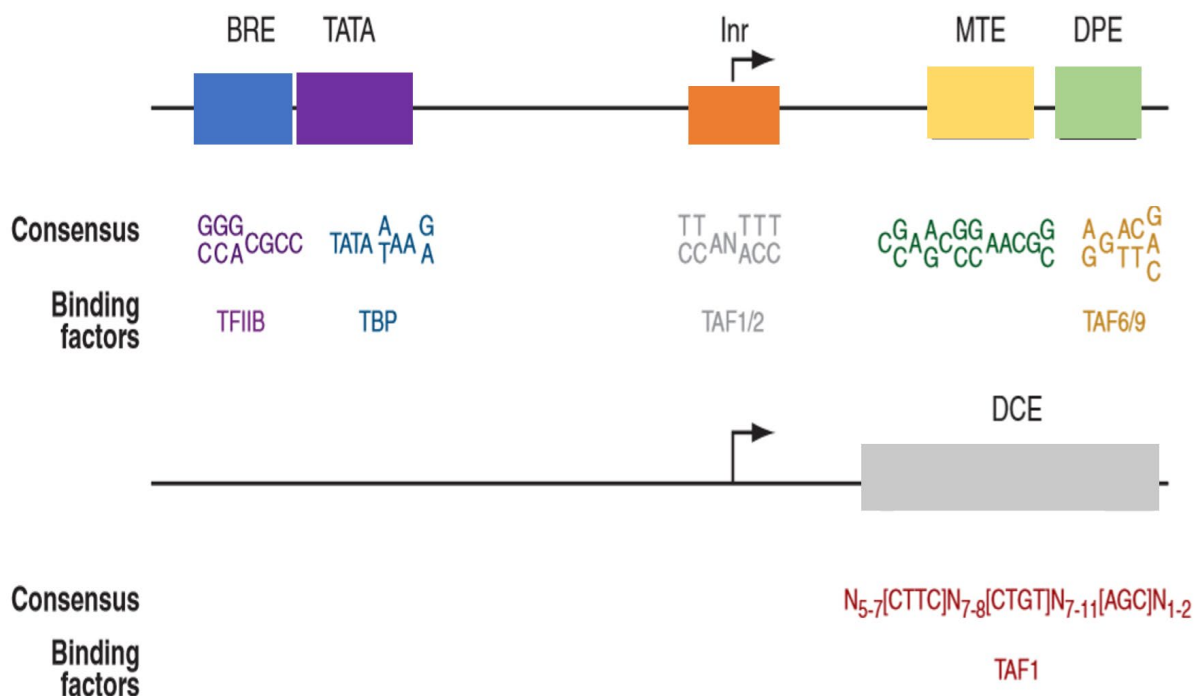
Figure 5. Core promoter elements. Core promoter elements are composed of a number of elements, including BRE, MTE, DPE, DCE, TATA box, and Inr. This figure represents the consensus sequence, the relative position, and the transcription factors that bind to these elements. DCE can be present in promoters containing a TATA box and/or Inr, however it does not occur with a DPE or MTE (adapted from Maston *et al.,* 2006).

Of these four core promoter elements, the Inr is the most common element, occurring in nearly half of all promoters. By contrast, DPE and BRE are found in a quarter of promoters, and the TATA box is found in one eighth of promoters. Nearly a quarter of all promoters have none of these four elements. Recent reports also suggest the existence of ATG deserts [149].

Core promoters are diverse in their context and organisation. First, PIC assembly does not depend on a single interaction point, rather than many of the core promoter elements interact with TFIID. Second, various core promoters may interact preferentially with TFIID complexes having different subunit compositions [150, 151]. Such variation may have functional significance, thus the core promoter can contribute to the regulatory specificity of a gene (**Figure 6**) [147, 152].

1.10.2 Proximal promoter elements

Proximal promoters are regions upstream from the promoter and contain multiple binding sites for activators. Linker-scanning mutagenesis [153] is a technique used to characterise vertebrate promoter elements. This revealed that there are multiple regulatory elements in the regions adjacent to the TSS. This study showed that the synergistic nature of transcriptional regulation is embodied in the promoter structure itself [153].

An interesting feature of ~60% of human genes is that their promoter is located near a CpG island. CpG islands are defined as short stretches of DNA, between 500 bp to 2 kb in length, that have a high G+C nucleotide content and high frequency of the CpG dinucleotide. Many CpG dinucleotides are scattered through the genome, which are methylated at the fifth carbon position. They are associated with many regulated genes as well as many housekeeping genes. The presence of a CpG island is one of the most reliable indicators for predicting the presence of a gene [154]. Interestingly, a correlation exist between the presence of CpG islands and the core promoter: BREs are more common in promoters associated with CpG islands, whereas TATA boxes are more common in promoters that do not have a CpG island nearby [155].

Methylation at CpG dinucleotides block the ability of transcription factors to bind their recognition sequences, this process represses transcription. In addition, MeCP2 binds methylated CpG dinucleotides and recruits histone-modifying complexes that establish a repressive chromatin structure [156]. Proximal promoter elements block the local region from being methylated and thus these regions are inappropriately silenced (**Figure 6**).

1.10.3 Enhancers

Enhancers can act upon distinct enhancer elements at different times, in different tissues or in different stimuli. Enhancers are composed of clusters of TFBSs that enhance transcription. The organisation and orientation of TFBSs within an enhancer can be critical to its regulatory activity [159, 160].

Enhancers are similar to proximal promoter elements. The same activators that bind enhancer elements also bind proximal promoter elements in different genes. Enhancers are long-distance elements that control transcription and can be situated distally from core promoter. They can be situated upstream and/or downstream of a promoter, or beyond the 3' end of the gene [161].

Enhancer and core promoter elements come into close proximity by "looping out" the intervening DNA. Recent studies suggest that DNA-looping may be a general mechanism by which enhancers function [162]. Studies suggest that PIC assembly may begin at a distal enhancer, this would allow more precise control of the regulation of transcription.
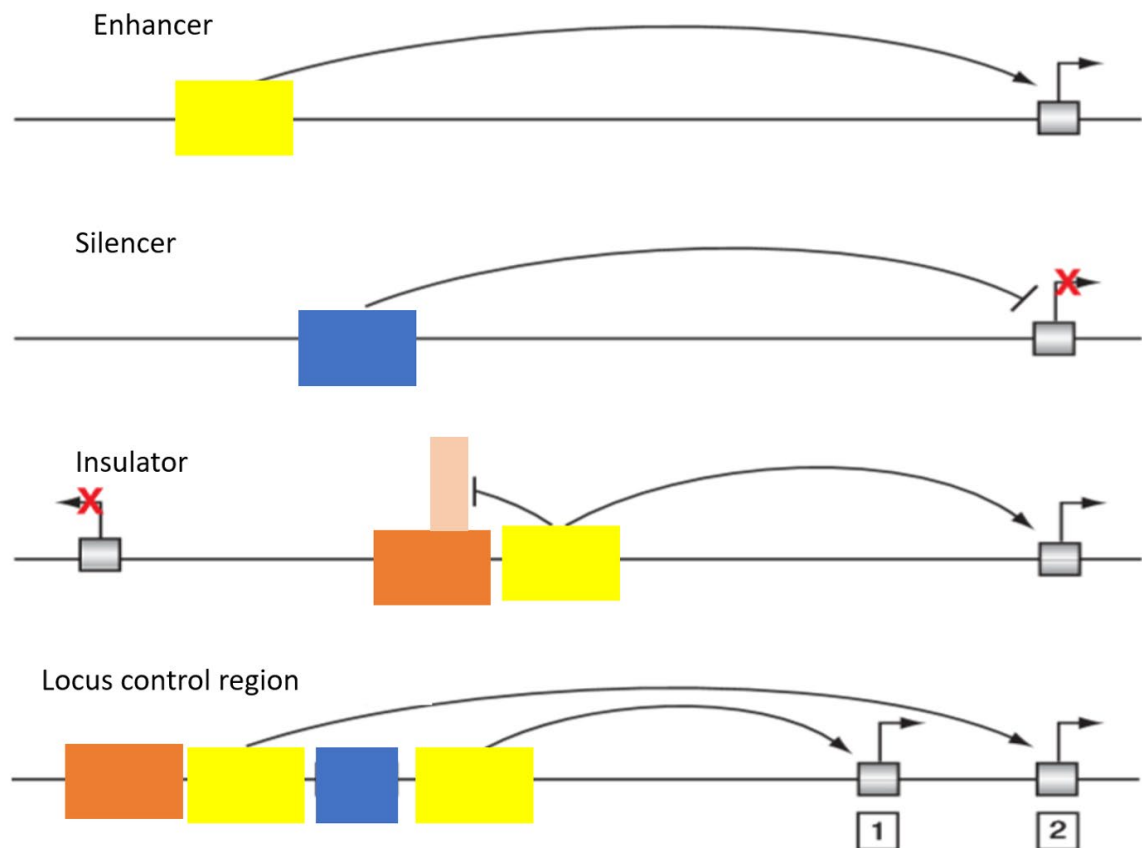
Figure 6. Distal transcriptional regulatory elements. A, B) Enhancers and silencers activate and repress transcription, respectively. C) Insulators block genes from being affected by transcriptional regulatory elements. D) Locus control regions are composed of multiple regulatory elements that confer temporal- and/or spatial-specific gene expression to nearby genes (adapted from Maston *et al.,* 2006).

Gene transcription starts sites with regulatory events at promoters, where TFs bind to cis-regulatory sequences at core promoters immediately upstream of TSSs and promote RNA polymerase II (RNAPII) transcription preinitiation complex assembly, assisted by general TFs and co-activators [157, 158]. RNAPII is paused around the TSSs after RNA of approximately

50 bp has been transcribed and extra signals are required for subsequent RNAPII escape into effective elongation along the gene body. In addition, gene expression depends on the inputs from distal cis-regulatory elements, including insulators and enhancers. Enhancers tend to be active in a cell type specific manner, whereas promoters are ubiquitously used [159-161]. Enhancers can influence different steps of the transcription cycle, from RNAPII recruitment, promoter-proximal pause-release, to transcription elongation, undergoing the importance of these regulatory DNA elements in the control of gene expression (**Figure 7 A and B**) [162].
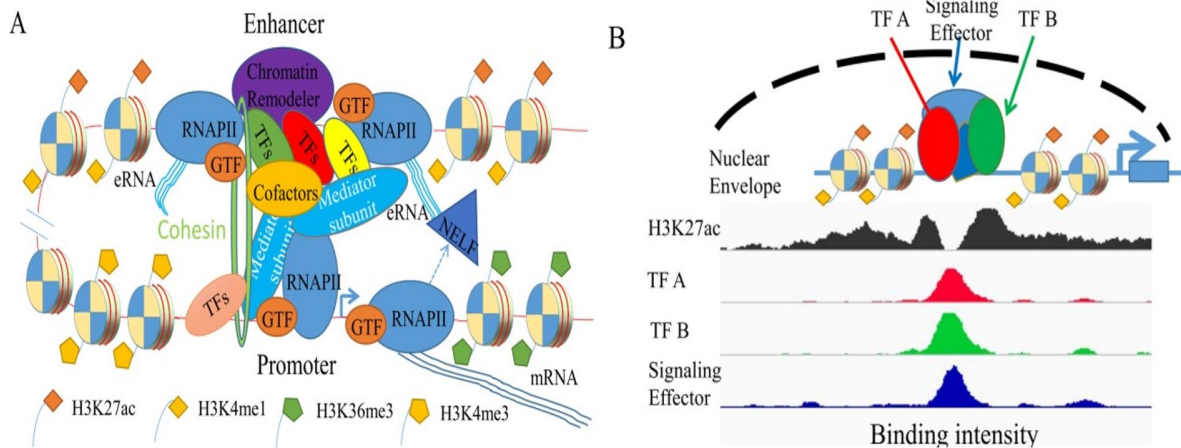
Figure 7. The role of enhancers in gene activation. Multiple factors and processes are involved in the gene activation. A) TFs and many other cofactors, including chromatin regulators and mediator/cohesion complexes mediate cognate enhancer–promoter interaction. Transcribed enhancer RNAs (eRNAs) from active enhancers in turn regulate different stages of transcription, including release of paused RNAPII and enhancer–promoter looping. Gene transcription is associated with distinct chromatin structures, such as the enrichment of histone H3 lysine 36 trimethylation (H3K36me3) at gene bodies, histone H3 lysine 4 trimethylation (H3K4me3) at promoters, and histone H3 lysine 27 acetylation (H3K27ac) and histone H3 lysine 4 monomethylation (H3K4me1) at enhancers. B) Clusters of TFs binding sites at enhancers, including super-enhancers, serve as regulatory hubs to synthesise information from multiple sources of stimuli. TFs, including signalling terminal effectors, associate with each other and bind to (super-) enhancers, which tend to show stronger enhancer activity than typical enhancers (adapted from Hu *et al.,* 2017).

Enhancers play an important role in cell-type specific gene expression through successful and specific communication with their cognate gene promoters. To exert their functions three requirements have to be satisfied: i) enhancers should be accessible, which requires the genome to adopt appropriate local chromatin structure to expose TF binding motifs at enhancers, ii) enhancers and promoters have to interact in a cognate fashion, which requires the genome to adopt appropriate higher order chromatin structure, so that enhancers and target promoters are

placed in physical proximity and importantly, (iii) both local chromatin accessibility and higher order chromatin organisation have to be compatible with different cell types and developmental stages (**Figure 6**) [163].

1.10.4 Silencers

Silencers confer a negative effect on transcription of a target gene. They share most of the properties to enhancers [164]. Typically, they function independently of orientation and distance from the promoter. They can be situated as part of a proximal promoter, distal enhancer, or as an independent regulatory module. Also, silencers can be situated far from their target gene, in its 3'-untranslated region or in its intron. Finally, silencers may play an important role in binding to DNA [165].

Repressors have several functions, including blocking the binding of a nearby activator [165] or competing for the same site [166]. Alternatively, a repressor can establish a repressive chromatin structure through the recruitment of chromatin-stabilising factors or histone modifying activities. Repressors can block transcription activation by inhibiting PIC assembly [167].

For many genes, the transcriptional state is repression, while activation occurs only under specific conditions. Polycomb group response elements (PREs) act as either silencers or antisilencers depending on the protein that is bound and the switch from repression to activation depends on the presence of noncoding transcription across the PRE element. Transcription induces chromatin modifications that prevent access of repressive complexes to DNA (**Figure 6**).

1.10.5 Insulators

Insulators block genes from being affected by the transcriptional activity of neighboring genes. Insulators have two main properties: a) they can block enhancer-promoter communication and b) they can prevent the spread of repressive chromatin. Typically, insulators are ~0.5-3 kb in length and function in a position-dependent, orientation-independent manner [168].

Insulator elements have also emerged as a recurrent feature of a number of imprinted loci in the human genome [169]. The most well-characterised example is imprinting-control region located upstream of the *H19* gene that modulates allele-specific transcription of *H19* and *Igf2* [170]. The number of insulator elements in the human genome is not known and might be found in regions with high density of coding or regulatory information [169].

Insulator activity can be mediated by a number of *trans*-acting factors. In vertebrates, such activity is mediated by CTCF. CTCF play role in many different loci, including mammalian H19/Igf2 ICR [171] and chicken globin 5'HS4. CTCF can be regulated by number of means, including post-transcriptional modifications, DNA methylation and interactions with cofactors [172].

The mechanism(s) by which insulators carry out their enhancer-blocking and/or heterochromatin-barrier activity is not known. There are two categories used to classify insulator function. The first category includes interactions between insulators and transcriptional activators [173]. Enhancer-blocking activity is explained by insulator-bound activator that is unable to interact with its target promoter. Heterochromatin-barrier is explained by the recruitment of histone-modifying activities or gene-activating factors, blocking the spread of repressive chromatin.

The second category associates insulators with chromatin organisation. Insulators interact with each other and/or with a nuclear attachment substrate, thus resulting in the formation of physically isolated chromatin loops. Positioning an insulator between an enhancer and its target promoter results in enhancer-blocking activity due to the physical obstruction between the two elements, preventing their communication (**Figure 6**).

1.10.6 Locus control region

LCRs are groups of regulatory elements, which control a gene cluster or an entire locus. They direct tissue-specific, physiological expression of a linked transgene in a copy-number or position-independent manner. LCRs are composed of multiple *cis*-acting elements, including insulators, silencers, enhancers, nuclear-matrix, or chromosome scaffold-attachment regions. These regulatory elements are bound by chromatin modifiers, repressors, coactivators, and transcription factors. Each of the components differentially affects gene expression. LCRs are marked by a DNase I hypersensitivity sites and provide an open-chromatin domain for genes to which they are linked.

LCRs have been identified in a broad spectrum of mammalian loci. The first LCR identified was the mammalian β-globin locus, which contains five differentially genes. The β-globin LCR [174] is situated ~6–25 kb upstream of the gene cluster and confers high-level erythrocyte-specific gene expression. The β-globin LCR activity is orientation-dependent [175]. LCRs have been identified in a broad spectrum of mammalian loci. The first LCR identified is

mammalian β-globin locus, which contains five differentially genes. The β-globin LCR [174] is situated ∼6–25 kb upstream of the gene cluster and confers high-level erythrocyte-specific gene expression. The β-globin LCR activity is orientation-dependent [175].

A series of recent studies with β-globin LCR have revealed substantial evidence for a "looping" model, similar to the enhancer looping mechanism [176]. Such long-range physical interactions result in the clustering of sequences into an "active chromatin hub", the formation of which is thought to be crucial for establishing an open-chromatin domain [177]. These long-range interactions are observed when the locus is transcriptionally active, providing substantial evidence that they play a role on gene activation. (**Figure 6**) [178].

1.11 CAD

Cardiovascular disease (CVD) is a group of diseases including coronary artery disease (CAD) and acute coronary syndrome (ACS). CAD is characterised by atherosclerosis, whereas ACS is associated with myocardial infarction (MI) regardless of the presence of CAD. CAD is a major cause of death and disability in the developed countries [179]. CAD mortality has gradually declined over the last decades in western countries but still causes about one-third of all deaths in people older than 35 years of age.

1.11.1 Pathophysiology of CAD

CAD represents the clinical manifestation of atherosclerosis, which is also the underlying process responsible for ischaemic stroke and peripheral vascular disease. Atherosclerosis is characterised by the progressive deposition of lipids and fibrous matrix in the arterial wall (**Figure 8**) [180]. The underlying pathology of CAD incorporates a series of events, including loss of the normal barrier function of the endothelium, lipoprotein abnormalities, recruitment of monocytes and lymphocytes to the artery wall and smooth muscle proliferation.
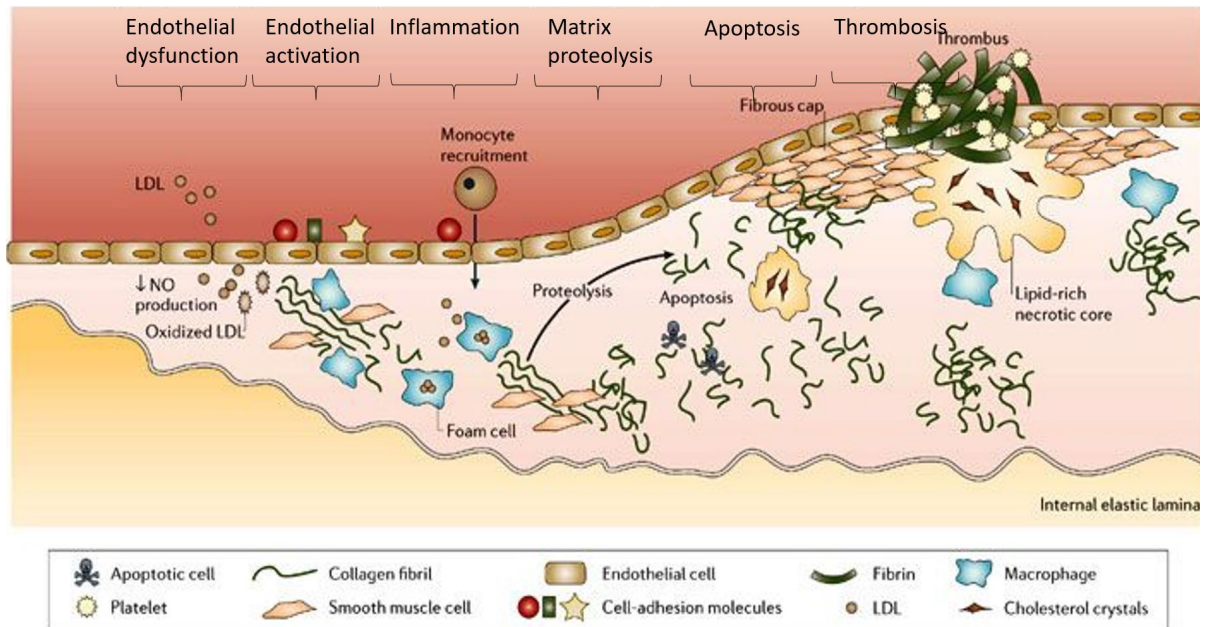
Figure 8. Molecular and cellular processes involved in the atherosclerotic plaque development. Low density lipoprotein (LDL) cholesterol enters dysfunctional endothelium, which is oxidised by smooth muscle cells and macrophages. Release of cytokines and growth factors, and upregulation of adhesion molecules, attracts monocytes. The atherosclerotic plaque development is initiated by the proliferation of foam and smooth muscle cells. This process includes a series of events including inflammatory cells infiltration, smooth muscle cells death through apoptosis and matrix degradation through proteolysis, which lead to progressive deposition of fibrous matrix and lipids in the arterial wall. Thrombosis is caused by plaque rupture, which may lead to vessel occlusion (adapted from Choudhury, 2004).

The clinically important lesions are focal and tend to occur at specific sites (**Figure 9**). Early lesions appear as fatty streaks that later develop into atherosclerotic plaques, which is characterised by a lipid core and an overlaying fibrous cap. A series of complex events influence disease outcome such as plaques can progressively narrow the lumen, which reduces the blood flow or occlude vessels through acute thrombosis leading to MI. Therefore, the clinical entity of CAD results from a particularly complex disease process and extensive genetic heterogeneity among patients [180].
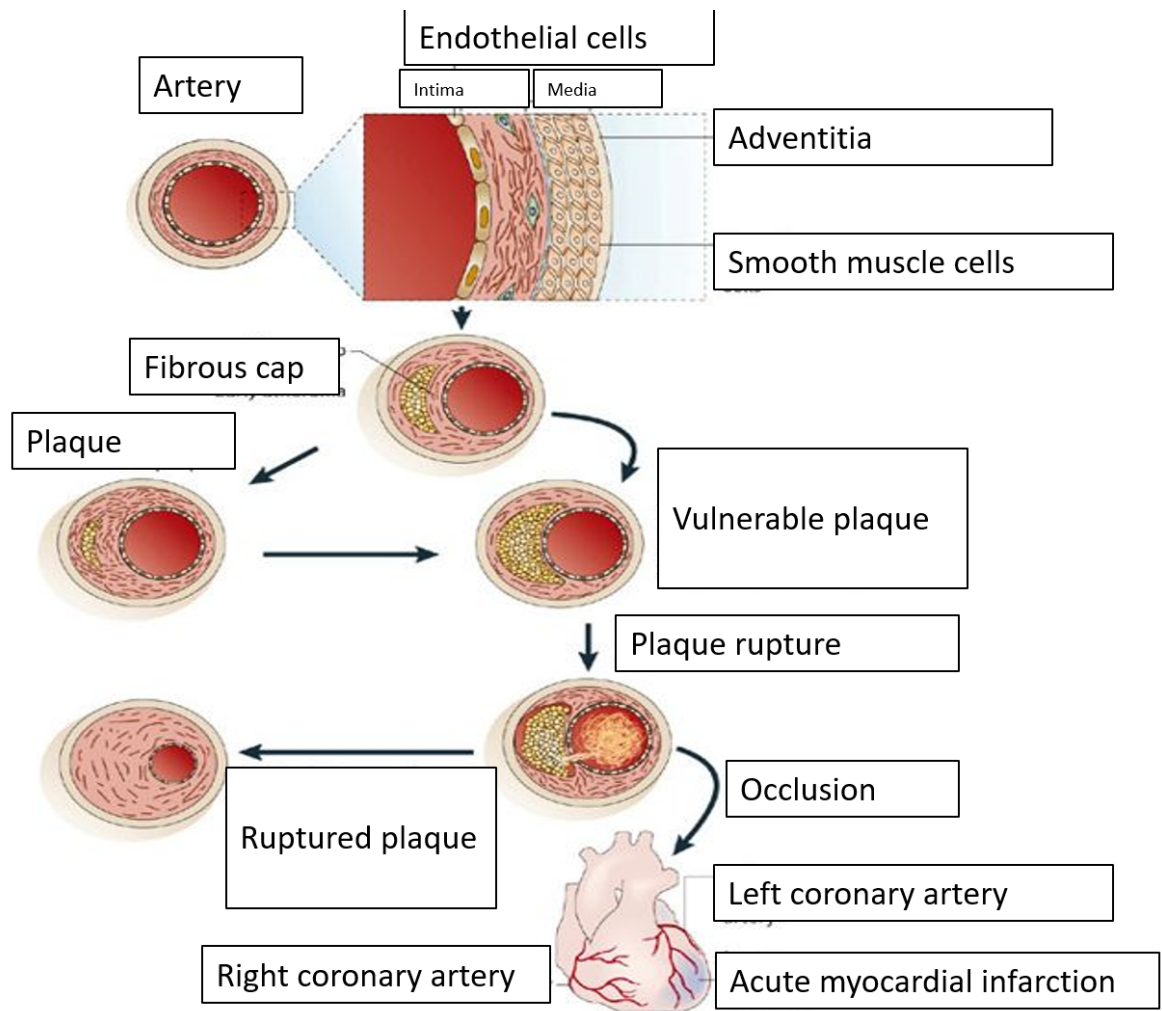
Figure 9. Stages and clinical outcomes in atherosclerosis. Sequential episodes of plaque rupture and healing lead to the progressive development of atheromatous plaques. As a plaque encroaches on the luminal area, blood flow becomes limiting to the heart and this result in chest pain on exertion, known as chronic stable angina. Alternatively, plaque rupture can cause an acute occlusion owing to the thrombus or blood clot formation, known as an acute coronary syndrome. Partial or transient occlusion results in a partial-thickness myocardial infarction with troponin release into the blood, while complete occlusion results in a full-thickness myocardial infarction with scar formation (adapted from Watkins *et al.,* 2006).

1.11.2 Prevalence of CAD

The prevalence of CAD varies between geographical locations, ethnicity, and gender [181]. The fist prospective study of CAD prevalence was conducted in 1946, where the relationships between lifestyle, diet, CAD, and stroke were elucidated [182].

The Monitoring Trends and Determinants in Cardiovascular Diseases (MONICA) project provided information on the incidence of cardiovascular diseases in various populations

over a period of 10 years. This study evaluated CAD trends in approximately 15 million men and woman in the age of 25-64 years. This study reported the survival and CAD event rates [183]. 37 populations in the MONICA Project showed substantial contributions from changes in survival, but the major determinant of decline in CHD mortality is whatever drives changing coronary-event rates.

The INTERHEART study explored the wide array of risk factors associated with MI in different ethnic groups and geographical locations, including 52 countries such as Asia, Africa, Europe, Australia, North and South America and the Middle East [184]. Abnormal lipids, smoking, hypertension, diabetes, abdominal obesity, psychosocial factors, consumption of fruits, vegetables, and alcohol and regular physical activity account for most of the risk of myocardial infarction worldwide in both sexes and at all ages in all regions. These finding suggests that approaches to prevention can be based on similar principles worldwide and have the potential to prevent most premature cases of myocardial infarction.

In the 1960s, "a natural experiment" named as the Nippon-Honolulu-San Francisco (Ni-Hon-San) was conducted in Japanese men aged 45-69 years, living in Hawaii and California and native Japanese men in the same age group, living in Japan [185]. This study proved that changes in lifestyle were associated with acculturation that can explain changes in CAD risk. Higher cholesterol levels and higher CAD mortality rates was observed in Japanese residing in USA compared to Japanese residing in Japan [185, 186]. In 2010, the prevalence of CAD in the USA was found to be higher in people aged over 65 (19.8%), followed by people in the age group of 45-64 years (7.1%) and then those in the age group of 18-44 years (1.2%) [187].

In 2016, American Heart Association reported that in the USA, 15.5 million people above 20 years of age suffer from CVD [188]. CVD prevalence was found to increase with age in both men and women [188]. Lifestyle and environmental factors linked to people's location and habitat are known to impact disease prevalence. For example, in Delhi, during the period from 1984 to 1987, the prevalence of CAD was reported to be 9.7% in urban and 2.7% in rural populations in the age group of 24-64 years of age [189].

1.11.3 Risk factors

Epidemiological studies have identified the most important risk factors associated with CAD to be smoking [190], type 2 diabetes (T2D) [191], hyperlipidemia [192] and hypertension.

Smoking accounts for about 30-40% of deaths related to CAD [190]. Case-control and cohort studies involving 20 million people reported a higher CAD incidence in smokers than nonsmokers. The risk of developing CAD increases with smoke inhalation, number of cigarettes and duration of smoking. Smoking contributes to CAD morbidity and mortality either directly or indirectly. Smoking promotes occlusion as it produces endothelial denudation and platelet adhesion to sub-intimal layers, thereby increasing lipid infiltration and platelet-derived growth factor (PDGF)-mediated proliferation of smooth muscle cells.

Diabetes mellitus or T2D, is another CAD risk factor [193, 194]. Diabetes is often associated with hyperlipidemia, which is characterised by increased levels of triglycerides and decreased levels of high density lipoprotein (HDL) cholesterol [193]. Low levels of HDL cholesterol, high levels of VLDL cholesterol and high levels of total VLDL triglycerides have been reported as risk factors for CAD in patients with type 2 diabetes.

The Prospective Cardiovascular Münster study reported a strong association between hypertension and CAD [195]. Transmural pressure increases in the arteries due to the deposition of lipids and atherosclerotic plaque formation. This results in increased mechanical stress and endothelial permeability, contributing to increased CAD risk [196]. Hypertension is associated with metabolic disorders such as hyperinsulinemia and dyslipidemia, which contribute to CAD risk [197].

In addition to the main CAD risk factors outlined above, additional factors include obesity, height, and levels of uric acid. Obesity involves an excessive amount of body fat [198]. Excess body fat in the abdominal visceral can lead to atherosclerotic disease due to dysregulation of the adipocyte-derived endocrine factors in over-nutrition [199].

Hyperuricemia is defined as excess of serum urate concentration [200], which when presented at a concentration more than 6.8 mg/dl, cause hyperuricemia [200]. Many studies have reported the involvement of uric acid in the stimulation of vascular smooth muscle cell proliferation, reduction in vascular nitric oxide production and insulin resistance. Also,

serum uric acid has been associated with arterial intima-medium thickness, which causes atherosclerosis [201].

Various physiological changes as a result of stress increase the risk of CAD [202]. These changes include reduced insulin sensitivity, elevated blood pressure, endothelial dysfunction and elevated homeostasis [203]. In a case-control study of INTERHEART, job strain was reported to be associated with a higher risk of MI in men then in women. Some studies reported that the risk of job strain is higher is younger employees than in older employees [204].

1.11.4 CAD treatments

1.11.4.1 Treatment using angioplasty and stent placement

Recent advances in the field of therapeutics have suggested novel approaches in the management and treatment of CAD [205]. Coronary angioplasty is procedure used during heart attack to open a blocked artery and reduce the amount of damage to the heart [206]. A tiny balloon is inserted to enlarge the clogged artery. A stent (wire mesh tube) is used to keep the artery open. There are two types of stents including bare metal and drug-eluting stents [207]. This process reduces chest pain and breath shortness.

1.11.4.2 Treatment using recombinant fibroblast growth factor 2 (FGF2)

FGF2 has the capability of stimulating growth and migration of cell types. This factor promotes vascular tree branching. FGF2 triggers signals and increases expression of receptors such as tyrosine kinase and syndecan 4 heparin sulfate core protein. Preclinical studies have demonstrated that FGF2 in chronic ischemia model by ventricular function and augmentation of coronary flow. It has been reported that a single intracoronary infusion of recombinant FGF2 (rFGF2) proved to be tolerated and safe [208].

1.11.4.3 Treatment using antiplatelet agents

Antiplatelet agents including aspirin, sulfinpyrazone and nonsteroidal anti-inflammatory agents are used in CAD treatment [209]. These drugs alter platelet function by inhibiting an enzyme called cycloxygenase [209]. Platelet secret thromboxane A2 protein, which is formed by arachidonic acid metabolism and catalysed by cyclo-oxygenase 1 (COX-1). This COX-1 is activated by aspirin, resulting in reduced platelet function [210]. Another antiplatelet drug, namely thienopyrimidines are biologically inactive prodrugs, which are

metabolised to the active form by cytochromes P450 system in the liver. Thienopyridines could block the P2Y12 receptor present on the platelet surface. When adenosine diphosphate binds to this PYY12 receptor this results in activation of glycoprotein IIb/IIIa receptor, which are involved in platelet aggregation [210]. Glycoprotein IIb/IIIa receptors are repressors for fibrinogen, which are activated on platelet activation. Examples of such blockers include tirobiban, abciximab and eptifibatide.

1.11.4.4 Treatment using other therapeutic agents

β-Blockers can be used for asthma, rhythm management, hypertension, and angina. These therapeutic agents can reduce myocardial oxygen demand and elevate myocardial oxygen supply [211]. It was shown that patients receiving β-blockers showed better prognosis, indicating the benefits of this drugs. Hence, β-blockers therapy improved long-term survival in older patients [212]. Nitrates including nitroglycerin or nitroglycerin sprays have been used for immediate relief of angina [213]. These are recommended as additives if β-Blockers alone are inefficient [214]. Calcium antagonists are used to relief symptoms by decreasing the myocardial oxygen demand and increasing myocardial oxygen supply [211]. Patients with chronic angina or chest pain were treated with ranolazine. Ranozaline alters the transcellular sodium current, decreasing the intracellular calcium resulting in lesser ischemia.

1.12 Genetics of CAD

Family and twin studies [215] have provided evidence that CAD has a heritable component [215]. However, the delineation of the genetic architecture of CAD has been challenging, with the initial candidate gene approaches showing inconsistent results [216]. GWAS assess the association with CAD and hundreds of thousands of genetic variants distributed across the whole genome. Therefore, GWAS represent an essentially unbiased approach that is not limited by the current pathophysiological understanding of CAD.

In 2007, the first GWAS for CAD were performed (**Figure 10**) [217-219]. The main finding was a locus on chromosome 9p21, which harbours the most strongly associated common CAD risk variant to date [220]. Subsequent studies demonstrated that this locus is also implicated in other vascular phenotypes such as abdominal aortic aneurysm [221], peripheral artery disease [221], and coronary artery classification [222]. Large consortia

such as the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) [223] and Coronary Artery Disease (C4D) Genetics [224] have identified through GWAS meta-analyses multiple additional loci associated with CAD and MI. To date, over 160 genetic loci associated with CAD with a genome-wide level (p<5x10^{-8}) have been reported [225]. These loci explain approximately 10% of the heritability of CAD [226, 227]. Heritability estimates based on more relaxed criteria of statistical significance for inclusion of risk variants increase the variance explained to >20%. For example, a study using 304 independent variants at 5% false discovery rate (FDR) reported a heritability estimate of 21.2% [228], however most of the CAD heritability remains unexplained. Interestingly, most of the GWAS loci were not related to traditional risk factors [224, 227] underscoring the potential of GWAS to identify previously unknown genomic loci and biological pathways that contribute to susceptibility [226, 227].

Recent data emerging from theoretical models and empirical observations through GWAS demonstrate that hundreds of genetic loci contribute to complex traits in humans. There are two questions that arise: 1) can additional genetic loci be identified by follow-up of the most significantly associated variants and 2) can genetic fine mapping refine association signals at established genetic loci. Subsequently, addressing these two questions can improve understanding of the genetic architecture of complex traits and their shared genetic determinants.

Addressing these two questions requires genotyping a very large number of individuals. To address these two questions in the context of T2D, CAD and other cardiometabolic traits, the Metabochip [229] was designed as a custom genotyping array that provided accurate and cost-effective genotyping of nearly 200,000 SNPs chosen from relevant GWAS meta-analyses. The success of low-cost custom arrays such as the Metabochip led to adopting this approach for assessing a comprehensive set of rare likely functional non-synonymous variants for association to complex traits. The Exome Chip led to the identification of low-frequency missense variants in *SVEP1* (p.D2702G; MAF 3.60%; OR 1.14) and *ANGPTL4* (p.E40K; MAF 2.01%; OR 0.86), which encodes angiopoietin-like 4, as well as confirmed low-frequency missense variants in the genes *LPA* and *PCSK9* [230].

In the past few years, improved statistical methods for conducting large-scale GWAS meta-analyses, improved imputation panels and large Biobank projects have accelerated new

findings for CAD. In 2015, the 1000 Genomes Project [231] released whole genome sequence data for the HapMap population panels and a comprehensive catalogue of common human genetic variation data. The Haplotype Reference Consortium (HRC) [232, 233] combined extensive whole-genome sequencing data to facilitate genotype imputation. The HRC improved the concordance between assayed and imputed genotypes for low-frequency variants [232]. UK Biobank (UKBB) has linked electronic health records to 502,712 participants with genome-wide imputed data.

At the start of this study in 2017, studies using the interim UKBB data release, previous CARDIoGRAMplusC4D summary statistics and other smaller data sets reported collectively 15 new loci associated with CAD [228, 234, 235]. More recently, an additional 64 novel CAD risk loci were reported by analysing of the full UKBB data set (34,541 CAD cases and 261,984 controls) with previously reported data [236]. Altogether, 163 loci associated with CAD have been reported by the above studies (**Figure 11**). Biobank Japan has very recently reported additional CAD loci. Most CAD risk variants have been discovered under additive model of inheritance however, for 25% of all CAD risk variants it seems that a dominant (26 SNPs) or a recessive model (12 SNPs) might offer a better fit [236]. A Runs of Homozygosity (ROH) study by Christofidou *et al.,* [237], reported an excess of homozygosity in CAD and suggested an important role of ROH in the pathogenesis of atherosclerosis.
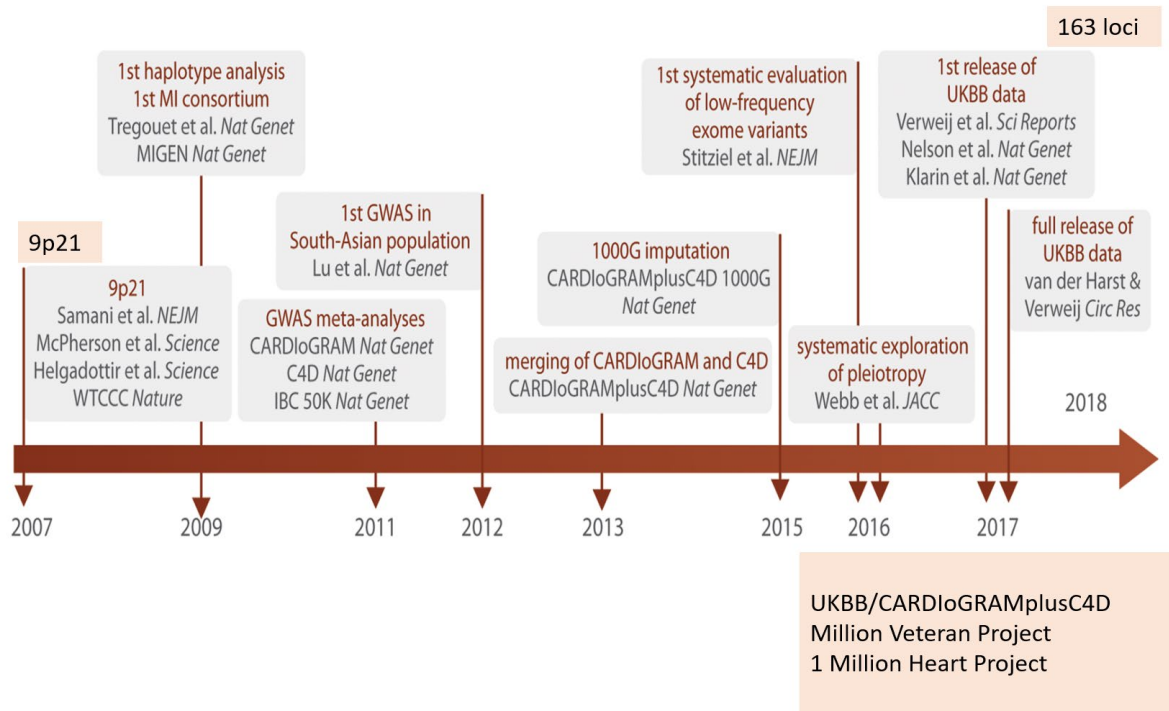
Figure 10. Cardiovascular genome research from 2007 to 2017 (adapted from Erdmann *et al.,* 2018).
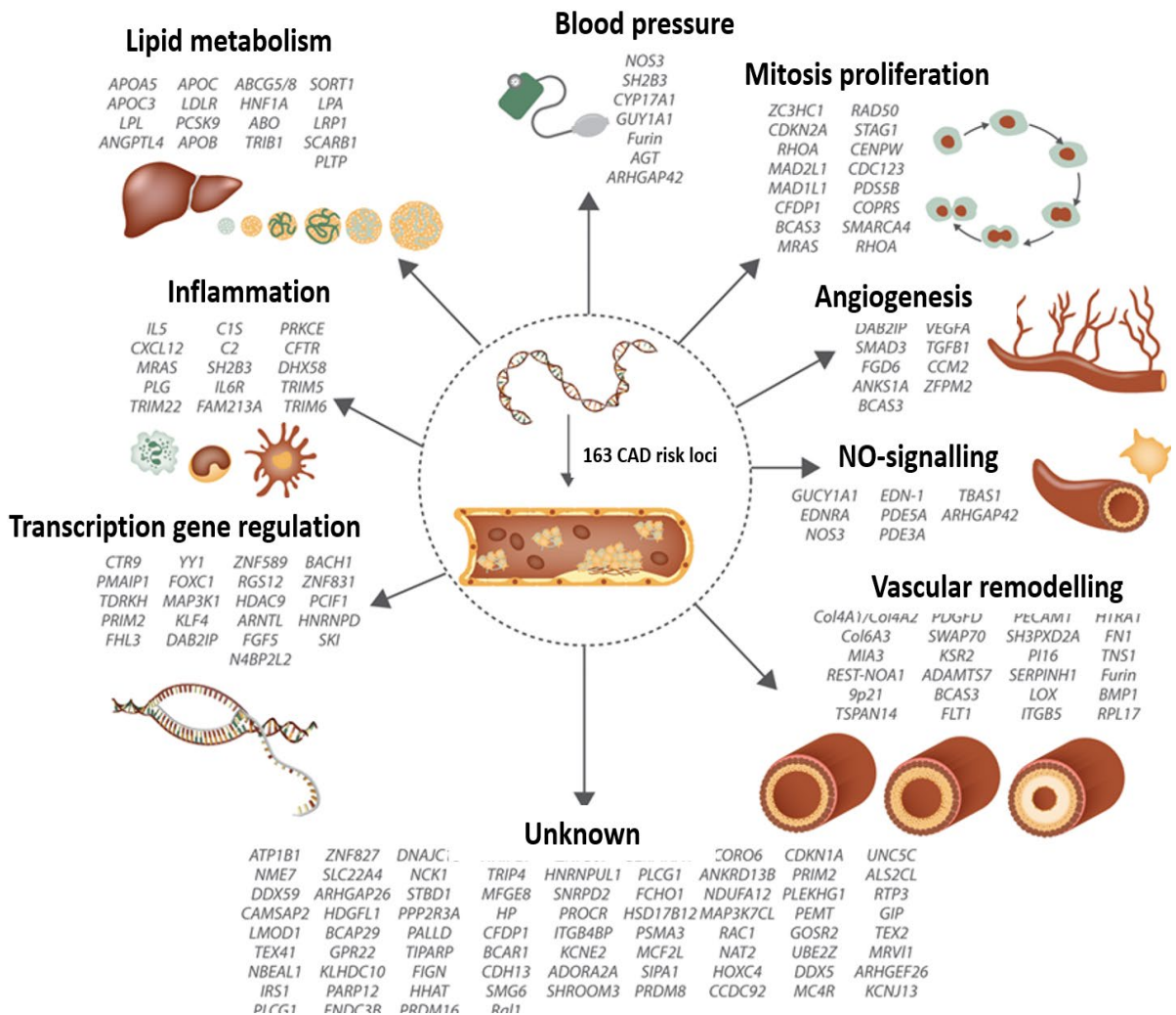
Figure 11. Genes mapped to 163 CAD risk loci and pathophysiological pathways in atherosclerosis (adapted from Erdmann *et al.,* 2018).

The human autosomal genome is covered by modern GWAS arrays and subsequent imputation, however the X- and Y-chromosomes have been neglected due to a number of analytical challenges. The first GWAS for CAD for X-chromosomal variants in more than 43 000 cases and 58 000 controls from 35 international study cohorts showed no genome-wide significant association for any variant [238]. Recently, a study on Y-chromosome showed that genetic variation within the male-specific region confers risk for CAD. This study confirmed that the increased risk for CAD in men cannot be fully explained through common autosomal genetic risk factors.

1.13 UK Biobank

UKBB holds self-reported disease outcomes, extensive health, and lifestyle questionnaire data from 502,713 participants who are tracked through their National Health Service records and national registries. Individuals aged 40-69 years from England, Scotland, and Wales between 2006 and 2010 (94% of self-reported European ancestry).

In July 2015, UKBB released imputed genotypes to the 1000 Genomes for 152,249 participants profiled on a SNP array with 820,967 common variants optimised for imputation, validated rare coding variants, and set of phenotype-associated variants and their proxies. The UK Biobank CardioMetabolic Consortium CHD working group was set up to assess the available self-reported and linked hospital record data to define the relevant case and control subgroups to undertake analyses of CAD risk.

HARD CAD included fatal or nonfatal MI, PTCA or CABG. SOFT CAD included HARD CAD, chronic ischemic heart disease (IHD) and angina. Controls were defined as individuals who were not a SOFT case after exclusions (listed below). All conditions were defined by either self-reported (questionnaire), hospital episode or death registry data. Exclusions were made for aneurysm and atherosclerotic cardiovascular disease using hospital admissions or cause of death codes. Susceptibility effect sizes in MI cases and an inclusive CAD definition were very similar to those in the earlier GWAS [239]. Detailed clinical information in UKBB might enhance the search for new loci by further broadening the CAD phenotype to increase sample size [228].

The July 2015 release of UKBB comprised 10,801 genotyped individuals with an inclusive CAD phenotype (SOFT) that incorporates self-reported angina. Of those 6,482 had more stringent defined CAD phenotype (HARD) that incorporated revascularisation and MI. Cases with the SOFT and HARD phenotypes were analysed separately against 137,914 controls for 9,149,595 variants present either in the CARDIoGRAMplusC4D 1000 Genomes–imputed GWAS [239] or the MIGen/CARDIoGRAM Exome chip study [240, 241]. The SOFT definition of CAD was selected for the primary analysis on the basis of power calculations. This study found four variants associated with both the SOFT and HARD phenotypes, one SOFT-only variant, and two HARD-only variants reaching genome-wide significance, all located in known CAD-associated loci. This study led to the identification of 13 genome-wide significant CAD-associated loci in the combined discovery and replication sample that were

new at the time of analysis. Twelve of the 13 genome-wide significant CAD-associated loci made the 5% FDR threshold [239]. The identified 304 independent variants, clustering in 243 putative CAD loci, associated at 5% FDR, obtained a heritability estimate of 21.2%. The 243 putative CAD loci are implicated in pathways, including blood vessel morphogenesis, development, and angiogenesis [228].

1.14 Vascular pathways

1.14.1 Vascular genes identified in GWAS

Pathway analyses of GWAS-identified CAD loci have highlighted genes involved in lipid metabolism, nitric oxide, and inflammation [227]. Using the DEPICT software, Nelson *et al.,*[228], considered a set of 357-genes (linked to associated variants at 5% FDR) and reported enrichment in canonical pathways related to lipid metabolism, extracellular matrix, inflammation, and nitric oxide production as before, but also angiogenesis and signalling by the proangiogenic growth factor vascular endothelial growth factor A (VEGFA). Blood vessel development, which includes angiogenesis was among the top ten gene sets (**Figure 12**) [228].
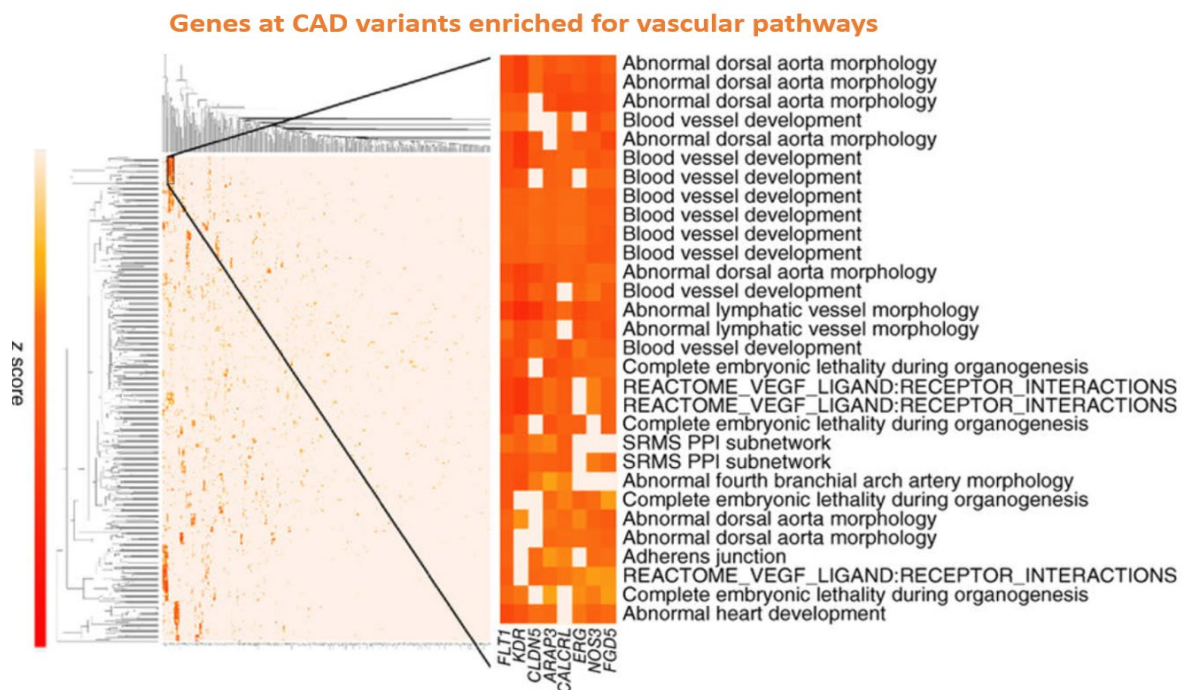
Figure 12. DEPICT gene set enrichment results represented by heat map. 556 gene sets that had evidence of enrichment at 1% FDR. The x axis shows the gene names predicted to be included in the gene sets showing along in the y axis. Higher z scores (shown in red), where z score is a value corresponding to a gene's inclusion in the gene set. Highlighted pathways in the cluster include angiogenesis, blood vessel development and morphogenesis (adapted from Nelson *et al.,* 2017).

1.14.2 Vascular endothelial growth factor (VEGF) and related pathways

VEGFs play an important role in vascular development during blood-vessel formation (angiogenesis) and embryogenesis (vasculogenesis). Five VEGF ligands have been identified in mammals. These ligands bind to three receptor tyrosine kinases (RTKs) including VEGFR1, VEGFR2, and VEGFR3 as well as to co-receptors such as heparan sulphate proteoglycans (HSPGs) and neuropilins.

VEGFs share regulatory mechanisms with other well-characterised RTKs, such as the epidermal growth-factor receptors and the platelet-derived growth factor receptors (PDGFRs). These mechanisms include creation of docking sites for signal transducers, activation of the tyrosine kinase, and receptor dimerisation. The VEGFRs induce cellular processes that are common to many growth-factor receptors, including cell proliferation, survival, and migration. However, VEGFRs are unique due to their ability to regulate vascular permeability that leads to oedema and swelling of tissues and transduce signals that form the three-dimensional vascular tube. VEGFR1 is a positive regulator of macrophages and monocyte migration as well as a positive and negative regulator of VEGFR2 signalling. Negative regulation is exerted by an alternatively spliced soluble VEGFR1 that binds to VEGF and thereby prevents VEGF from bindings to VEGFR2. VEGFR2 is implicated in vascular-endothelial cell biology, whereas VEGFR3 is important for lymphatic-endothelial-cell development [242].

The VEGF family members are dimeric glycoproteins of approximately 40 kDa. In mammals, the VEGF family consists of five members, including VEGFA, VEGFB, VEGFC, VEGFD and placenta growth factor (PLGF). In addition, VEGFE exist in parapoxvirus and VEGFF exists in snake venom. VEGFA, VEGFB and PLGF bind to VEGFR1, VEGFA and VEGFE bind to VEGFR2, and VEGFC and D bind to VEGFR3.
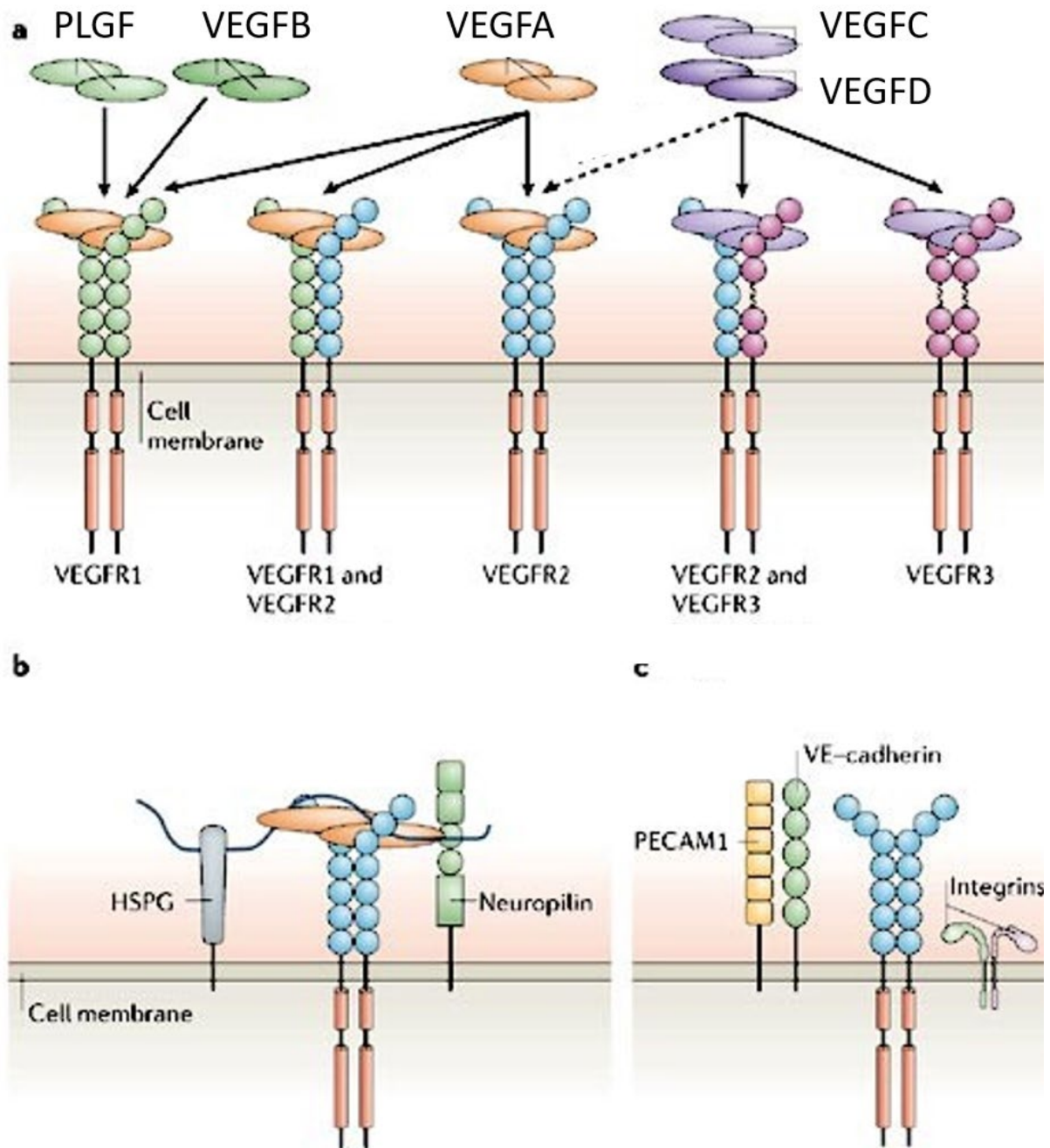
Figure 13. VEGF receptor-binding properties and signalling complexes. A) VEGFs bind to the three VEGF homodimers and heterodimers. Proteolytic processing of VEGFC and VEGFD allows binding to VEGFR2. B) VEGFR signalling is modulated by different co-receptors. VEGFs and VEGFRs bind to co-receptors such as neuropilins and HSPGs. This can influence VEGFR-mediated responses, affecting the half-life of the receptor complex. C) Mechanosensory complex formation. Blood flow activates VEGFRs by the formation mechanosensory complexes that consist of platelet-endothelial-cell adhesion molecule-1, vascular endothelial cadherin, VEGFRs and integrins (adapted from Olsson *et al.,* 2006).

VEGFAs are related to the PDGF family with intrachain and interchain disulfide bonds between eight cysteine residues in conserved positions. The crystal structure of VEGFA revealed two monomers that are organised in an anti-parallel fashion to form a dimer [243], with the receptor-binding sites located at each pole of the dimer (**Figure 13 A, B and C**) [244] .

VEGFA family members give rise to isoforms with different biological activities. The human isoforms are denoted VEGFA121, VEGFA145, VEGFA165, VEGFA189 and VEGFA206. The activities of the VEGFA isoforms are dictated by their abilities to interact with VEGFR co-receptors, such as HSPGs and neuropilins. Another splice variant of VEGFA, known as VEGFA165b negatively regulate VEGFR activity [245].

The activity of the VEGFA family members is regulated by proteolytic processing. This allows specific interactions with different types of receptors. For example, in humans, processed VEGFC and VEGFD bind to VEGFR2 and VEGFR3. Furthermore, proteolytic processing of VEGFA splice variants affects their ability to interact with the VEGF co-receptors HSPGs and neuropilins [246].

The VEGFRs are member of the RTK superfamily and belong to the same subclass as receptors for PDGFs and FGFs. The VEGFRs consists of an approximately 750-amino-acid-residue extracellular domain which is organised into seven immunoglobulin-like folds. In VEGFR3, the fifth Ig domain is replaced by a disulfide bridge. The extracellular domain is followed by a single transmembrane region, a membrane domain, a split tyrosine-kinase domain that is interrupted by a 70-amino-acid kinase insert and a C-terminal tail (**Figure 14**). Structural and functional studies have demonstrated how the distinct domains contribute to VEGFR activity. The crystal structure of the extracellular domain of VEGFR1, alone and in complex with ligand, shows that the Ig domain-2 constitutes the ligand-binding site on the receptor [247]. In addition, biochemical analyses showed that the Ig domain-3 in VEGFR2 is important for the determination of ligand-binding specificity [248]. Alternative splicing or proteolytic processing of VEGFRs give rise to secreted variants of VEGFR1 [249] and VEGFR2 [250], and in humans, to a C-terminal truncated VEGFR3 [251] and VEGFR2 [250].
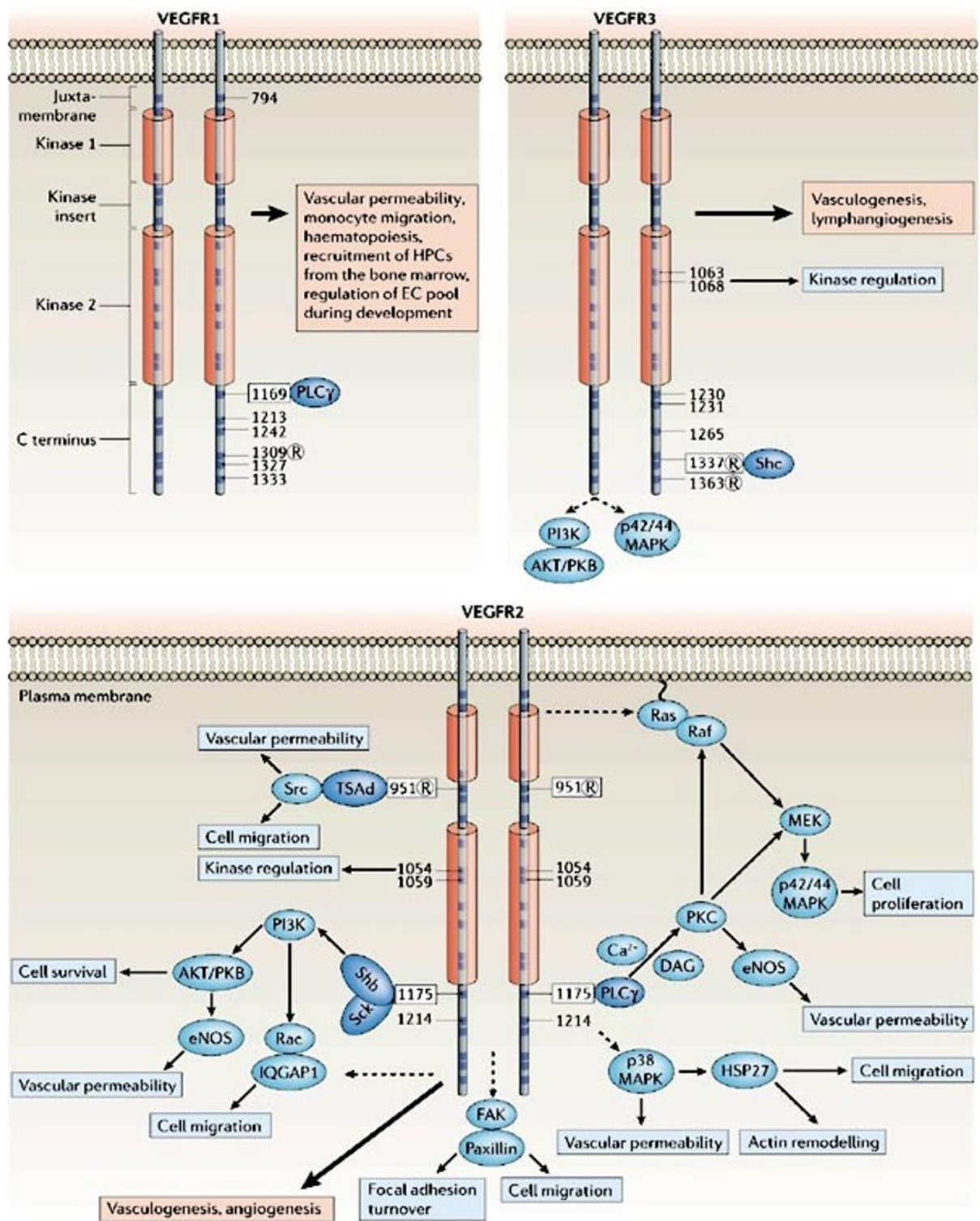
Figure 14. VEGFR phosphorylation sites and signal transduction. Intracellular domains of dimerised and activated VEGFR1, VEGFR2, and VEGFR3 with tyrosine-phosphorylation sites that are indicated by numbers (adapted from Olsson *et al.,* 2006).

VEGFR2, which mediates proliferation of endothelial cells, can be suppressed by VEGFR1 [252, 253] and this effect is dependent on phosphoinositide 3-kinase (PI3K) [253]. There are also reports demonstrating that VEGFR1 might instead promote VEGFR2 activity, for example, during pathological conditions [254]. Activation of VEGFR1 by PLGF results in increased phosphorylation of VEGFR2 [255], possibly through displacement of VEGF that is bound to VEGFR1, thereby making it available for VEGFR2.

VEGFR2 interacts with several SH2-domain-containing molecules. Phosphoinositide specific phospholipase Cγ (PLCγ) binds to phosphorylated Tyr1175 and mediates the activation of the mitogen-activated protein kinase (MAPK)/extracellular-signal-regulated kinase-1/2 (ERK1/2) cascade and proliferation of endothelial cells [256]. PLCγ activates protein kinase C (PKC) by the generation of diacylglycerol and increased concentrations of intracellular calcium. Mice that express a mutated Tyr1173Phe VEGFR2 die because of vascular defects that resemble the defects of $VEGFr2^{-/-}$ mice [257]. These data show an essential function of the Tyr1173 residue during vascular development. PLCγ activates PKC by the generation of diacylglycerol and increased concentrations of intracellular calcium.

The adaptor molecule Shb binds to phosphorylated Tyr1175 [258]. VEGFA-induced migration and activation of PI3K is inhibited by small interfering RNA (siRNA)-mediated knockdown of Shb in endothelial cells [258]. The serine/threonine kinase AKT/PKB mediates survival of the endothelial cells [259]. AKT/PKB also regulates nitric oxide (NO) production by direct phosphorylation and activation of endothelial NO synthase (eNOS). Phosphorylated Tyr1175 has also been shown to interact with Sck/ShcB [260], that plays an important role for normal development [261].

VEGFA stimulates Ras activation in HUVECs [262, 263]. Ras-independent induction of the Raf-MEK-MAPK pathway after VEGFA-stimulation, through PLCγ-activated PKC, has been shown in primary liver sinusoidal endothelial cells [264]. Conflicting results exist in the literature with respect to the interaction of VEGFR2 with ShcA or Grb2, which recruit the Ras-activating nucleotide-exchange factor son of seven to the receptor [260, 265].

Another important phosphorylation site in VEGFR2 is Tyr951, which is a binding site for the signalling adaptor TSAd (T-cell-specific adaptor) [266]. The phosphorylated Tyr951TSAd pathway regulates endothelial-cell migration [266] [267]. Reduced vascularisation and growth of tumours in $Tsad^{-/-}$ mice showed that this is an important

pathway for endothelial cells that are engaged in active angiogenesis [266]. VEGF induces the formation of a complex between TSAd and Src [266], which indicates that TSAd might regulate Src activation and vascular permeability downstream of VEGFR2.

Mice that express mutant Tyr1212Phe (corresponding to the human Tyr1214) VEGFR2 are viable and fertile [257]. Phosphorylation of Tyr1212/1214 has been implicated in VEGF-induced actin remodelling through the sequential activation of cell division control protein 42 homolog (CDC42) and p38 MAPK Mice that express mutant Tyr1212Phe (corresponding to the human Tyr1214) VEGFR2 are viable and fertile [257]. Inhibition of the p38 MAPK augments VEGF-induced angiogenesis [268, 269]. Moreover, p38 MAPK induces phosphorylation of the heat-shock protein-27, a molecular chaperone that positively regulates VEGF-induced actin reorganisation and migration [270, 271].

Focal-adhesion kinase and its substrate paxillin is implicated VEGFA-induced migration through VEGFR2. Ras GTPase-activating-like protein IQGAP1 (IQGAP1), which binds to and activates Rac1 by inhibiting its intrinsic GTPase-activity [272], which regulates endothelial-cell motility. IQGAP1 co-localizes with phosphorylated VEGFR2 at the leading edge of migrating endothelial cells and IQGAP1 knockdown by siRNA prevents VEGF-induced migration [273].

VEGFR3 plays an important role in vasculogenesis. VEGFR3 is regulated by phosphorylation of Tyr1063 and Tyr1068. Phosphorylation of the Tyr1337 is required for association of the Shc Grb2 complex to VEGFR3 [274]. In response to VEGFC VEGFR3 forms homodimers or heterodimers with VEGFR2 [275]. These heterodimeric receptors form both in lymphatic endothelial cells and in certain endothelial cells, which express both receptor types [276]. Importantly, the dimerisation partner directs the use of potential phosphorylation sites, which is a reflection of the different substrate specificities of kinases. Therefore, in the heterodimer, VEGFR3 is not phosphorylated at the two C-terminal tyrosine residues Tyr1337, which is the Shc-binding site, and Tyr1363 [275]. VEGFR3 mediates activation of the ERK1/2 in a PKC-dependent manner and activation of the PI3K/AKT/PKB pathway [277]. These pathways might be important during embryonic development, when VEGFC guides migration and sprouting of lymphendothelial precursor cells from restricted regions of the cardinal vein [278]. VEGFR3 uses other signal transducers including PLCγ, SHP2 [279] STAT3 and STAT5 [280]. VEGFR3 is modulated by neuropilin-2. *Neuropilin-*

$2^{-/-}$ mice fail to form normal lymphatic vessels and capillaries [281]. VEGFR3 is modulated by neuropilin-2. *Neuropilin-2*$^{-/-}$ mice fail to form normal lymphatic vessels and capillaries [281].

1.14.3 Therapeutics that target the vasculature

Therapeutics that target the vasculature may provide potential drug targets for future therapeutics. Vascular targeted therapies can include agents that inhibit angiogenesis and agents that directly interfere with established vasculature [282]. Anti-angiogenic agents and the vascular disrupting agents differ in their mechanism of action and therapeutic application. A combination of these two agents may give the potential of targeted therapy for treatment of vascular diseases including CAD [282].

1.14.4 Prioritisation of functional variation from accessible chromatin

The use of accessible chromatin maps may help identification of important regulatory elements, which will help to localise functional variants. A study by Moyerbrailean *et al.* integrated DNaseI footprinting data with sequence-based TF motif models to predict the impact of genetic variants on TF binding across 153 tissue and 1,372 motifs [283]. This study led to the identification of 5.8 million genetic variants in footprints, 66% are predicted to affect TF binding. This study used allele-specific hypersensitivity (ASH), which revealed that the latter group showed evidence for ASH (3,217 SNPs a 20% FDR), suggesting that 97% of the genetic variants in footprinted regions are silent. This study helped to fine map 86 GWAS SNPs with a 2-fold increase in the posterior odds. Information provided by the tissue-specificity and the identity of TF binding sites can aid in identification of the underlying mechanisms supporting the association [283].

A study by Gaulton *et al.,* [125], examined the map of open chromatin in human pancreatic islets using FAIRE-seq. This study led to identification of ~80,000 open chromatin sites. Approximately 3,300 physically linked clusters of islets-selective open chromatin sites were identified by comparing FAIRE-seq data in islets to non-islet cells lines. This study intersected T2D-associated variants with open chromatin sites, leading to the identification of rs7903146, a Transcription Factor 7 Like 2 (*TCF7L2*) intronic variant associated with T2D in islet-selective open chromatin sites. The variant rs7903146 exhibited allelic imbalance and altered enhancer activity, indicating that variation at this locus acts in cis

with local chromatin. These findings implicate the tissue-specific organisation of cis-regulatory elements.

A study by Bysani *et al* examined the impact of T2D on open chromatin in human pancreatic islets from diabetic and non-diabetic donors using ATAC-seq [284]. This study led to the identification of 57,105 and 53,284 open chromatin regions in islets of non-diabetic and diabetic donors, respectively. The majority of the open chromatin regions mapped near transcription start sites, enriched in enhancer regions and regions with islet-specific transcription factors, including *PDC1*, *NKX6.1*, *NKX2.2*, *MAFB*, and *FOXA2*. This study intersected T2D variants with open chromatin, which led to the identification of 13 variants associated with T2D, including rs7903146, rs2237897, rs757209, rs11708067 and rs878521 near *TCF7L2*, *KCNQ1*, *HNF1B*, *ADCY5* and *GCK*, respectively. Additionally, this study identified 67 variants in LD with known T2D variants near *GIPR*, *KCNJ11*, *GLIS3*, *IGF2BP2*, *FTO* and *PPARG*. There was enrichment of open chromatin near highly expressed genes in human islets. 1,078 open chromatin regions were annotated to 898 genes differing in prevalence between diabetic and non-diabetic islets donors. Several open chromatin regions were annotated to candidate genes for T2D, including *HHEX*, *HMGA2*, *GLIS3*, *MTNR1B* and *PARK2* and some were annotated to genes associated with T2D, including *WFS1* and *ANK1*. Enhancers and motifs specific to key transcription factors, including *PDX1*, *MAFA*, *NEUROD1*, *FOXA2*, *FOXO1,* and *BACH2* were enriched in differential open chromatin regions of diabetic and non-diabetic donors. This study provided insights into how T2D alters the open chromatin landscape in human pancreatic islets [284].

2. Aims

Genetic variation at non-protein coding regions of the genome comprises the majority of association signals in GWAS for complex diseases. In most cases, neither the causative variant at the GWAS locus nor the molecular mechanism(s) is known. Many of the genetic variants are thought to exert their effect on phenotype through altering gene regulatory elements and thereby gene expression levels. Along with established mechanisms such as lipid metabolism and inflammation, vascular morphology may play an important role in the genetic basis of cardiovascular disease and provide novel therapeutic targets.

The aims of this thesis are:

1. To construct open chromatin maps in vascular cells and characterise the genomic regulatory landscape of these cells using the ATAC-seq.

2. To examine the effects of VEGFA on the regulatory landscape.

3. To prioritise coronary artery disease and blood pressure risk variants for functional studies based on open chromatin maps in vascular cells.

3. Materials and methods

3.1 Cell types

Human Carotid Artery Endothelial Cells (HCtAEC, Catalog No. 3014-05a, Lot. No. 2366, Cell Applications, Inc) are primary human endothelial cells isolated from healthy human carotid artery. HCtAEC were cultured in Endothelial Cell Growth Medium and cryopreserved at the third passage.

Human Coronary Artery Endothelial Cells (HCAEC, Catalog No. 300-05a, Lot No, 1522, Cell Applications Inc) are primary human endothelial cells isolated from healthy coronary arteries. HCAEC were cultured in Endothelial Cell Growth Medium and cryopreserved at second passage.

Human Coronary Artery Smooth Muscle Cells (HCASMC, Catalog No. 350-05a, Lot. No. 2139, Cell Applications, Inc) are primary human smooth muscle cells derived from tunica medium of healthy and fibrous plaques free coronary arteries. HCASMC were cultured in Smooth Muscle Cell Growth Medium and cryopreserved at second passage.

3.2 Cell culture

Cryopreserved vials were removed from liquid nitrogen and thawed in a 37°C water bath. HCtAECs, HCAECs, and HCASMCs were resuspended and transferred from the vial into a T-75 flask containing 15 ml of appropriate cell medium. The flasks were placed in a 37ºC, 5% CO2 humidified incubator for 24 hours. The volume of medium was doubled when the culture was 60% confluent. Cells were subcultured when the culture reached 80% confluency. The medium was removed from culture flasks and the monolayer of cells washed with 10 ml of phosphate buffered saline (PBS). 3 ml of trypsin was added to the cells and the flask incubated at 37°C, 5% CO2 for 3 minutes. 5 ml of serum-containing medium was added to the flask to quench the action of the trypsin. The cell suspension was centrifuged at 1000 rpm for 3 minutes to pellet the cells. The supernatant was aspirated without disturbing the cell pellet. The cells were resuspended in 5 ml of medium and transferred into the T-175 flask containing 30 ml of medium.

3.3 VEGFA

15 ml of serum-free medium was added to the flasks. The flasks were placed in a 37ºC, 5% CO2 humidified incubator for 24 hours. 3 µl of 10 µg recombinant VEGFA (Gibco™

PHC9394) was added to the flasks. The flasks were placed in a 37ºC, 5% CO2 humidified incubator for 1 hour.

3.4   Cell counting

Cell counting was performed using the C-Chip (DHC-N01, NAoEnTek). 10 µl of trypan blue solution (SV30084.01) and 10 µl of the cell containing medium was mixed in a 1.5 ml Eppendorf tube. 10 µl of sample was loaded into the chip sample injection area, so that it filled the chamber by capillary action. The cells were visualised and counted using a GXM 1903726 microscope.

Cell per ml = average count per square x dilution factor x volume factor

3.5   Optimised Omni-ATAC transposition reaction

50,000 cells were centrifuged at 10000 rpm at 4°C for 5 minutes. All supernatant was aspirated and 50 µl cold ATAC-Resuspension Buffer (RSB) (containing 0.1% NP40, 0.1% Tween20, and 0.01% Digitonin) was added and mixed by pipetting. The tube was incubated on ice for 3 minutes. 1 ml of cold ATAC-RSB (containing 0.1% Tween-20) was added. The cells were centrifuged at 10000 rpm at 4°C for 10 minutes. The supernatant was aspirated and the cell pellet resuspended in 50 µl of Transposition Mixture (25 µl 2xTD buffer, 2.5 µl transposase, 16.5 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20, 5 µl H2O). The tube was incubated at 37°C for 30 minutes in a thermomixer at 1000 rpm.

3.6   Pre-amplification of transposed fragments

The cell suspension was purified using the Qiagen MinElute PCR Purification Kit (cat. No. 28004, Qiagen) following the manufacturer's instructions. 250 µl of buffer PB was added to the PCR reaction. The column was placed in a 2 ml collection tube. The sample was applied to the column and centrifuged for 1 minute until the entire sample passed through the column. Flow-through was discarded and the column was placed back into the same collection tube. 750 µl of buffer PE was added to the column and centrifuged for 1 minute. Flow-through was discarded and the column was placed back in the same collection tube. The column was centrifuged in a 2 ml collection tube for 1 minute to remove the residual ethanol from buffer PE. Each column was placed in a clean 1.5 ml microcentrifuge tube. 10 µl buffer EB was added to the center of the membrane to elute DNA. The column was left for 1 minute and centrifuged

for 1 minute. Transposed DNA fragments were amplified by combining the following in a 0.2 ml PCR tube: 10 µl Transposed DNA, 10 µl Nuclease Free H2O, 2.5 µl 25 µM Custom Nextera PCR Primer 1, 2.5 µl 25 µM Custom Nextera PCR Primer 2, 25 µl NEBNext High-Fidelity 2x PCR Master Mix (Appendix, **Table 32**). **Table 1** shows the thermal cycles followed for the pre-amplification of transposed fragments.

Table 1. Thermal cycles.

Cycle	Temperature	Time
1 cycle	72°C	5 minutes
	98°C	30 sec
5 cycles	98°C	10 sec
	63°C	30 sec
	72°C	1 minute
Hold	4 °C	∞

3.7 Final amplification and cleanup

The final PCR product was purified using a Zymo DNA Clean and Concentrator-5 Kit (cat. D4014, Zymo Research). 250 µl of DNA Binding buffer was pipetted with 50 µl of DNA sample. Samples were mixed by vortexing. The mixture was transferred to a provided column in a collection tube. Samples were centrifuged for 30 seconds and flow-through was discarded. 200 µl DNA Wash Buffer was pipetted to the column. Samples were centrifuged for 30 seconds, and the wash step was repeated. 10 µl DNA Elution Buffer was pipetted directly to the column matrix and incubated at room temperature for 1 minute. The column was transferred to a 1.5 ml microcentrifuge tube and centrifuged for 30 seconds to elute the DNA.

3.8 Quantify library concentration

The library concentration was quantified using the Qubit™ 1X dsDNA HS Assay Kits (cat. No. Q33230, Invitrogen). 10 µl of each standard and 190 µl of working solution were added to the appropriate standard tube. 1 µl of each sample and 199 µl of working solution were added to the appropriate sample tube. Each sample was mixed by vortexing for 3-5 seconds. All tubes were left to incubate at room temperature for 2 minutes prior to analysis.

3.9 Sequencing

NGS was performed by Bart's and the London Genome Centre at Queen Mary, University of London. The sample requirements were 5 µg DNA at 50 ng/µl. Genome Centre performed all processing steps from receiving DNA through to delivery of quality controlled FASTQ files. The Illumina NGS method is based on sequencing-by-synthesis (SBS), and reversible dye-

terminators that enable the identification of single bases as they are introduced into DNA strands (**Figure 15**).
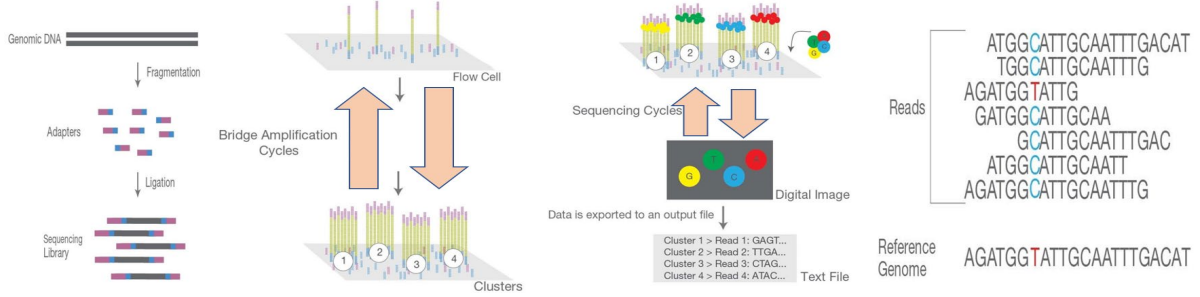
Figure 15. Illumina NGS workflow. The core principle of Illumina NGS contains several steps, including library preparation, cluster generation, sequencing, and alignment and data analysis (adapted from https://www.illuminutesa.com/).

Adapter-ligated ATAC-seq fragments were PCR amplified and gel purified. The DNA fragments randomly attached to the sequencing lanes, which match the adapters at the ends of the DNA fragments. Flow cells adsorbed the DNA and support the amplification of the bridge PCR on the surface of the DNA. The adapters on the flow cell surface were used as a template. After continuous amplification and mutation cycles, each DNA fragment was clustered in bundles. Once cluster generation was complete, these templates were ready for sequencing. The sequencing method was based on sequencing-by-synthesis. DNA polymerase, connector primers and 4 dNTPs with base-specific fluorescent markers were added to the reaction system. All unused free dNTPs and DNA polymerase were eluted. Buffer solution for fluorescence excitation was added, the fluorescence signal excited by laser and recorded by optical equipment. The optical signal was converted into sequencing base by computer. A chemical reagent was added to quench the fluorescence signal and remove the dNTP 3′-OH protective group so that the reaction can be performed (https://www.illumina.com/).

3.10 ATAC-seq data analysis

3.11 Quality control

FastQC [285] was used for the quality control of FASTQ raw reads. The workflow includes checking various parameters, including per base sequence quality, per sequence quality scores, per base sequence content, per sequence GC content, per base N content, sequence length distribution, duplicate sequences, overrepresented sequences, adapter content, kmer content, and per tile sequence quality.

3.12 Trimming

Trimmomatic [286] was used to trim and crop FASTQ files and remove adapters. Paired end mode was used as it maintains correspondence of read pairs and uses the additional information contained in paired end reads to find adapter or PCR primer fragments introduced by the library preparation process. For paired-end, two input and 4 output files were specified, 2 for the paired output where both reads survived the processing, and 2 for corresponding unpaired output where a read survived, but the partner read did not. The following trimming steps were applied to all HCtAEC samples at 1M reads per sample. Cutting Illumina-specific adapter sequences from the read. The maximum number of mismatches to allow was two when searching for the adapters in the sequences with or without the match score. The palindrome clip threshold was 30, which specifies the minimum match score for paired end adapters. The minimum match score was 10 for adapters in single end reads. Bases off the start of a read were cut if below a threshold quality of 3. It scans from the 5' end and clips the read once the average quality within the window falls below a threshold (a sliding widow trimming approach of 10:20 was set), cutting the specified number of bases from the start of the read (12 bases were cut from the start of the read), and dropping the read if it is below a specified length (specified length of 36 was set). The implemented trimming steps to all HCtAEC samples at 30M read per sample were: cutting the read to a specified length by removing bases from the end (74 bases were removed from the end), cutting the specified number of bases from the start of the read (15 bases were cut from the start of the read), performing a sliding window trimming approach. It starts scanning at the 5' end and clips the read once the average quality within the window falls below a threshold (a sliding widow trimming approach of 5:28 was set). The implemented trimming steps to all HCtAEC samples at 50M read per sample were: cutting the read to a specified length by removing bases from the end (60 bases were removed from the end), followed by the same parameters used for 30M reads per sample.

3.13 Read alignment

Bowtie2 [287] was used to align reads to the human reference genome (hg19). Paired-end reads were stored in a pair of files, one file containing the mate 1s and the other containing the mates 2s. The first mate in the file for mate 1 forms a pair with the first mate in the file for mate 2, the second with the second. The file with the mate 1s was specified using the -1 argument and the file with the mate 2s was specified using the -2 argument. Bowtie2 generated a sequence

alignment map (SAM) alignment for a pair, it generated two records, one for each mate. The first record describes the alignment for mate 1 and the second record describes the alignment for mate 2. In both records, some of the fields of the SAM record describe various properties of the alignment. For instance, the 7th and 8th fields indicate the reference name and position where the other mate aligned, and the 9th field indicates the inferred length of the DNA fragment from which the two mates were sequenced.

3.14 Alignment adjustments

Picard tools (https://broadinstitute.github.io/picard/index.ht ml) were used to manipulate high-throughput sequencing data formats including SAM and/or binary version of a SAM file (BAM) files. Picard *SortSam* was used to sort SAM files by coordinate. For a coordinate sorted SAM file, reads alignments were sorted first by the reference sequence name using the reference sequence dictionary. Alignments within these subgroups were secondary sorted using the left-most mapping positions of the read. Subsequently, the alignments were listed arbitrarily. The SAM file was specified using the INPUT argument, the output file was specified using the OUTPUT argument, and the sort order output file was specified using the SORT_ORDER argument. Picard *MarkDuplicates* was used to identify duplicate reads. The input BAM file was specified using the INPUT argument, the output BAM files was specified using the OUTPUT argument, the duplication metrics were specified using the METRICS_FILE argument, duplicates were removed by specifying the REMOVE_DUPLICATES argument. Picard *MergeSamFiles* was used to merge multiple BAM files into a single file. The input BAM file was specified using the INPUT argument and the output BAM was specified using the OUTPUT argument.

3.15 Peak calling

Model-based Analysis for ChIP-Seq (MASC2) [288] was used to identify regions of genomic enrichment. The input read alignments were specified using *-t* argument, the format of the tag file was specified using the *-f BAM* argument, the reference genome was specified using the *-g hs*, the output files were specified using the *-n* argument, the *q*-value cutoff to call significant regions was specified using the *-q 0.05* argument, since the DNA wrapped on nucleosome is 147 bp this was specified using the *--nomodel --shift 37 --extsize 73*.

3.16 Comparing peak files

Bedtools [289] was used to combine overlapping features in an interval file into a single feature, which spans all of the combined features. The maximum distance between features allowed for features to be merged was specified using *-d 100* argument. The columns from the input file to operate upon were specified using *-c 1,2,3,4*. The operation that should be applied to *-c* was specified using *-o distinct, minutes, max, distinct* arguments. Bedtools intersect was used to identify overlaps between two sets of genomic features. The original A and B entries plus the number of base pairs of overlap between the two features, where only A features with overlap are reported was specified using *-wo* argument. Bedtools subtract was used to identify features in B that overlap A by at least the number of base pairs given by the *-f* argument. If an overlapping feature is found in B, the overlapping portion is removed from A and the remaining portion of A is reported. If a feature in B overlaps all of a feature in A, the A feature will not be reported. If a feature in B does not overlap a feature in A by at least the *-f* argument, the A feature will be reported in its entirety.

3.17 Displaying annotations in the University of California, Santa Cruz (UCSC) Genome Browser

To construct an annotation file and display it in the UCSC Genome Browser [290], the following steps were performed: i) the data was formatted as a tab-separated file using a formats, including bigwig (reads) and narrowPeak (peaks); ii) the data were added to configure genome position; iii) a track line was added to define the display attributes for the annotation data set.

3.18 ATAC peak annotation, comparison, and visualisation

ChIPseeker [291] was used to retrieve the nearest genes around a peak, annotate genomic region of the peak, employ statistical methods to estimate the significance of overlap among ATAC-seq peak data sets, and incorporate gene expression omnibus database to compare datasets with those deposited in database. Several visualisation functions were implemented to summarise the coverage of the peak experiment, average profile and heatmap of peaks binding to TSS regions, genomic annotation, distance to TSS and overlap of peaks. To annotate the location of a given peaks in terms of genomic features, *annotatePeak* assigns peaks to genomic annotation in "annotation" column of the output, which includes whether a peak is in the TSS, Exon, 5' UTR, 3' UTR, Intronic or Intergenic. The function *plotAnnoPie(peakAnno)* was used to visualise the genomic annonation. The function *plotAvgProf2* was used to generate average

profile of ATAC-seq peaks binding to TSS region. The function *peakHeatmap* was used to generate heatmap of ATAC-seq peaks binding to TSS regions. The function *compareCluster* was used for comparing biological themes among gene clusters.

3.19 Motif discovery

The TF binding analysis was performed using Hypergeometric Optimization of Motif EnRichment (HOMER) v3.1 considering the known motifs. The following ATAC-seq datasets were used: (i) HCAEC ATAC-seq peaks; ii) VEGFA stimulated HCAEC ATAC-seq peaks and iii) HCASMC ATAC-seq peaks. The TF binding motif enrichment was performed using the *findMotifsGenome.pl* script and the top known motifs, based on *p*-value less than 0.05, were reported. HOMER goes through series of steps to find quality motifs [292].

1. Verifies peak/BED file.
2. Extracts sequences from the genome corresponding to the regions in the input file, filtering sequences that are >70% "N".
3. Calculates GC/CpG content of peak sequences.
4. Prepares the genomic sequences of the selected size to serve as background sequences.
5. Randomly select background regions for motif discovery. Atomizes sequence bias.
6. Checks enrichment of known motifs.
7. Performs *de novo* motif fining.

3.20 Correlation analysis

multiBamSummary [293] was used to compute the read coverages for genomic regions for two or more BAM files. The analysis was performed for the entire genome by running 'bins' mode. The output of multiBamSummary is a compressed numpy array (.nzp), which was used to calculate and visualise the pairwise correlation values between the read coverages using 'plotCorrelation'. This tools analyses and visualizes sample correlations based on the output of multiBamSummary. Pearson method is available to compute correlation coefficients. Results were saved as multiple scatter plots depicting the pairwise correlations or as a clustered heatmap, where the colors represent the correlation coefficients, and the clusters are constructed using complete linkage.

3.21 Calculation of the fraction of reads in peaks (FRiP) score

The BEDTools [289] intersect function was used to identify all reads in peaks. The analysis was performed by overlapping reads in peaks. The FRiP score was calculated by dividing reads in peaks by total reads used for calling peaks.

3.22 Overlapping ATAC-seq peaks with ENCODE Open Chromatin by DNaseI HS and FAIRE

Overlap was determined using the BEDTools [289] package's intersect function with default arguments (minimum 1 bp overlap). The analysis was performed by overlapping ATAC-seq peaks with HUVEC Open Chromatin by DNaseI HS from ENCODE/OpenChrom (Duke University) and FAIRE from ENCODE/OpenChrom (UNC Chapel Hill).

3.23 ATAC-seq overlap with Chromatin State Segmentation by HMM from ENCODE/Broad

Overlap was determined using the BEDTools [289] package's intersect function with default arguments (minimum 1 bp overlap). The analysis was performed by overlapping ATAC-seq peaks with HUVEC Chromatin State Segmentation by HMM from ENCODE/Broad [294] [295].

3.24 ATAC-seq overlap with Histone Modifications by ChIP-seq from ENCODE/Broad Institute

Overlap was determined using the BEDTools [289] package's intersect function with default arguments (minimum 1 bp overlap). The analysis was performed by overlapping ATAC-seq peaks with HUVEC Histone Modifications by ChIP-seq from ENCODE/Broad Institute [296] [295, 297-299].

3.25 RNA purification

RNA-extraction was performed using RNeasy Mini Kit (cat. nos. 74104) as described in the manual. Cells were harvested as described previously and 600 µl of buffer RLT was added. 70% ethanol was added to the lysate and mixed. 700 µl of sample was transferred to an RNeasy Mini spin column placed in a 2 ml collection tube. The sample was centrifuged for 15 seconds at 8000xg and the flow through was discarded. 700 µl Buffer RW1 to the RNeasy spin column. The sample was centrifuged for 15 seconds at 8000xg and the flow through was discarded. 500 µl Buffer RPE was added to the RNeasy spin column. The sample was centrifuged at 8000xg and the flow through discarded. 500 µl Buffer RPE was added to the RNeasy spin column. The

sample was centrifuged for 2 minutes at 8000xg. The RNeasy spin column was placed in a new 1.5 ml collection tube. 30-50 µl RNase-free water was added directly to the spin column membrane. The sample was centrifuged for 1 minute at 8000xg to elute the RNA. The RNA was stored at -80°C.

3.26 Quality control

FastQC [285] was used for the quality control of FASTQ raw reads. The workflow includes checking various parameters, including per base sequence quality, per sequence quality scores, per base sequence content, per sequence GC content, per base N content, sequence length distribution, duplicate sequences, overrepresented sequences, adapter content, kmer content, and per tile sequence quality.

3.27 Alignment

Spliced Transcripts Alignment to a Reference (STAR) was used to map the trimmed reads were mapped to the human reference human genome (GRCh37.85) [300]. STAR allows a stepwise approach for the reads alignment. STAR command-line options used, include *runMode*, which indicates genome generate mode, *genomeDir*, which shows the path to store genome indices, *readFilesIn*, which shows the path to FASTQ files, and *runThreadN*, which shows number of threads.

3.28 Counting reads in features

Htseq-count [301] was used to count the reads overlapping with known genes. Htseq-count command-line options used, include *--stranded=no*, which for stranded=no, a read is considered overlapping with a feature regardless of whether it is mapped to the same or the opposite strand as the feature, *--mode=intersection-nonempty*, which indicates mode to handle reads overlapping more than one feature, *-r pos*, which for paired-end data, the alignment have to be sorted either by read name or by alignment position, this indicates how the input data has been sorted.

3.29 Pathway enrichment analysis

Reactome pathway database [302] was used to perform pathways analysis of the top up regulated genes upon VEGFA-stimulation in HCAEC from the RNA-seq data. The "analysis tool" was used from the Reactome pathway database. This tool merged pathway identifier

mapping, over-representation, and expression analysis The first step was selecting the set of 148 (1.5-fold) up-regulated genes upon VEGFA-stimulation in HCAEC. The second step was selecting the "Project to human", which includes all non-human identifiers that are converted to their human equivalents. The last step was the analysis of enriched pathways.

3.30 CAD- and BP- associated variants within ATAC-seq peaks

The unstimulated and VEGFA-stimulated HCAEC, and HCASMC ATAC-seq peaks were intersected with the CAD association data [303] that includes 181,522 CAD cases UKBiobank/CARDIoGRAMplusC4D and additional studies, 897 1% FDR CAD risk variants and proxies (N = 21,461; r^2 >0.8). Similar analysis was performed for the unstimulated and VEGFA-stimulated HCAEC, and HCASMC ATAC-seq peaks, which were intersected with the BP association data [304] that includes > 1M individuals, 901 BP risk variants and proxies (N = 14,168; r^2 >0.8).

3.31 Prioritisation of CAD- and BP-associated variants

RegulomeDB (https://regulomedb.org/): includes data from resources such as ENCODE and the Roadmap Epigenomics Project for annotating variants within regulatory elements [305], including different regulatory features: TFBSs, chromatin states and eQTL. This information is combined to calculate a score for variant prioritization. RegulomeDB was used to annotate the CAD- and BP-associated variants based on their regulatory function using classification scheme [306].

HaploReg (https://pubs.broadinstitute.org/): combines information from ENCODE and the RoadMap Epigenomics Project [307] for annotating regulatory variants [308]. The tool combines variants into haplotype blocks, so that SNPs correlated with the query SNP can also be examined. This is achieved using 1000 Genomes Project data, which provides sequence data for many individuals from different populations, for calculating the LD between different variants and producing haplotype blocks. Haploreg was used to examine the annotations of the CAD- and BP-associated variants using information, including promoters, enhancers, DNaseI, proteins bound, eQTL tissues, transcription factor motifs changed, GENCODE genes, and dbSNP functional annotations [307].

GTEx (https://gtexportal.org/home/): encompasses a large community of research groups with the shared aim of demonstrating the relationship between variants and human traits or diseases by analysing changes in gene expression [309]. This project determines eQTL by combining

genetic variation with gene expression in post-mortem tissues. It can be used to determine which pathways are affected in disease. They provide an expression atlas for the identification of putative regulatory regions and determining eQTLs associated with disease. GTEx was used to examine whether the CAD- and BP-associated variants are an eQTLs in three vascular tissues, including coronary artery, tibial artery, and aortic artery [309].

3.32 Motif-based sequence analysis

CentriMo [310] was used to identify motifs that show whether the local enrichment is significant relative to control sequences. CenriMo takes a set of motifs and equal-length sequences and plots the positional distribution of the best match of each motif. The motifs are DNA or RNA-biding motifs, and the sequences were 25 bp sequences aligned on CAD- and BP-associated variants. CentriMo reports the relative enrichment in the control set of sequences. CentriMo uses Fisher's exact test to evaluate the significance of the number of best matches in the primary set compared to the control set of sequences. The first step was to extract sequences from a FASTA for each set of CAD and BP SNPs by extracting 25 bp region around the center of each SNP. This analysis was performed using bedtools getfasta, which extracts sequences from a FASTA file. The second step was selecting the primary and control sequence sets. For the CAD association data, I used the 1% FDR CAD SNPs (control sequences) compared to HCAEC SNPs (primary sequences), 1%FDR CAD SNPs (control sequences) compared to HCASCM SNPs (primary sequences), unstimulated HCAEC (control sequences) compared to VEGFA-stimulated HCAEC (primary sequences). For the BP association data, I used the BP SNPs (control sequences) compared to HCASMC SNPs (primary sequences), unstimulated HCAEC (control sequences) compared to VEGFA-stimulated HCAEC (primary sequences). The last step was to select "Eukaryotic DNA" and "Vertebrates (in vivo and in silico)" from the motifs database for enrichment.

1. For each motif, scan each sequence and record the 'best' site; in the event of k ties for the 'best' score in a sequence, $k > 1$, count each tie as $1/k$ sites rather than 1.

2. For each bin size (increased in steps of 2 to maintain symmetry), calculate the binomial *p*-value

3. Select the bin size with lowest *p*-value and plot the site probability for the distribution of location of 'best' sites.

4. If control sequences are provided, repeat the motif scan for the control sequences and count how many 'best' sites fall in the optimal bin determined from the primary sequences. Then

compute the E-value for a Fisher's exact test comparing the counts in that bin in the primary and control sequences.

4. ATAC-seq optimisation for human vascular endothelial cells

4.1 Introduction

Eukaryotic chromatin is tightly packed into an array of nucleosomes, each consisting of four histone proteins wrapped around by 147 bp of DNA and separated by linker DNA, as previously discussed in Chapter 1, section 1.3 [53, 97, 311]. These proteins can be post-translationally altered by covalent modifications or replaced by histone variants [53, 312-316]. Nucleosome positioning throughout a genome has a significant regulatory function and affects DNA dependent processes including, transcription, repair, replication and recombination [317]. Open chromatin regions of the genome are primary positions for the binding of regulatory elements [63, 318]. Changes in chromatin structures have been implicated in many aspects of human health due to mutations in chromatin remodelers [319-321]. Thus, current interest is placed on comparing genome-wide chromatin accessibility to examine epigenetic changes that accompany cell differentiation, environmental signalling and disease development [322].

Several high-throughput methods have been developed to assess chromatin accessibility, nucleosome positioning, and occupancy of DNA-associated proteins, as previously discussed in Chapter 1, section 1.8. High-throughput methods that directly measure chromatin accessibility include DNase-seq, FAIRE-seq and ATAC-seq, while MNase-seq indirectly evaluates chromatin accessibility [322].

As part of the ENCODE effort to explore regulatory elements across the genome, DNase-seq and FAIRE-seq have been used to identify accessible chromatin and generate genome-wide open chromatin maps in a variety of cell types, including HUVECs [122, 125, 323-325].

ATAC-seq is a newer method and has gained considerable interest for several reasons [322]. Firstly, ATAC-seq chromatin accessibility profiles are as sensitive or more sensitive in comparison to other methods, including FAIRE-seq, DNase-seq and MNase-seq, but require much less starting material [137, 142]. Secondly, the ATAC-seq protocol can be performed in much less time. Thirdly, it does not require fixation of cells, which tends to negatively impact the quality of the data [326, 327].

ATAC-seq uses hyperactive Tn5 transposase to fragment and integrate into active regulatory regions *in vivo*, as previously discussed in Chapter 1, section 1.8.4 [328, 329] [95]. During ATAC-seq, 500 to 50,000 unfixed nuclei are tagged *in vitro* with sequencing adapters by purified Tn5 transposase. Adapters are integrated into accessible chromatin regions during

PCR for library preparation followed by paired-end NGS. This method is used to study accessible chromatin regions, nucleosome positioning and TF footprints genome-wide [95].

ATAC-seq has enabled the generation of high-fidelity chromatin accessibility profiles for a variety of cell types [330-335]. However, different cell types and tissues often require different optimisation steps to generate high-quality chromatin accessibility data [336, 337]. Omni-ATAC is a modified ATC-seq protocol that may be able to improve data quality across all cell types. These improvements include i) the use of multiple detergents, including NP40, Tween-20 and digitonin, to improve permeabilisation across a wide array of cell types and to remove mitochondria from the transposition reaction (ii) a post-lysis wash step using Tween-20 to remove mitochondria and increase library complexity, and (iii) the use of PBS to increase the signal-to-background ratio [338].

The aim this chapter was to (a) optimise an ATAC-seq protocol for vascular endothelial cells considering two methodologies, ATAC-seq and Omni-ATAC and (b) establish an informatics pipeline for processing ATAC-seq data and calling peaks.

4.2   Methods

4.2.1   ATAC-seq optimisation in HCtAEC

Several parameters were controlled for during ATAC-seq optimisation, including ATAC-seq methodology, cell number, lysis time and mixing procedure (**Table 2**).

Table 2. ATAC-seq optimisation parameters using HCtAEC.

Sample ID	Protocol	Cell number	Experimental conditions
HCtAEC-1	ATAC-seq by Buenrostro *et al*., 2015	50K	Longer lysis - 45 minutes instead under 1 minute
HCtAEC-2	ATAC-seq by Buenrostro *et al*., 2015	50K	Manual mix - every 10 minutes, mix with pipette, total 30 minutes
HCtAEC-3	ATAC-seq by Buenrostro *et al*., 2015	100K	Double amount of cells, manual mixing every 10 mins, total 30 minutes
HCtAEC-4	ATAC-seq by Buenrostro *et al*., 2015	50K	Thermomix speed - 500 rpm for 10 secs, every 10 minutes
HCtAEC-5	ATAC-seq by Buenrostro *et al.,* 2015	50K	Manual mix - every 10 mins, mix with pipette - total 1 hour
HCtAEC-6	Omni-ATAC by Corces *et al*., 2017	50K	Omni lysis buffer 1+2, TD buffer made in house
HCtAEC-7	Omni-ATAC by Corces *et al.,* 2017	50K	Omni lysis buffer 1+2, Illumina TD buffer

HCtAEC-8	ATAC-seq by Buenrostro *et al.,* 2015	50K	Thermomix speed - 1000 rpm continuously for 30 minutes

4.2.2   ATAC-seq peak and intersection

DNA from HCtAEC was quantified using the Qubit and TapeStation as detailed in Chapter 3, section 3.8. FastQC [285] was performed as detailed in Chapter 3, section 3.11. The BEDTools [341] intersect function was used to identify all reads in peaks (HCtAEC ATAC-seq reads in HUVEC DNaseI-seq/FAIRE-seq peaks; HCtAEC reads in peaks (30M; 50M read depth)) as detailed in Chapter 3, section 3.21. multiBamSummary [293] was used to compute the read coverages in HCtAEC as detailed in Chapter 3, section 3.20.

4.3   Results

4.3.1   Quantification of ATAC-seq/Omni-ATAC DNA

The quantity of purified ATAC-seq/Omni-ATAC DNA was assessed using Qubit to determine the DNA yield in each of the eight tested conditions (**Table 3**). HCtAEC-6 sample (Omni-ATAC) resulted in a distribution of fragment sizes similar to desired ATAC-seq characteristics and size intervals of ~200 bp (**Figure 16 F**). All other HCtAEC samples (HCtAEC-1, HCtAEC-2, HCtAEC-3, HCtAEC-4, HCtAEC-5, HCtAEC-7, and HCtAEC-8) lacked a clear pattern of periodicity (**Figure 16 A, B, C, D, E, G, H**).

Table 3. DNA concentration of ATAC-seq/Omni-ATAC libraries determined by Qubit.

Sample ID	Concentration [pg/ul]
HCtAEC-1	4.0
HCtAEC-2	3.9
HCtAEC-3	2.9
HCtAEC-4	3.8
HCtAEC-5	4.2
HCtAEC-6	8.5
HCtAEC-7	7.0
HCtAEC-8	4.1

A

B

C

D

79

E

F
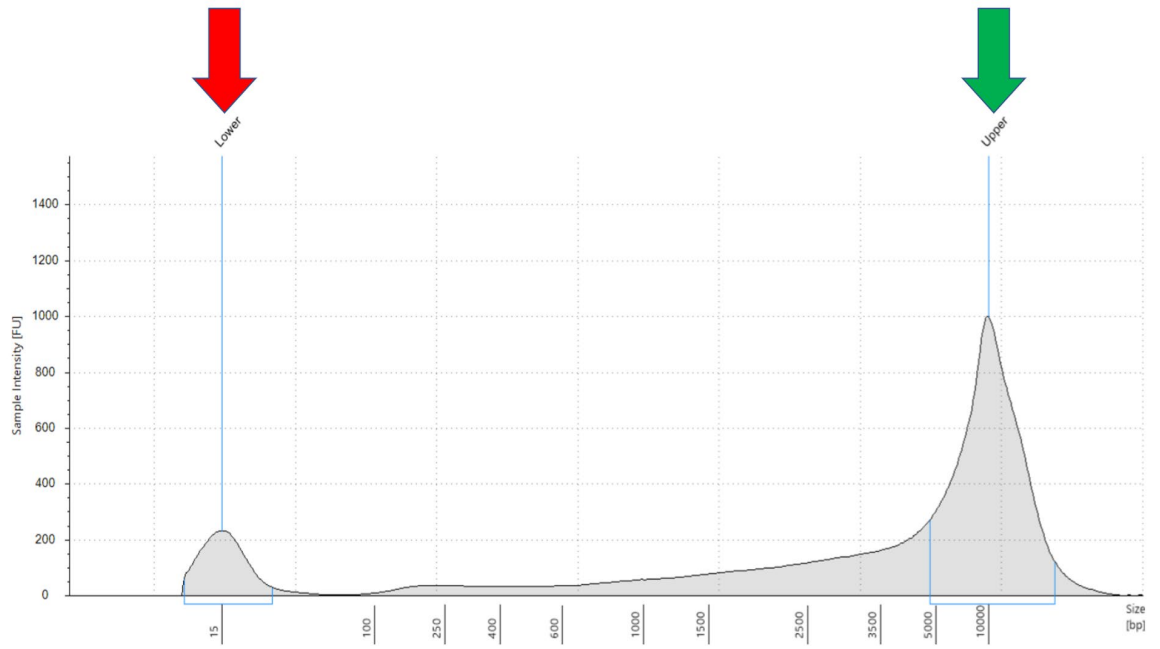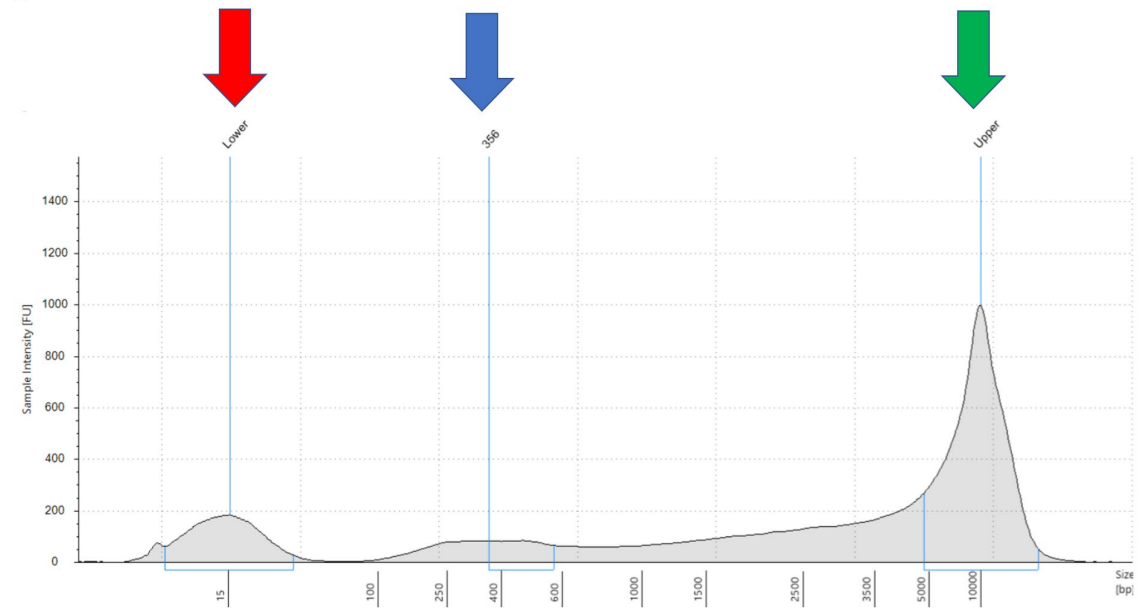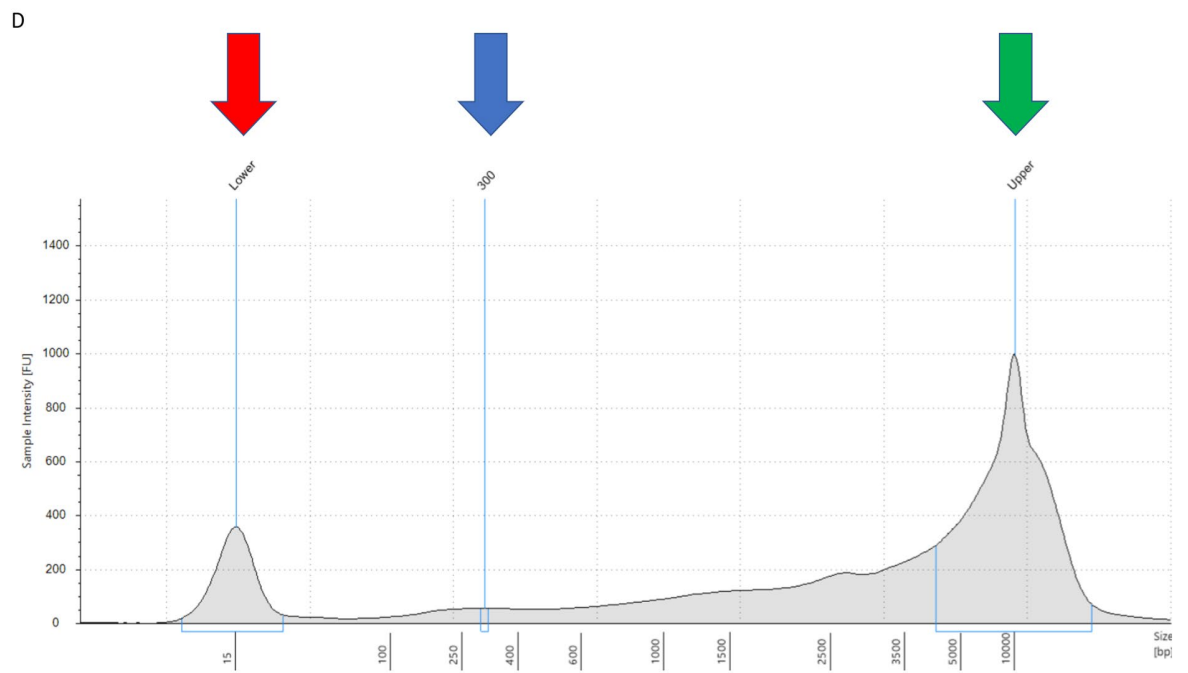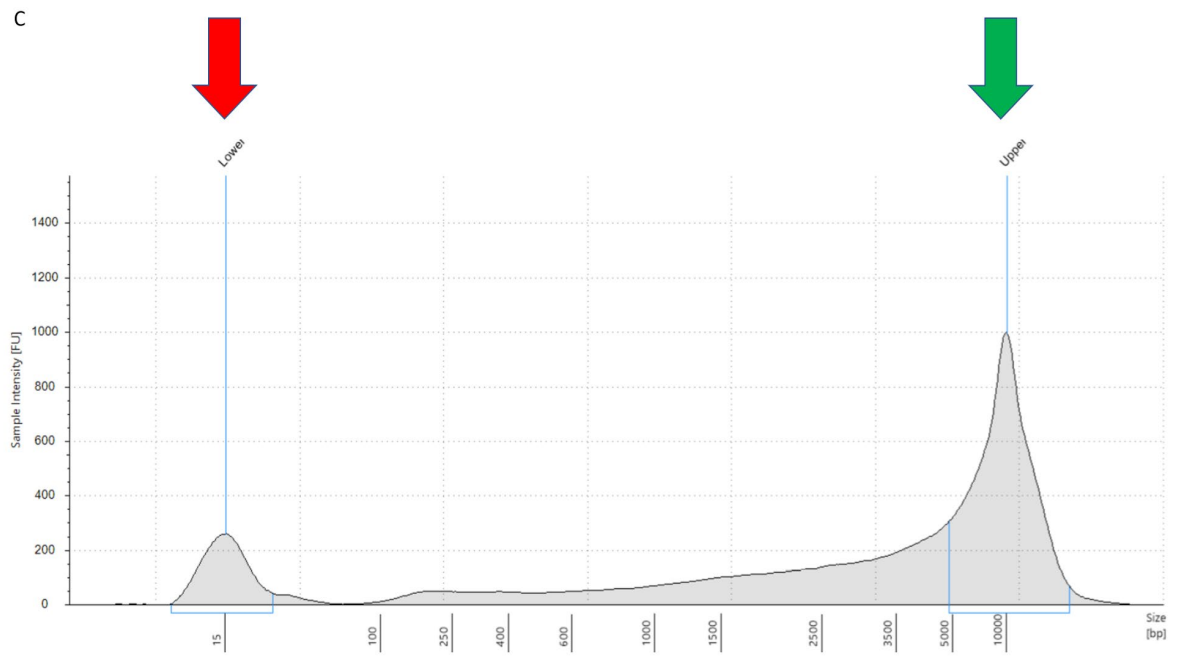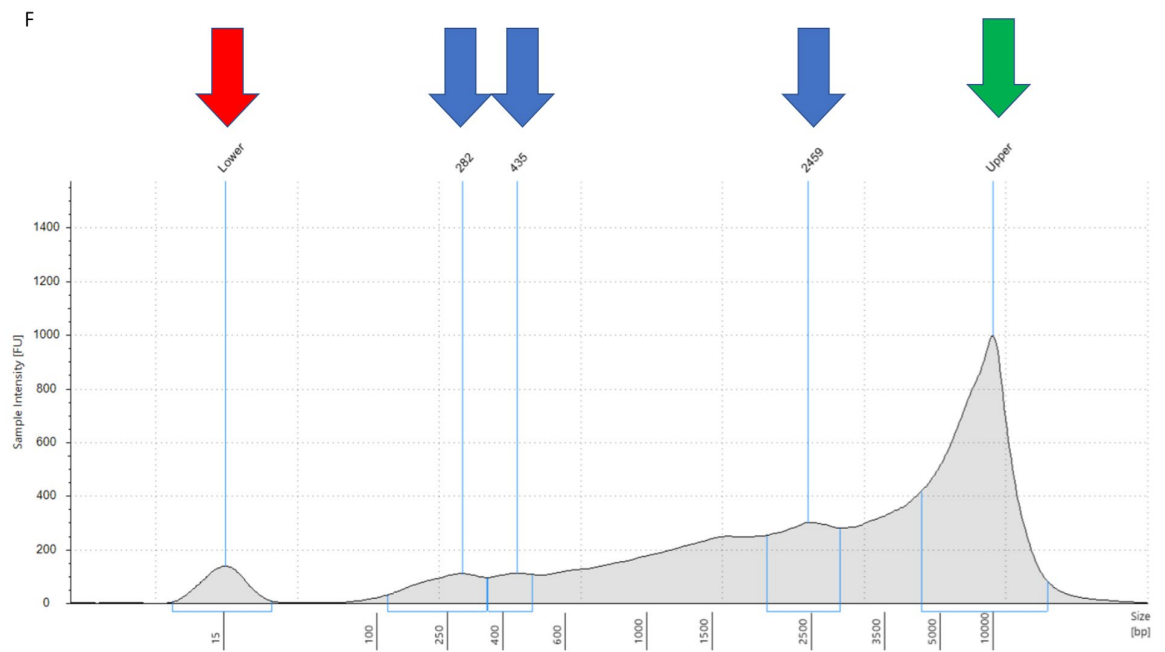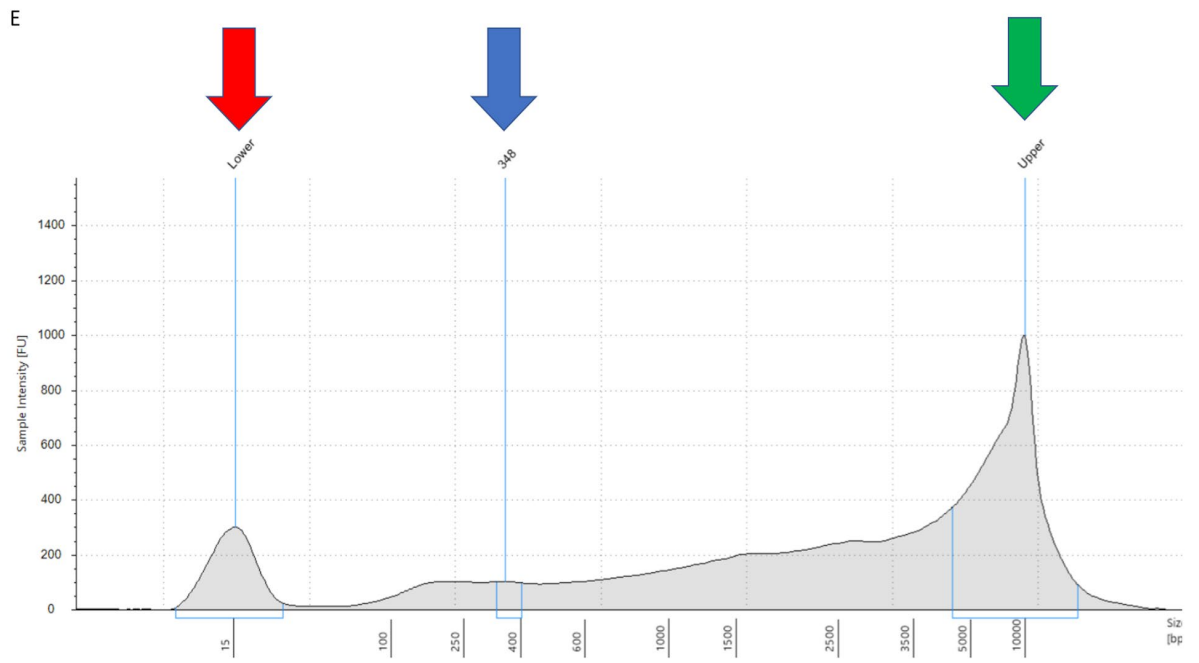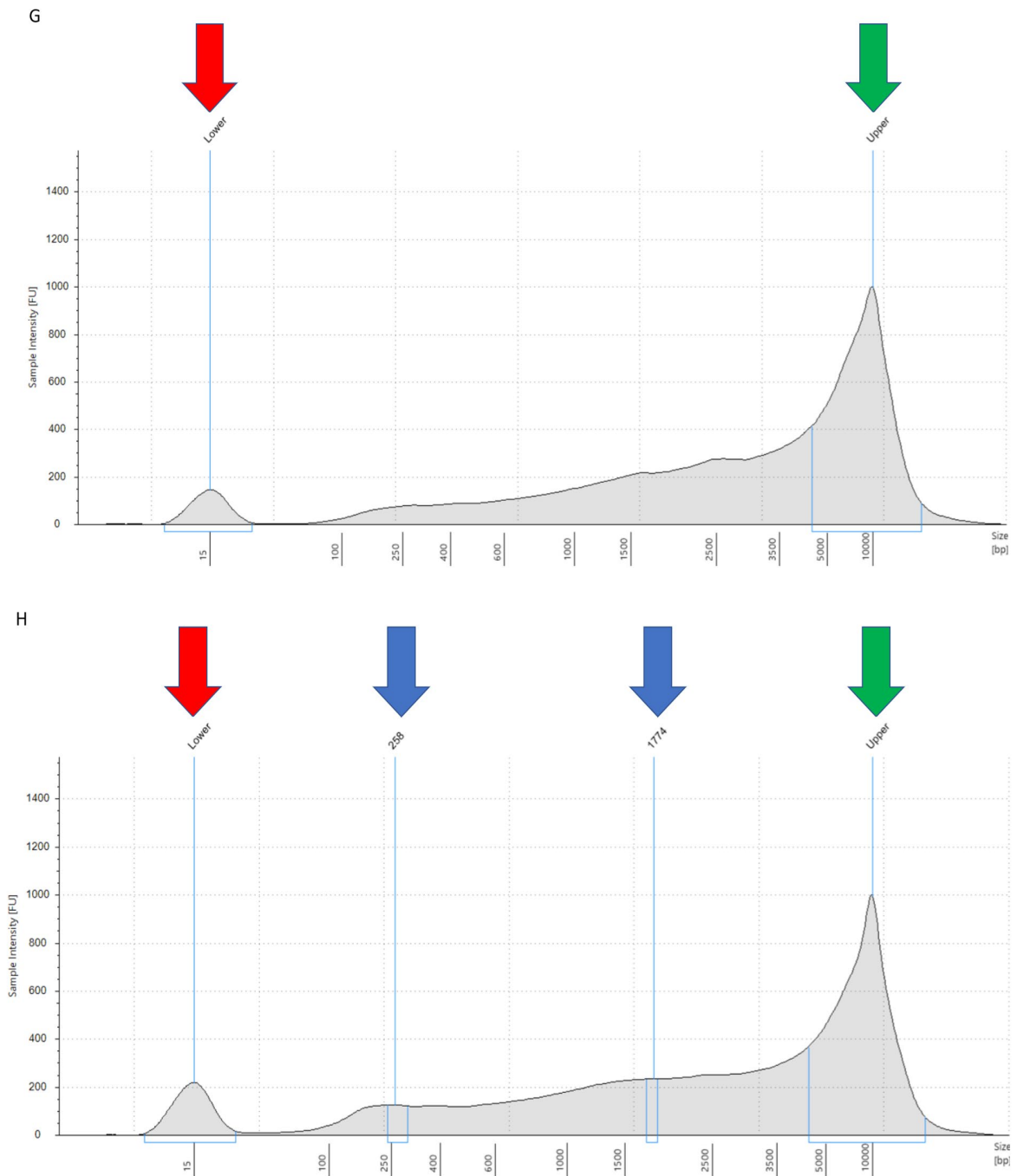
Figure 16. Electropherogram of fragment size distribution of ATAC-seq/Omni-ATAC libraries determined by TapeStation. The X-axis represents the product size (bp), and the Y-axis represents the arbitrary fluorescence intensity (FU). Red and green arrows represent lower and upper size markers, respectively. A) The figure shows a poor fragment size distribution plot of HCtAEC-1 sample. B) The figure shows a poor fragment size distribution of HCtAEC-2 sample. Blue arrow at 356 bp indicate fragment size of peak, which may represent mononucleosome. C) The figure shows a poor fragment size distribution of HCtAEC-3 sample. D) The figure shows a poor fragment size distribution of HCtAEC-4 sample. Blue arrow at 300 bp indicate fragment size of peak, which may represent mononucleosome. E) The figure shows a poor fragment size distribution of HCtAEC-5 sample. Blue arrow at 348 bp indicate fragment size of peak, which may represent mononucleosome. F) The

figure shows a good fragment size distribution of HCtAEC-6 sample. Blue arrows at 282 bp, 435 bp, and 2459 bp indicate fragment sizes of peaks, which may represent nucleosome-free, dinucleosome or multinucleated fragments, respectively. G) The figure shows a poor fragment size distribution of HCtAEC-7 sample. H) The figure shows a poor fragment size distribution of HCtAEC-8 sample. Blue arrows at 258 bp and 1774 bp indicate numbers indicate fragment sizes of peaks, which may represent nucleosome-free and multinucleated fragments, respectively.

4.3.2   FastQC analysis of ATAC-seq/Omni-ATAC reads

Due to the inconclusive nature of the TapeStation data, all samples initially underwent sequencing to 1M reads using Illumina's Hi-seq (NextSeq 500 High Output Run) to determine whether the ATAC-seq/Omni-ATAC library preparation was successful and which samples provided most definition. I set up the following pipeline to analyse data: raw sequencing data was examined using i) FastQC to identify potential problems, (i) Trimmomatic to trim, crop and remove adapters, (ii) Bowtie2 to align sequencing reads to the human reference genome (hg19), (iii) Picard to sort SAM/BAM files, merge multiple SAM and/or BAM files into a single file, and identify duplicate reads, (iv) MACS2 to identify regions of genomic enrichment, v) UCSC Genome Browser to interactively visualise genomic data (**Figure 17**).
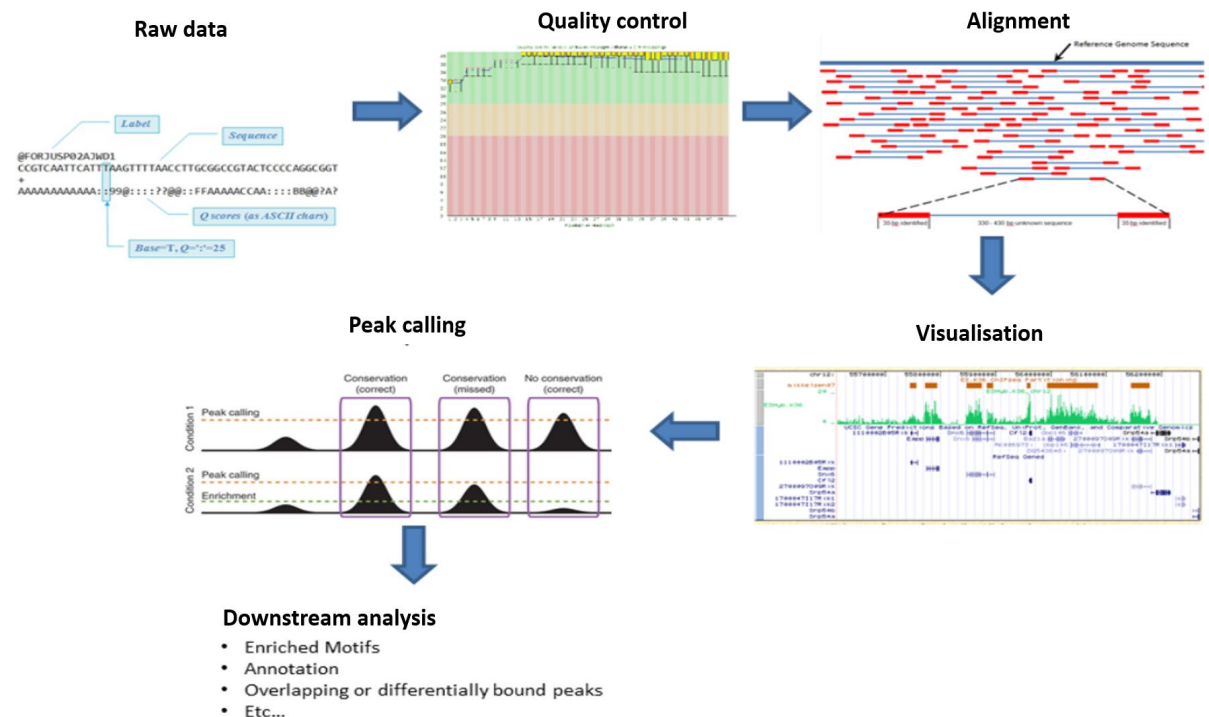
Figure 17. ATAC-seq data analysis. The figure illustrates all the steps, including (quality control, trimming, alignment, peak calling, and visualisation) processed using different bioinformatics tools (adapted from https://www.utsouthwestern.edu/labs/bioinformatics-lab/analysis/chip-seq/).

FastQC was used to check the quality of the raw data. First, per base sequence content was plotted as the proportion of the four nucleotides at each position across all reads. The expectation is that the proportion for the four nucleotides should remain relatively constant across all reads. For most libraries there was clear non-uniform distribution of bases for the first 13 nucleotides caused by transposase fragmentation (**Figure 18**).
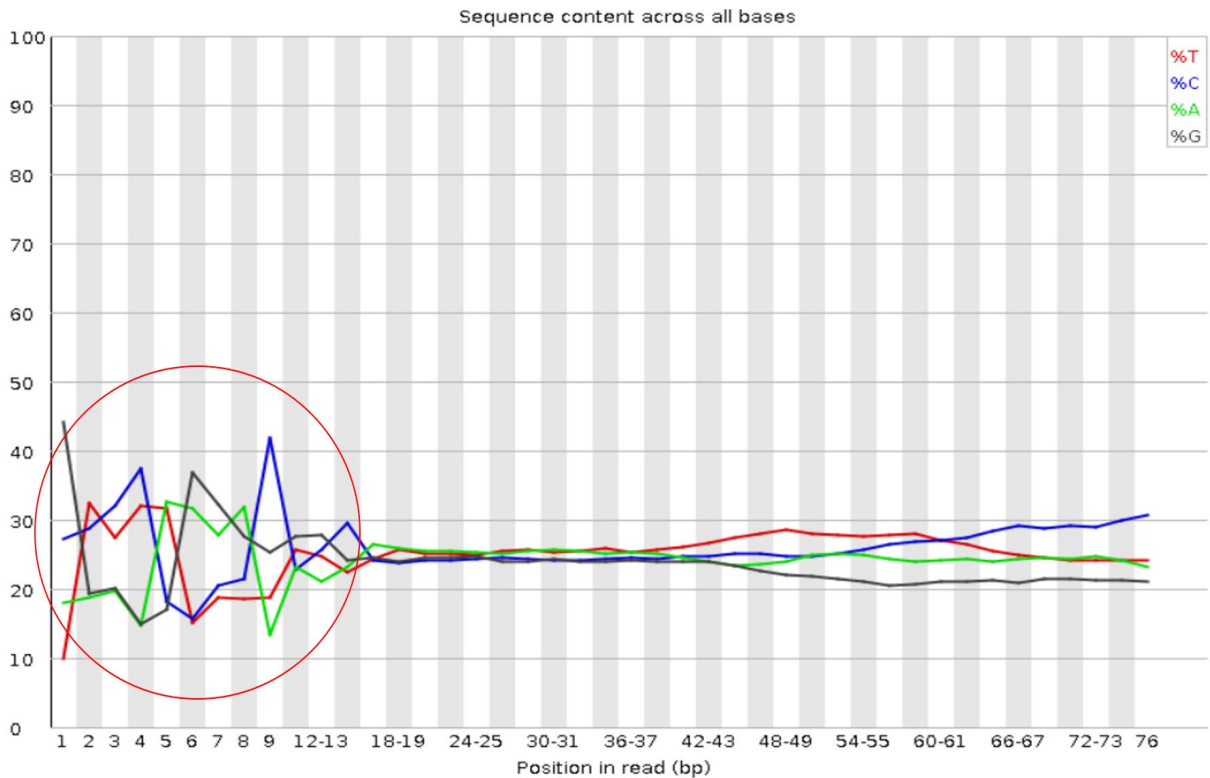
Figure 18. Representative per base sequence content plot of sample HCtAEC-6 at 1M reads per sample. The plot shows a biased sequence composition at the start of the read. Libraries produced by fragmentation using transposases produce bias in the start of the read.

Next, adapter content was plotted as the fraction of reads where the sequence library adapter sequence is identified at the indicated base position. Only adapters specific to the library type are searched. The expectation is that there should not be any adapter sequence present, however there might be a presence of adapter at the 3' end of the read (**Figure 19**).
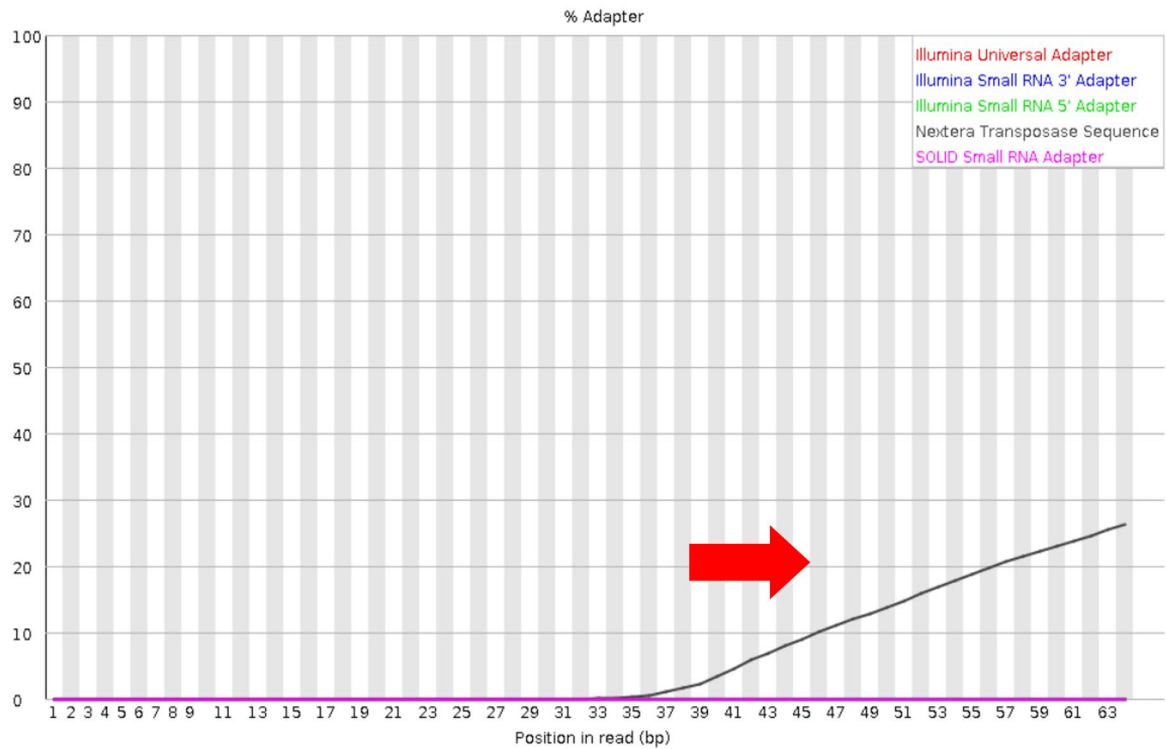
Figure 19. Representative adapter content plot of sample HCtAEC-6 at 1M reads per sample. The plot shows the presence of nextera transposase sequence in ~30% of all reads.

Lastly, Kmer content plots measure the count of each short nucleotide of length k (default = 7) starting at each position along the read. Any given kmer should be evenly represented across the length of the read. A list of kmers which appear to be specific position with greater than expected frequency are reported. The positions of the six most biased kmers are plotted. The libraries may have highly represented kmers that are derived from highly expressed sequences. The biased kmers are near the start of the read likely are due to slight, sequence dependent efficiency of DNA shearing or random priming (**Figure 20**).
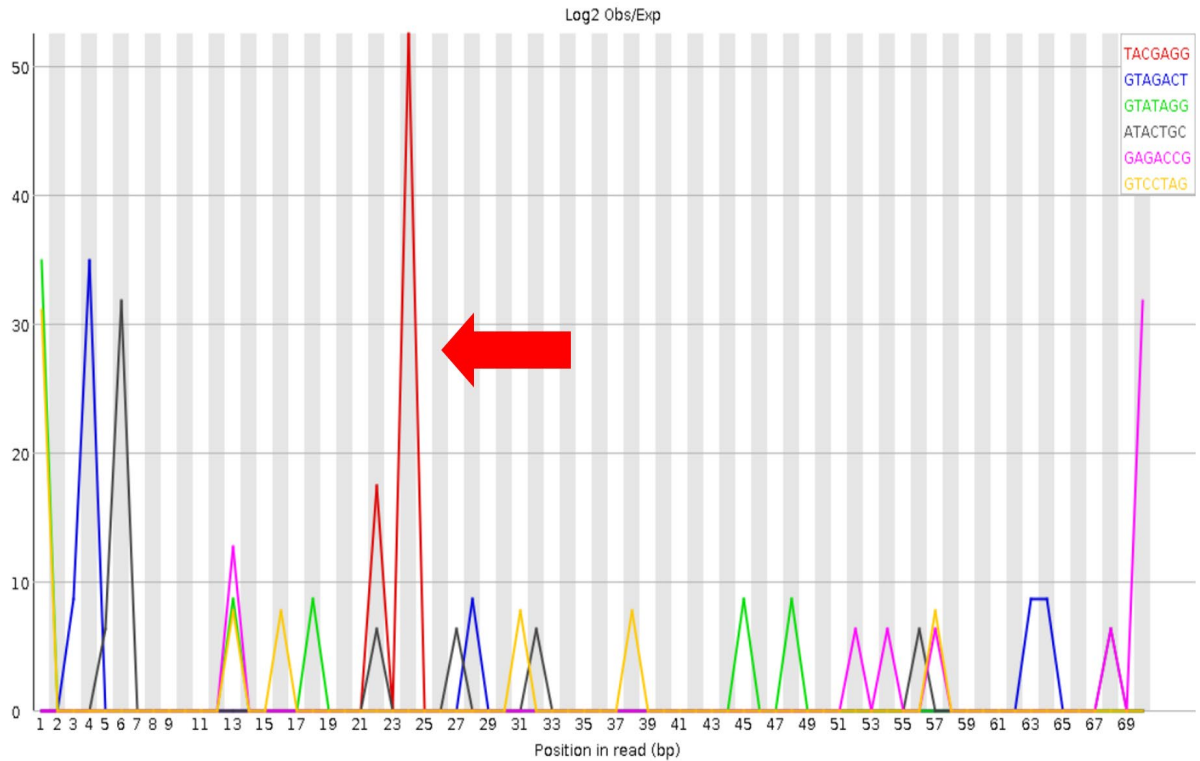
Figure 20. Representative Kmer content plot of sample HCtAEC-6 at 1M reads per sample. The plot shows Kmer bias at the start of the read.

I used Trimmomatic to trim, crop and remove adapters, which can cause a problem depending on the library preparation and downstream application. The optimal parameters I derived using the data from all HCtAEC samples at 1M reads depth, are: cut NexteraPE-PE.fa:2:30:10 from the read, cut bases off the start of the read, if below a threshold quality of 3, cut bases off the end of the read if below a threshold quality of 3, scanned at the 5' end and clipped the read once the average quality within the window falls below a 10:20, cut 12 bases from the start of the read, and dropped the read if below 36.

The per base sequence content plot after trimming showed that the proportion of the four bases remained constant over the read length %A=%T and %G=%C (**Figure 21**).
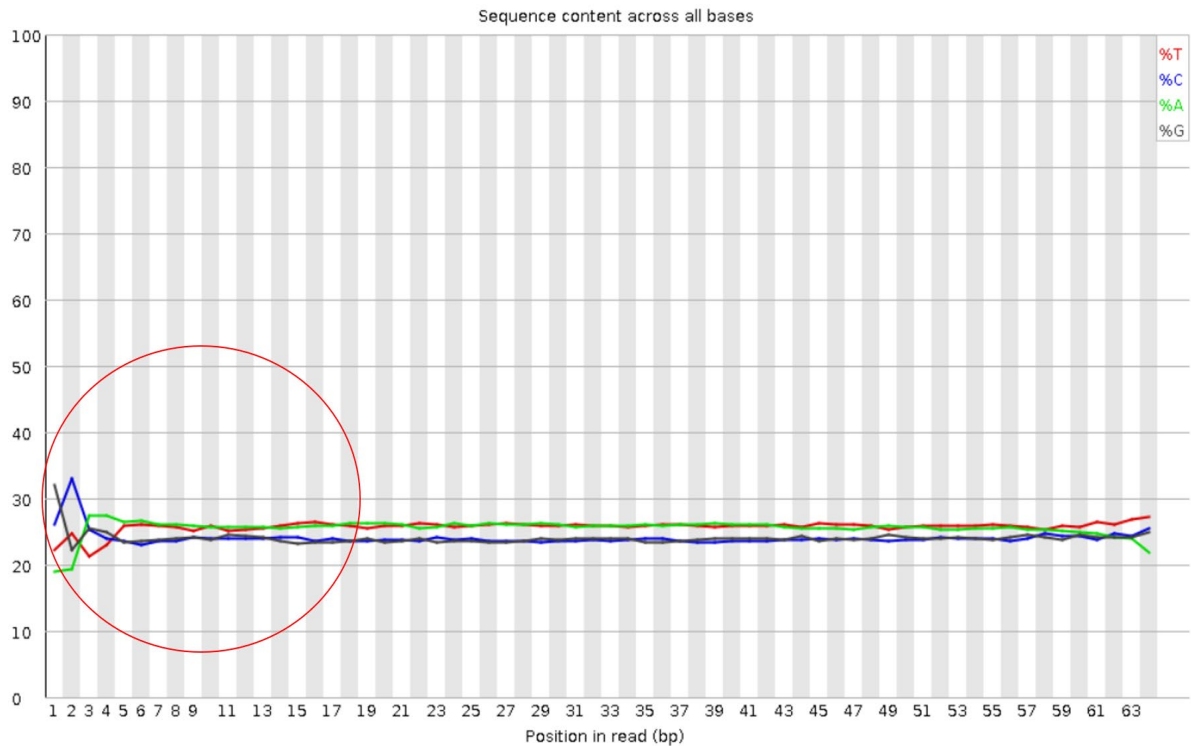
Figure 21. Representative per base sequence content plot of sample HCtAEC-6 at 1M reads per sample. The plot shows that there is little to no difference between the different bases of a sequence run, so the lines run parallel with each other.

The adapter content plot after trimming showed that there were no adapter sequence present (**Figure 22**).
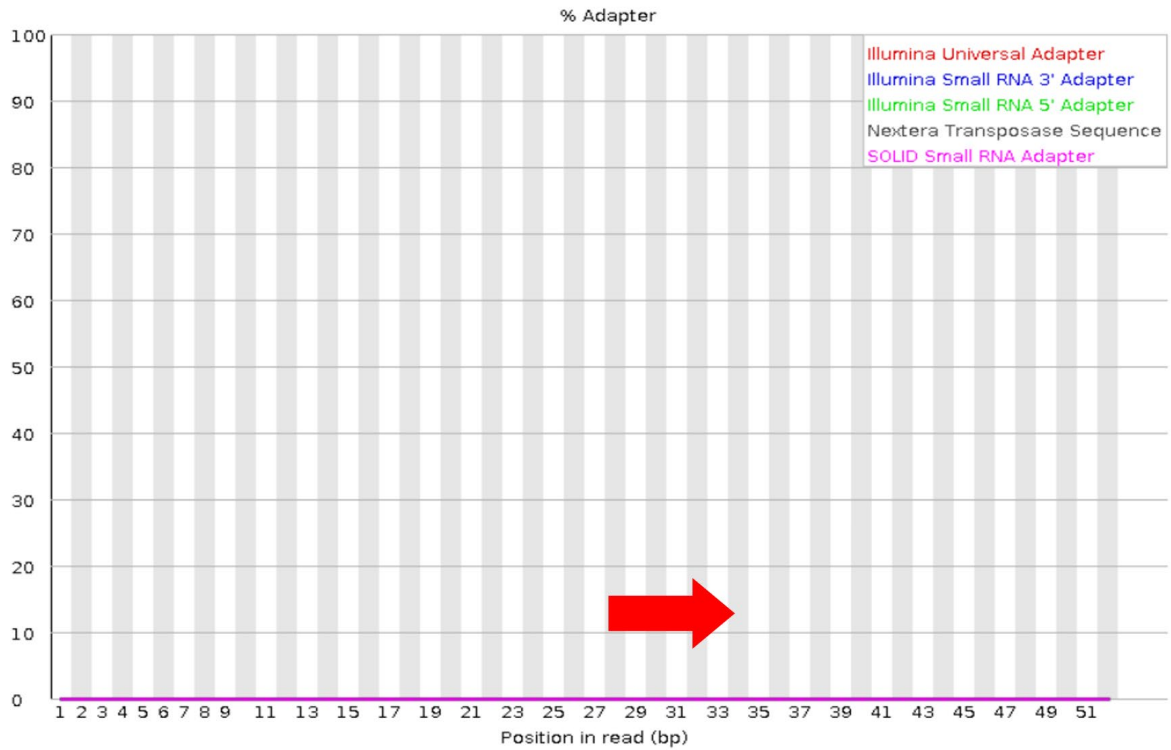
Figure 22. Representative adapter content plot of sample HCtAEC-6 at 1M reads per sample. The plot shows there are no adapter sequences.

The Kmer Content plot after trimming showed that there was some Kmer bias at the start of the library (**Figure 23**).
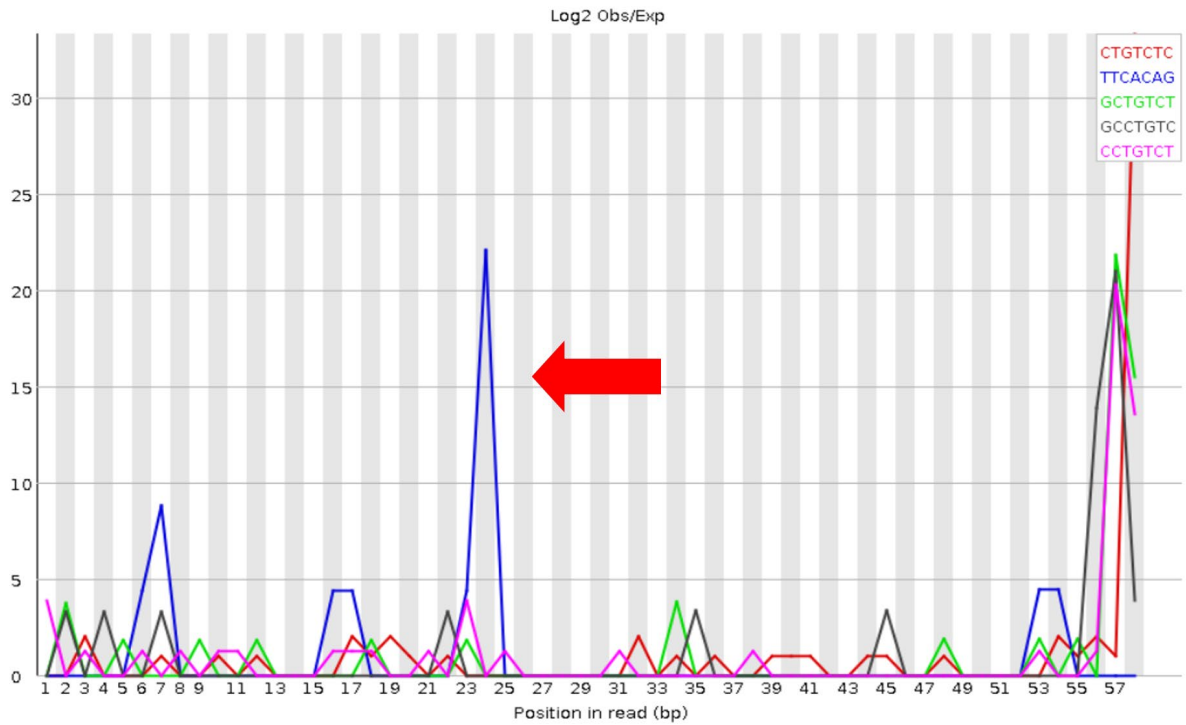
Figure 23. Representative Kmer content plot of sample HCtAEC-6 at 1M reads per sample. The plot shows kmer bias at the start of the read.

4.3.3   Comparison of ATAC-seq/Omni-ATAC profiles (1M reads depth)

The FRiP measures the signal-to-background ratio, since reads falling within regions of enrichment are informative of open chromatin, while reads falling outside regions of enrichment represent background noise. The FRiP was used to determine reads in peaks of the ATAC-seq/Omni-ATAC profiles in HCtAEC at 1M reads depth to prioritise the more reliable ATAC-seq/Omni-ATAC datasets. A conventional FRiP was not performed as 1M reads depth is not sufficient to generate uniform coverage of ATAC-seq/Omni-ATAC peaks across the genome. Instead, a modified FRiP was performed to examine ATAC-seq/Omni-ATAC reads in HUVEC accessible chromatin to prioritise the more reliable ATAC-seq/Omni-ATAC datasets. Mapped HCtAEC ATAC-seq/Omni-ATAC reads were overlapped with ENCODE HUVEC DNaseI-seq peaks. The fraction of HCtAEC ATAC-seq/Omni-ATAC reads overlapping with these ENCODE HUVEC DNaseI-seq peaks showed that two ATAC-seq samples HCtAEC-4 (0.04) and HCtAEC-5 (0.05), and two Omni-ATAC samples HCtAEC-6 (0.08), and HCtAEC-7 (0.07) had the highest FRiP scores (**Table 4**). Similar analyses using ENCODE HUVEC FAIRE-seq peaks, showed that the two ATAC-seq samples HCtAEC-4 (0.04) and HCtAEC-5 (0.05), and Omni-ATAC samples HCtAEC-6 (0.07) and HCtAEC-7 (0.06), had the highest FRiP scores (**Table 5**). Therefore, these four samples were selected to be sequenced to 30M reads depth.

Table 4. Fraction of HCtAEC ATAC-seq/Omni-ATAC reads in ENCODE HUVEC DNaseI-seq peaks.

Depth	Sample ID	Reads in peaks	Total reads used to call peaks	FRiP score
1M	HCtAEC-1	23844	862142	0.03
	HCtAEC-2	24902	1079575	0.02
	HCtAEC-3	22436	750431	0.03
	HCtAEC-4	87709	2157080	0.04
	HCtAEC-5	42610	811684	0.05
	HCtAEC-6	79770	1060773	0.08
	HCtAEC-7	65709	992692	0.07
	HCtAEC-8	34175	880995	0.04

Table 5. Fraction of HCtAEC ATAC-seq/Omni-ATAC reads in ENCODE HUVEC FAIRE-seq peaks.

Depth	Sample ID	Reads in peaks	Total reads used to call peaks	FRiP score
1M	HCtAEC-1	21551	862142	0.02
	HCtAEC-2	21120	1079575	0.02
	HCtAEC-3	19923	750431	0.03
	HCtAEC-4	76966	2157080	0.04
	HCtAEC-5	38144	811684	0.05
	HCtAEC-6	69961	1060773	0.07

| | HCtAEC-7 | 57641 | 992692 | 0.06 |
| | HCtAEC-8 | 29713 | 880995 | 0.03 |

The UCSC Genome Browser was used to visualise mapped reads from HCtAEC ATAC-seq/Omni-ATAC data in relation to ENCODE HUVEC DNase and FAIRE-seq data. Loci surrounding highly expressed genes in HUVECs were selected [339], including Nitric Oxide Synthase 3 (*NOS3*), SMAD Family Member 1 (*SMAD1*), and Bone Morphogenetic Protein 6 (*BMP6*) to visualise reads between the 8 different ATAC-seq/Omni-ATAC conditions [339]. Looking at the mapped reads around selected loci, including *NOS3*, *SMAD1*, and *BMP6*, I found that the four ATAC-seq samples HCtAEC-1, HCtAEC-2, HCtAEC-3, and HCtAEC-8, resulted in higher background noise, which is supported by the modified FRiP score analysis. There is less consistency in the accessible chromatin regions between the four ATAC-seq samples HCtAEC-1, HCtAEC-2, HCtAEC-3, and HCtAEC-8 and the ENCODE HUVEC DNaseI-seq/FAIRE-seq (**Figure 24 A, B, and C**).

Figure 24. Visualisation of ATAC-seq/Omni-ATAC reads in HCtAEC at 1M reads per sample with ENCODE HUVEC DNaseI-seq/FAIRE-seq reads at selected loci, including a) *NOS3*, b) *SMAD1*, and c) *BMP6* loci. ATAC-seq/Omni-ATAC reads in HCtAEC are displayed in black, while ENCODE HUVEC DNaseI-seq/FAIRE-seq reads are displayed in orange. The four ATAC-seq samples HCtAEC-1, HCtAEC-2, HCtAEC-3, and HCtAEC-8, resulted in higher background noise, which is supported by the modified FRiP score analysis.

4.3.4 Comparison of ATAC-seq/Omni-ATAC profiles (30M reads depth)

I performed the same analyses using HCtAEC-4, HCtAEC-5, HCtAEC-6, and HCtAEC-7 ATAC-seq/Omni-ATAC samples that had the highest FRiP scores but sequenced to 30M read depth. Mapped HCtAEC ATAC-seq/Omni-ATAC reads were overlapped with peaks. The fraction of HCtAEC ATAC-seq/Omni-ATAC reads overlapping with peaks showed that the two Omni-ATAC samples HCtAEC-6 (0.31) and HCtAEC-7 (0.25) had the highest FRiP scores (**Table 6**). Therefore, these two samples were selected to be sequenced to 50M reads depth.

Table 6. Fraction of HCtAEC ATAC-seq/Omni-ATAC reads in peaks.

Depth	Sample ID	Reads in peaks	Total reads used to call peaks	FRiP score
30M	HCtAEC-4	3959835	31898255	0.12
	HCtAEC-5	3591701	19794681	0.18
	HCtAEC-6	9774059	31135955	0.31
	HCtAEC-7	6034746	24515355	0.25

Looking at the mapped reads around the same loci, including *NOS3*, *SMAD1*, and *BMP6*, I found that the two ATAC-seq samples HCtAEC-4, HCtAEC-5, and the two Omni-ATAC samples HCtAEC-6, HCtAEC-7 resulted in lower background noise, which was supported by the conventional FRiP score analysis. There is more consistency in the accessible chromatin regions between the two ATAC-seq samples HCtAEC-4, HCtAEC-5, and the two Omni-ATAC samples HCtAEC-6, HCtAEC-7 and the ENCODE HUVEC DNaseI-seq/FAIRE-seq (**Figure 25 A, B and C**).
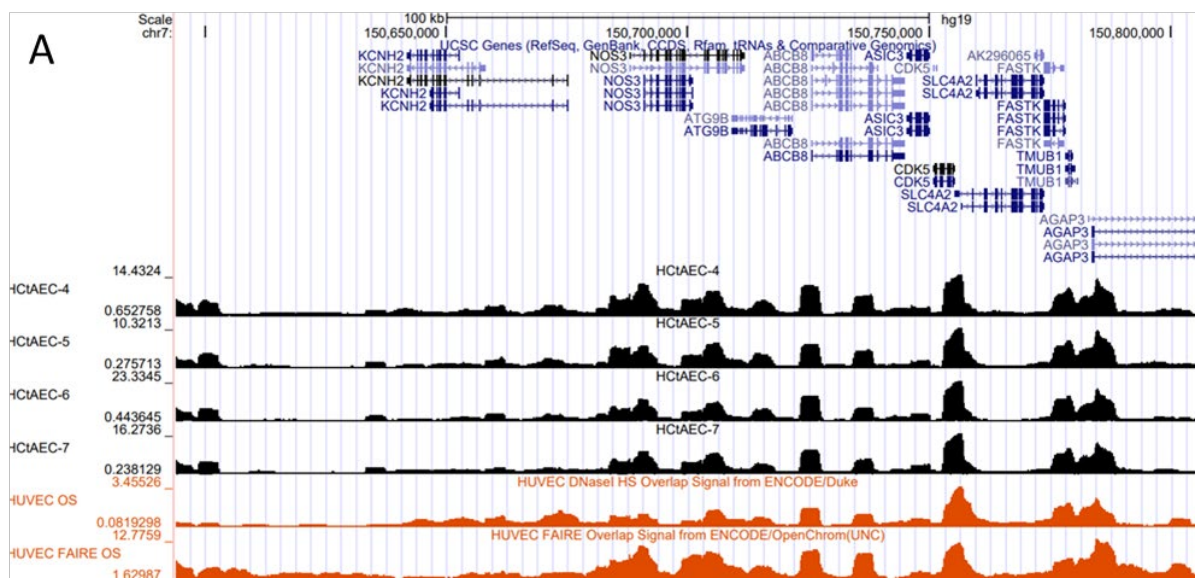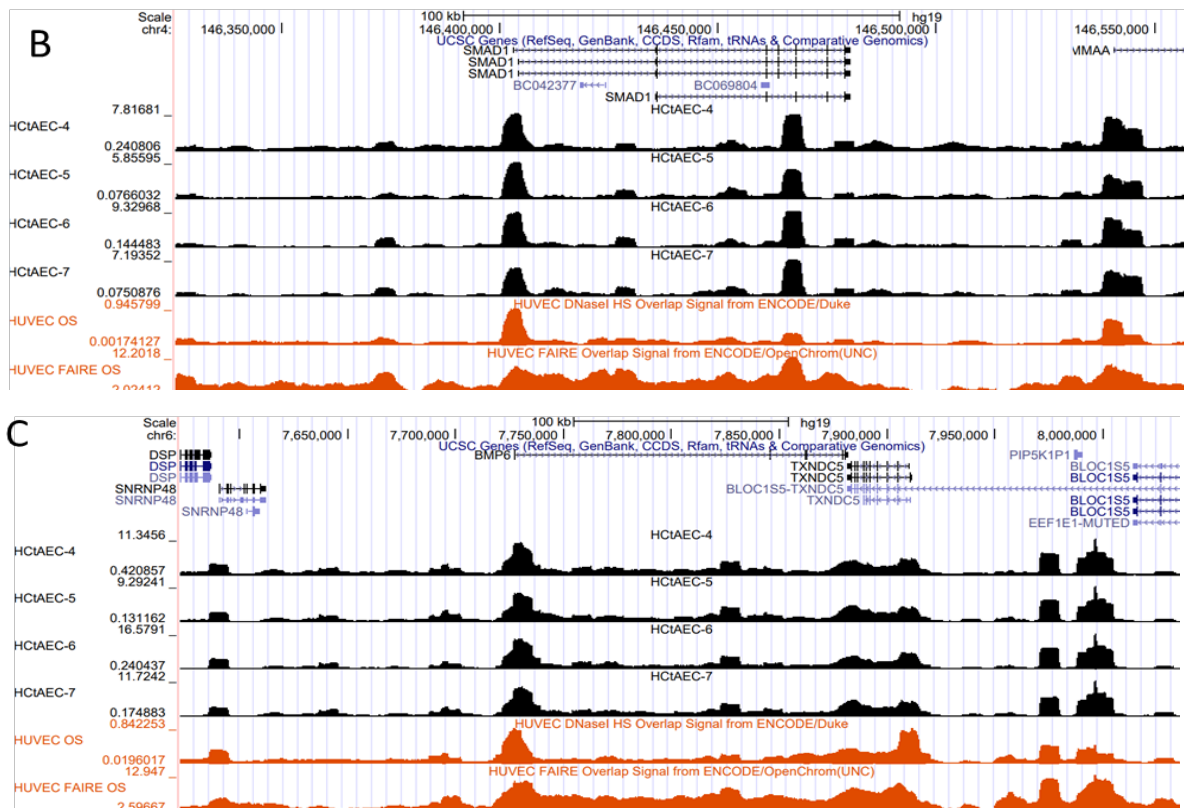
Figure 25. Visualisation of ATAC-seq/Omni-ATAC reads in HCtAEC at 30M reads per sample with ENCODE HUVEC DNaseI-seq/FAIRE-seq reads at selected loci, including a) *NOS3*, b) *SMAD1*, and c) *BMP6* loci. ATAC-seq/Omni-ATAC reads in HCtAEC are displayed in black, while ENCODE HUVEC DNaseI-seq/FAIRE-seq reads are displayed in orange. The two ATAC-seq samples HCtAEC-4, HCtAEC-5, and the two Omni-ATAC samples HCtAEC-6, HCtAEC-7 resulted in lower background noise, which is supported by the conventional FRiP score analysis.

4.3.5 ATAC-seq/Omni-ATAC (30M reads per sample) peak calling

MACS2 was used to call peaks in regions of genomic enrichment in vascular endothelial cells. I obtained 89,160 ATAC-seq peaks in HCtAEC-4, 71,263 ATAC-seq peaks in HCtAEC-5, 117,459 Omni-ATAC peaks in HCtAEC-6, and 96,023 Omni-ATAC peaks in HCtAEC-7. Looking at the mapped peaks around the same loci, including *NOS3*, *SMAD1*, and *BMP6*, I found additional ATAC-seq/Omni-ATAC peaks in HCtAEC-4, HCtAEC-5, HCtAEC-6, and HCtAEC-7 compared to ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks (**Figure 26 A, B and C**).
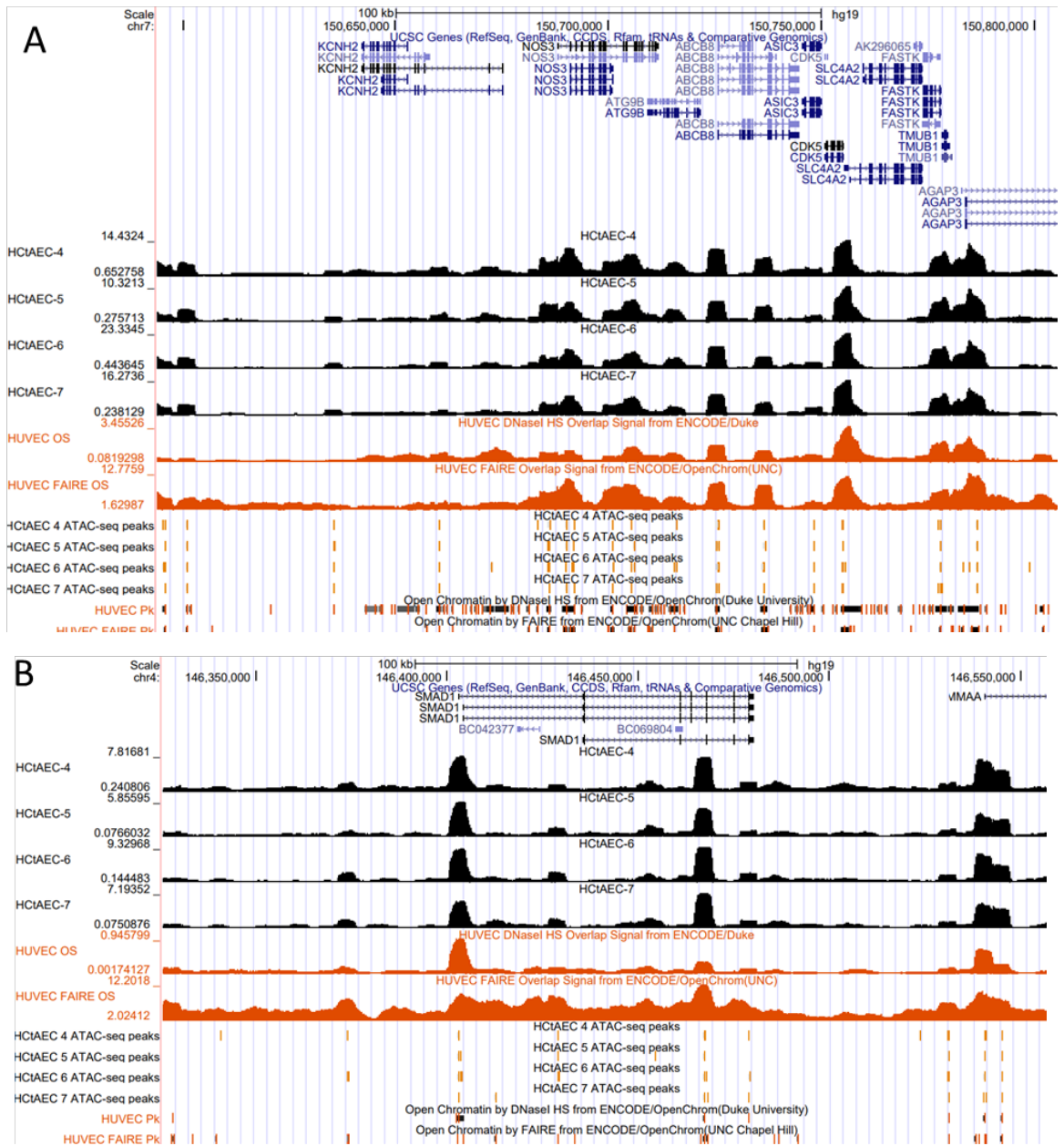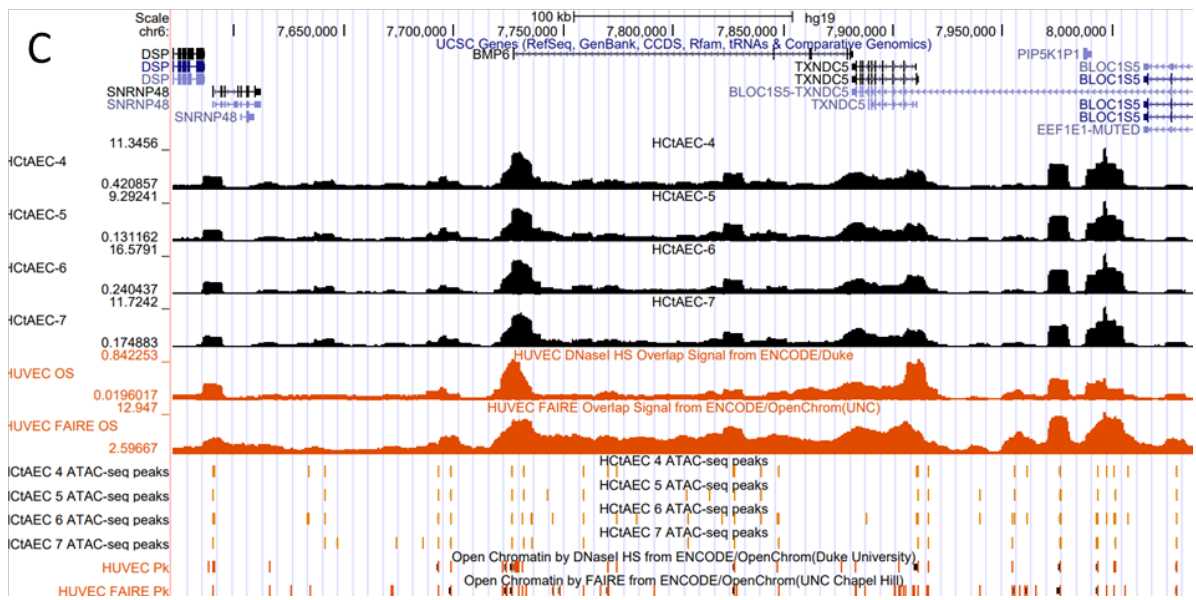
Figure 26. Visualisation of ATAC-seq/Omni-ATAC peaks in HCtAEC at 30M reads per sample with ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks at selected loci, including a) *NOS3*, b) *SMAD1*, and c) *BMP6* loci. ATAC-seq/Omni-ATAC peaks in HCtAEC are displayed in black, while ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks are displayed in orange. There are additional ATAC-seq/Omni-ATAC peaks in HCtAEC-4, HCtAEC-5, HCtAEC-6 and HCtAEC-7 samples compared to ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks.

4.3.6 Correlation analysis of HCtAEC Omni-ATAC samples (50M read depth)

Correlation analysis of two Omni-ATAC samples HCtAEC-6 and HCtAEC-7 was performed to confirm that these two samples were correlated. Pearson correlation coefficient plot of two Omni-ATAC samples HCtAEC-6 and HCtAEC-7 is 0.99, which indicates almost perfect correlation (**Figure 27**).

96

Figure 27. Pearson correlation coefficients plots of read coverages in Omni-ATAC samples HCtAEC-6 and HCtAEC-7. The y axis and x axis represent the number of fragments in Omni-ATAC samples HCtAEC-6 and HCtAEC-7, respectively. Each dot represents one genomic region.

4.3.7 Comparison of Omni-ATAC profiles (50M reads depth)

I performed the same analyses using the two Omni-ATAC samples HCtAEC-6 and HCtAEC-7 Omni-ATAC samples that has the highest FRiP scores but sequenced to 50M read depth. Mapped HCtAEC Omni-ATAC reads were overlapped with peaks. The fraction of HCtAEC Omni-ATAC reads overlapping with peaks showed that the two Omni-ATAC samples HCtAEC-6 and HCtAEC-7, met the recommended 0.2 standard set by ENCODE for ATAC-seq libraries (**Table 7**).

Table 7. Fraction of HCtAEC Omni-ATAC reads in peaks.

Depth	Sample ID	Reads in peaks	Total reads used to call peaks	FRiP score
50M	HCtAEC-6	20678424	57800069	0.36
	HCtAEC-7	14587197	49526467	0.29

Looking at the mapped reads around the same loci, including *NOS3*, *SMAD1*, and *BMP6*, I found that two Omni-ATAC samples HCtAEC-6 and HCtAEC-7 resulted in lower background noise, which is supported by the conventional FRiP score analysis. There is more consistency in the accessible chromatin regions between the two Omni-ATAC samples HCtAEC-6 and HCtAEC-7 and the ENCODE HUVEC DNaseI-seq/FAIRE-seq (**Figure 28 A, B, and C**).
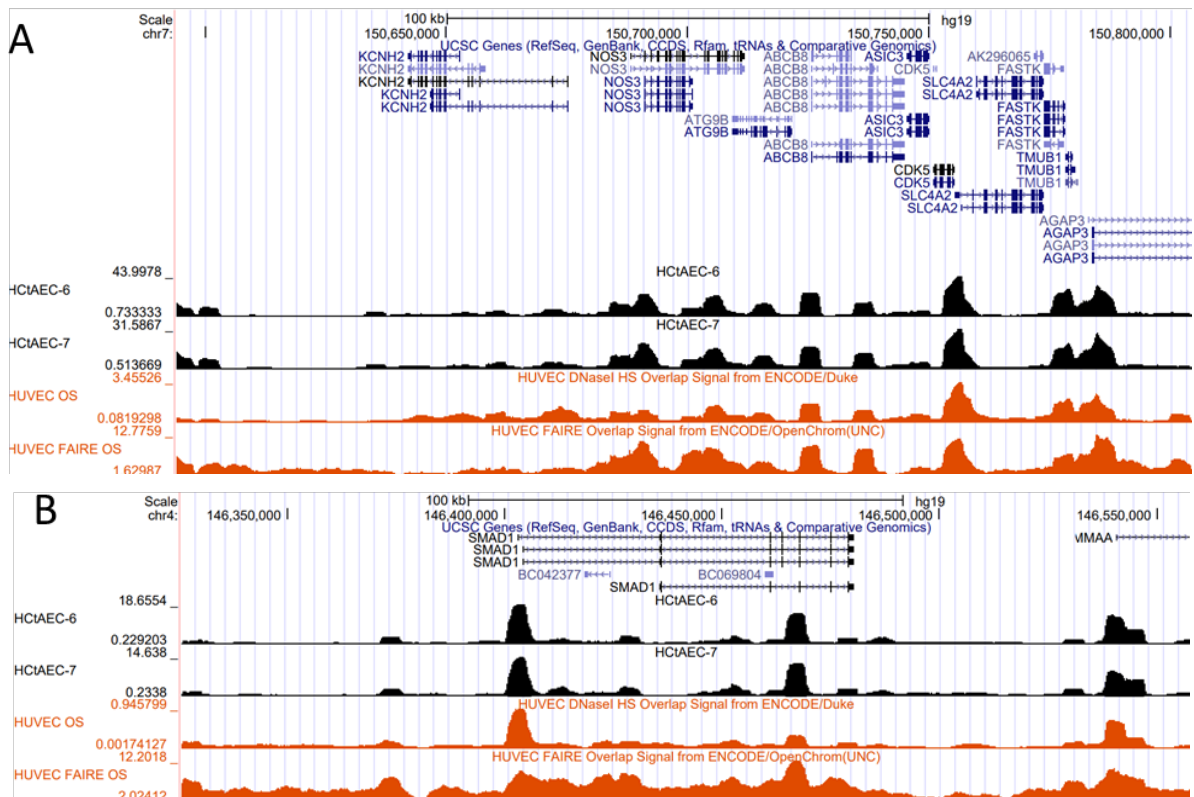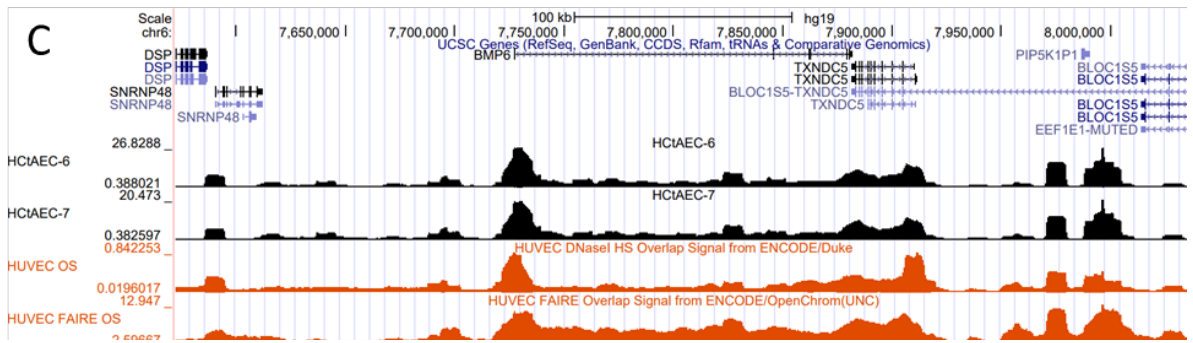
Figure 28. Visualisation of Omni-ATAC peaks in HCtAEC at 50M reads per sample with ENCODE HUVEC DNaseI-seq/FAIRE-seq reads at selected loci, including a) *NOS3*, b) *SMAD1*, and c) *BMP6* loci. Omni-ATAC reads in HCtAEC are displayed in black, while the ENCODE HUVEC DNaseI-seq/FAIRE-seq reads are displayed in orange. The two Omni-ATAC samples HCtAEC-6 and HCtAEC-7 resulted in low background noise, which is supported by the conventional FRiP score analysis.

4.3.8 Omni-ATAC (50M reads per sample) peak calling

I obtained 174,644 Omni-ATAC peaks in HCtAEC-6 and 157,226 Omni-ATAC peaks in HCtAEC-7 sample. Looking at the mapped peaks around the same loci, including *NOS3*, *SMAD1*, and *BMP6*, I found additional Omni-ATAC peaks in the two Omni-ATAC samples HCtAEC-6 and HCtAEC-7 compared to ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks (**Figure 29 A, B and C**).
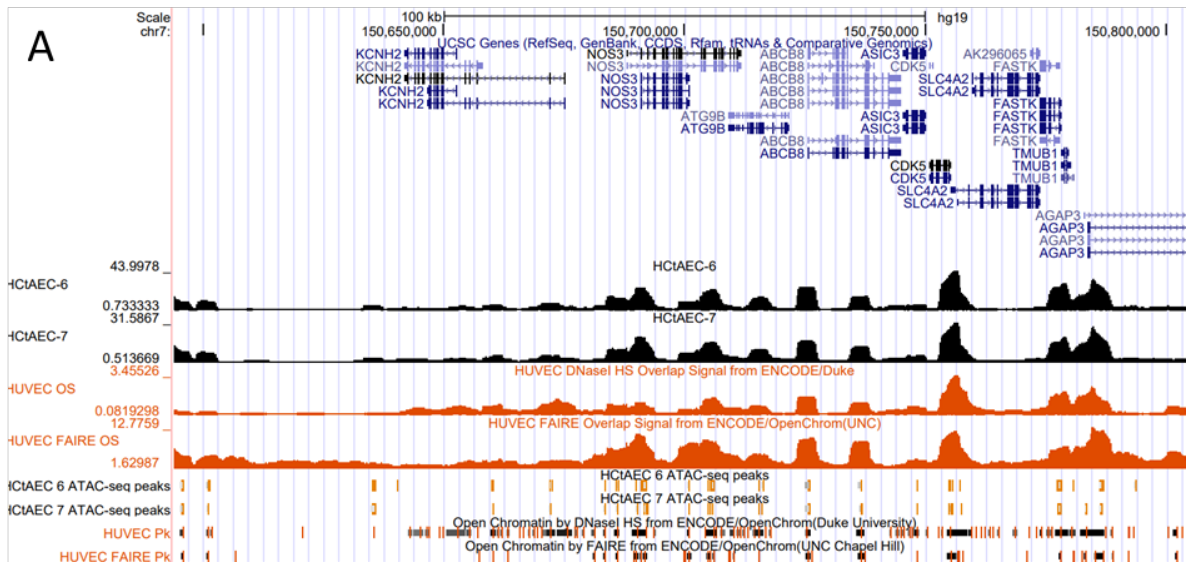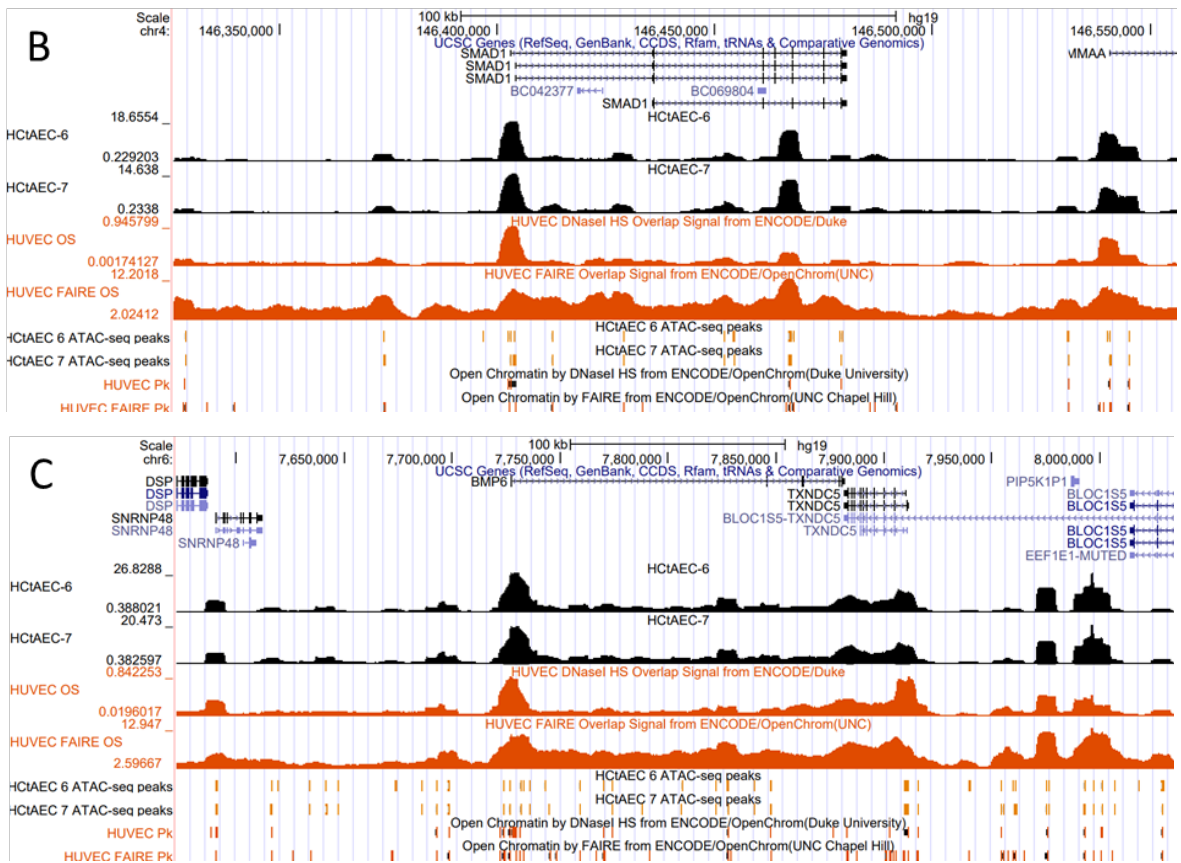
Figure 29. Visualisation of Omni-ATAC peaks in HCtAEC at 50M reads per sample with ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks at selected loci, including a) *NOS3*, b) *SMAD1*, and c) *BMP6* loci. Omni-ATAC peaks in HCtAEC are displayed in black, while ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks are displayed in orange. There are additional Omni-ATAC peaks in HCtAEC-6 and HCtAEC-7 samples compared to ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks.

In summary, the data demonstrated 50M is the optimal read depth based on the conventional FRiP score analysis and the conditions that produced Omni-ATAC samples HCtAEC-6 and HCtAEC-7 were selected to be the optimal (**Table 8**).

Table 8. Summary of the fraction of HCtAEC Omni-ATAC reads in peaks.

Depth	Sample ID	Reads in peaks	Total reads used to call peaks	FRiP score
30M	HCtAEC-6	9774059	31135955	0.31
	HCtAEC-7	6034746	24515355	0.25
50M	HCtAEC-6	20678424	57800069	0.36
	HCtAEC-7	14587197	49526467	0.29

100

4.4   Discussion

To determine the optimal experimental protocol for vascular endothelial cells, ATAC-seq samples were prepared following two approaches: a standard ATAC-seq protocol and an Omni-ATAC protocol. I tested eight conditions in total varying parameters such as cell number, lysis, and mixing procedure. In a sequential manner to select optimal conditions, I increased sequence read depth from 1M to 30M to 50M. I set up a robust quality control (QC)/analytical pipeline for ATAC-seq data to be able to perform comparisons to published data sets.

Opimising the cell number is critical for generating high-quality library. ATAC-seq is often robust to relatively minor variations of cell number. Too few cells cause over-digestion of chromatin and creates a larger fraction of reads that map to inaccessible regions of the genome (noise), while using too many cells causes under-digestion and creates high molecular weight fragments, which can be very difficult to sequence. Optimising the lysis duration is critical to ensuring that the majority of the cell membranes are lysed so that the transposition reagents can access the nucleus, while preventing over-lysis of the nuclear membrane, which can result in leakage of nuclear contents and subsequent increase in background signal. The different lysis conditions are important criteria for generating high-quality ATAC-seq libraries. Optimising mixing procedure is critical as this might increase fragment yield [135].

As part of the QC/analytical pipeline, FastQC reports indicated certain corrections were needed before proceeding with further analyses. I implemented trimming the first 13 nucleotides of the start of the read. Some of the most common reasons include overrepresented sequences, biased fragmentation, biased composition libraries or the presence of aggressively trimmed libraries. The adapter content plot showed the presence of nextera transposase in 30% of all reads. This was corrected by adapter trimming. The most common reasons include proportion of the insert sizes are shorter than the reads length. The Kmer content plot showed Kmer bias at the start of the read. This was corrected by trimming reads from the start and the end of the reads. The most common reasons for kmer bias include, the presence of overrepresented sequences, which appear as sharp spikes of enrichment at a single point in the sequence, rather than a progressive or broad enrichment. Kmer bias at the start of the reads could appear due to libraries which derive from random priming due to an incomplete sampling of the possible random primers [285].

The modified FRiP score from HCtAEC ATAC-/Omni-ATAC reads in HUVEC DNaseI-seq and FAIRE-seq peaks from ENCODE was used to prioritise conditions for increased sequence read depth. However, there are some limitations using the modified FRiP score instead of the conventional FRiP score, which include comparison different cell types and methods to determine chromatin accessibility. Samples HCtAEC-6 and HCtAEC-7 had the highest FRiP scores at 30M and were selected to sequence at 50M reads. Omni-ATAC [338] was the methodology of choice using 50K cells, and mixing using thermomixer with 1000 RPM continuously for 30 minutes with 2xTD buffer.

Overall, the HCtAEC ATAC-seq libraries showed strong overlap between HUVEC DNaseI and FAIRE-seq ENCODE data and were able to identify unique accessible chromatin peaks. Optimisation of primary endothelial cells for ATAC-seq will allow for further research into vascular-specific accessible chromatin and aid studies into gene regulation in these cells.

5.   **Maps of open chromatin sites in vascular endothelial and smooth muscle cells**

5.1   Introduction

Blood vessels walls are composed of two major cell types, endothelial cells (ECs) and vascular smooth muscle cells (VSMCs) (**Figure 30**) [340]. These cell types play an important role in atherosclerotic disease.
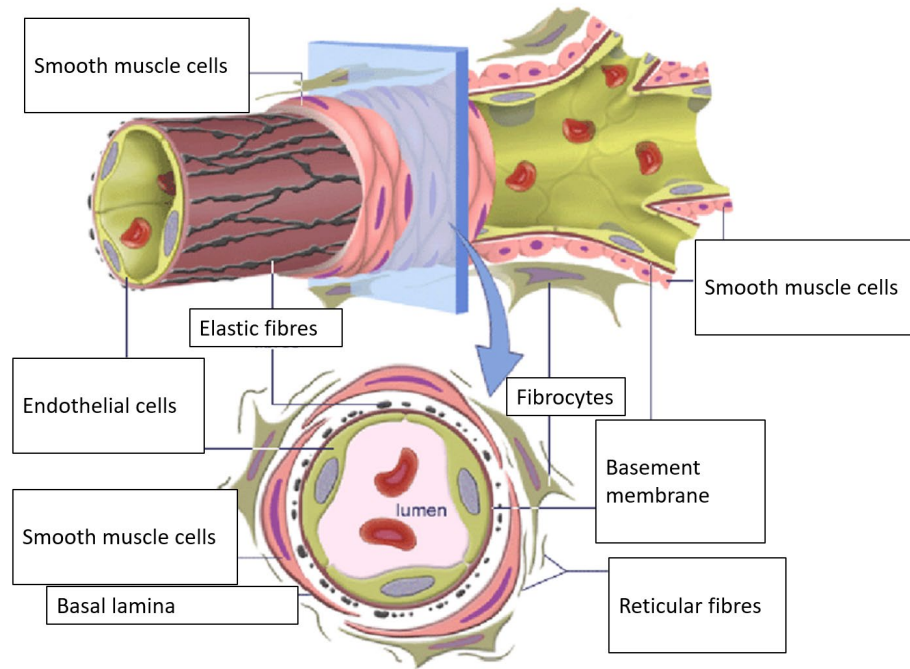
Figure 30. Blood vessel structure. Vascular endothelial cells (endothelium) form a thin layer on the interior surface of all the vessels, separated from the surrounding outer layers by a basement membrane (basal lamina). The largest blood vessels are veins and arteries, which have thick wall of connective tissue and many layers of smooth muscle cells (adapted from Goyal *et al.,* 2013).

Atherosclerosis is a complex process with initiation and progression of inflammatory responses that facilitate changes in the blood vessel walls [341, 342]. This causes the development of fatty streaks in the blood vessel walls, which gradually develop into characteristic plaques. Rupture of this plaque causes thrombosis, which results in partial or total occlusion. Endothelial dysfunction underlies vascular impairment, hypertension, hypercholesterolemia, and diabetes [343-345]. The effects of endothelial dysfunction on VSMC are reduction of nitric oxide bioavailability and augmentation of vasoactive constrictors released from the endothelium [343, 344]. VSMC proliferation is beneficial in atherogenesis, while dysregulated VSMC proliferation contributes to formation of plaque and inflammation [346, 347]. Endothelial dysfunction plays an important role in cellular glucose uptake, impairment of nitric

oxide-dependent vasodilatation, inflammation enhanced oxidative stress, leading eventually to atherosclerosis.

HUVECs have been used as a model in vascular endothelium research and have been useful to study monocyte adhesion to the endothelium, endothelial damage and repair, the potential impact of atherosclerosis in early stages and atherosclerosis progression [348]. Cultured HUVECs are subject to contact inhibition and limitations in growth density, as cell proliferation is strongly inhibited once the culture has reached confluence [349]. Because different components of the extracellular matrix can promote the formation of capillary-like structures in HUVECs, they are useful as angiogenesis models, although they do not reflect all the steps of physiological angiogenesis. These models are suitable for *in vitro* screening of angiostatic molecule activity, to evaluate the ultrastructure of capillary formation, the matrix synthesis of endothelial cells and tubular morphogenesis. Despite the widespread use of HUVECs in vascular research, it is now known that vascular cells show a high degree of heterogeneity, allowing biological adaptation to local needs [80].

The aim of this study was to explore the chromatin landscape of vascular endothelial and smooth muscle cells and how these might relate to vascular biology. ATAC-seq was performed to generate genome-wide open chromatin maps of HCAEC and HCASMC, representing cell types which both play an important role in CAD pathophysiology.

5.2   Methods

5.2.1   ATAC-seq and peak intersection

DNA from HCAEC and HCASMC was quantified using the Qubit and TapeStation as detailed in Chapter 3, section 3.8. multiBamSummary [293] was used to compute the read coverages of HCAEC and HCASMC samples as detailed in Chapter 3, section 3.20. The BEDTools [289] intersect function was used to identify all reads in peaks (HCAEC ATAC-seq reads in peaks; HCASMC ATAC-seq reads in peaks) as detailed in Chapter 3, section 3.21. ATAC-seq peaks in HCAEC and HCASMC were examined for overlap with DNaseI HS from ENCODE/OpenChrom (Duke University) and FAIRE from ENCODE/OpenChrom (UNC Chapel Hill), respectively. ATAC-seq peaks in HCAEC were examined for overlap with HUVEC Chromatin State Segmentation by HMM using data from ENCODE/Broad with BEDTools intersect function with default arguments (minimum 1 bp overlap) as detailed in

Chapter 3, section 3.23 [294, 295]. ATAC-seq peaks in HCAEC were examined for overlap with HUVEC histone modifications signatured using ChIP-seq data from ENCODE/Broad Institute with BEDTools intersect function with default arguments (minimum 1 bp overlap) as detailed in Chapter 3, section 3.24 [295, 297-299, 350]. Annotation of the ATAC-seq peaks in HCAEC and HCASMC was performed using ChIPseeker as detailed in Chapter 3, section 3.18. ATAC-seq peaks in the HCAEC and HCASMC were evaluated for biological themes among gene clusters, using compareCluster from ChIPseeker as detailed in Chapter 3, section 3.18 [291]. TF binding motif analysis of ATAC-seq peaks in HCAEC and HCASMC datasets was performed using HOMER as detailed in Chapter 3, section 3.19 [292]. Overlap of ATAC-seq peaks in HCAEC and HCASMC with DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3) was performed to identify unique and vascular-specific accessible chromatin peaks in HCAEC and HCASMC compared to the full set of 125 cell types. Annotation of the ATAC-seq peaks in the unique and vascular-specific HCAEC and HCASMC was performed using ChIPseeker as detailed in Chapter 3, section 3.18 [291]. ATAC-seq peaks in the unique and vascular-specific HCAEC were evaluated for biological themes among gene clusters, using compareCluster from ChIPseeker as detailed in Chapter 3, section 3.18. TF binding motif analysis of ATAC-seq peaks in the unique and vascular-specific HCAEC was performed using HOMER as detailed in Chapter 3, section 3.19 [292].

5.3 Results

5.3.1 ATAC-seq

ATAC-seq in three HCAEC and two HCASMC samples was performed as detailed in Chapter 3, section 3.5. **Table 9** shows the amount of purified ATAC-seq DNA per sample. As before, HCAEC and HCASMC samples resulted in fragments with no clear size periodicity (**Figure 31 A and B**).

Table 9. DNA concentration of ATAC-seq libraries determined by Qubit.

Sample ID	Concentration [pg/ul]
HCAEC-1	2.7
HCAEC-2	2.2
HCAEC-3	99.4
HCASMC-1	1.0
HCASMC-2	0.9
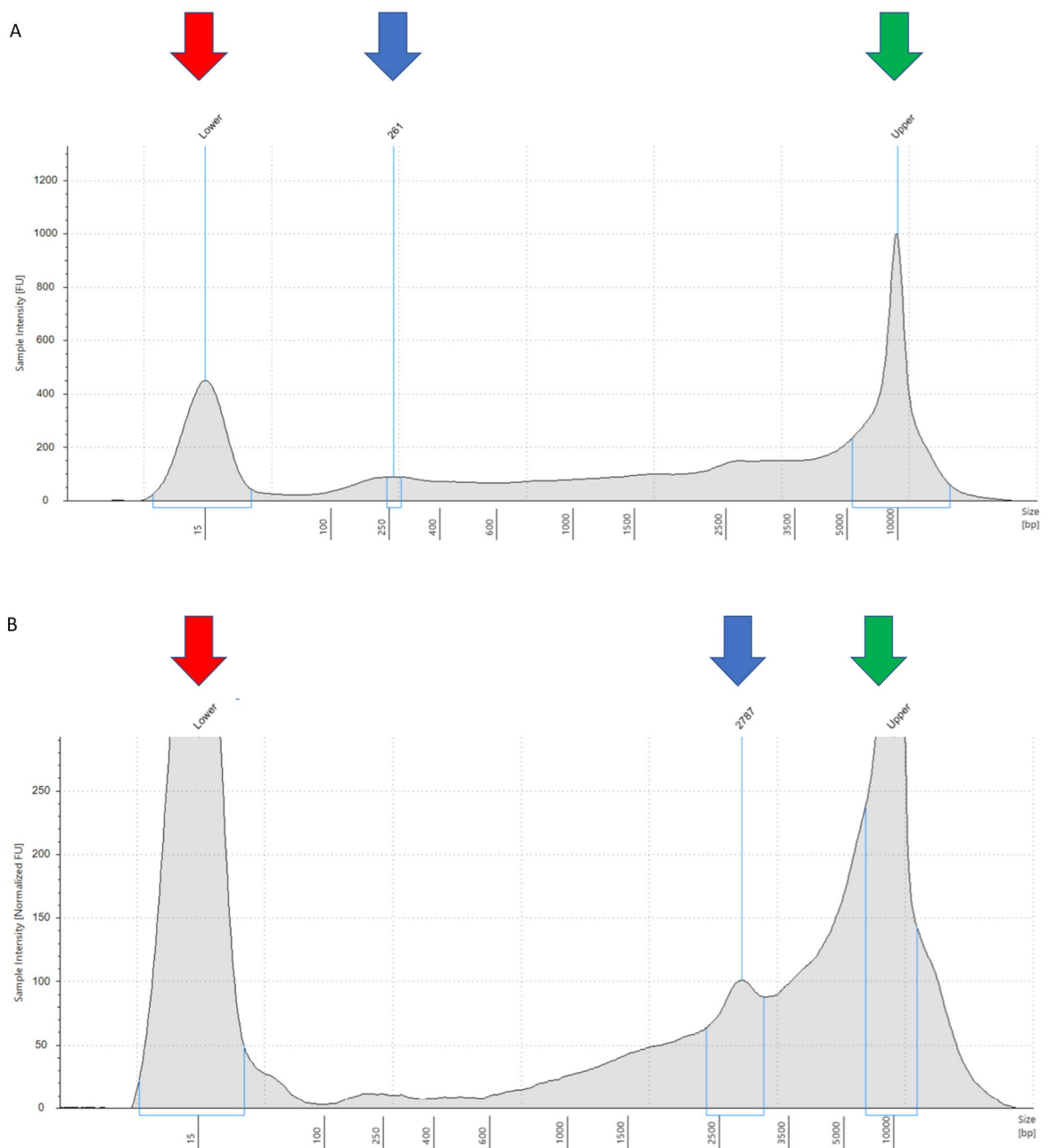
Figure 31. Electropherogram of fragment size distribution of ATAC-seq libraries determined by TapeStation. The X-axis represents the product size (bp) and the Y-axis represents the arbitrary fluorescence intensity (FU). Red and green arrows represent lower and upper size markers, respectively. A) The figure shows poor fragment size distribution of HCAEC sample. Blue arrow at 261 bp indicate fragment size of peak, which may represent mononucleosome fragments. B) The figure shows poor fragment size distribution of HCASMC sample. The blue arrow at 2787 bp indicate fragment size of peak, which may represent multinucleated fragments.

FastQC analysis did not identify any sub-standard sample (Appendix, **Figure 76** and **Figure 78**) show representative example of the FastQC reports generated.

107

5.3.2 Correlation analysis of HCAEC and HCASMC ATAC-seq samples

Pearson correlation coefficients from the three HCAEC ATAC-seq samples ranged from 0.92 to 0.98, demonstrating the desired strong correlation between samples (**Figure 32**). This was also observed from the two HCASMC ATAC-seq samples with a correlation coefficient of 0.98 (**Figure 33**).
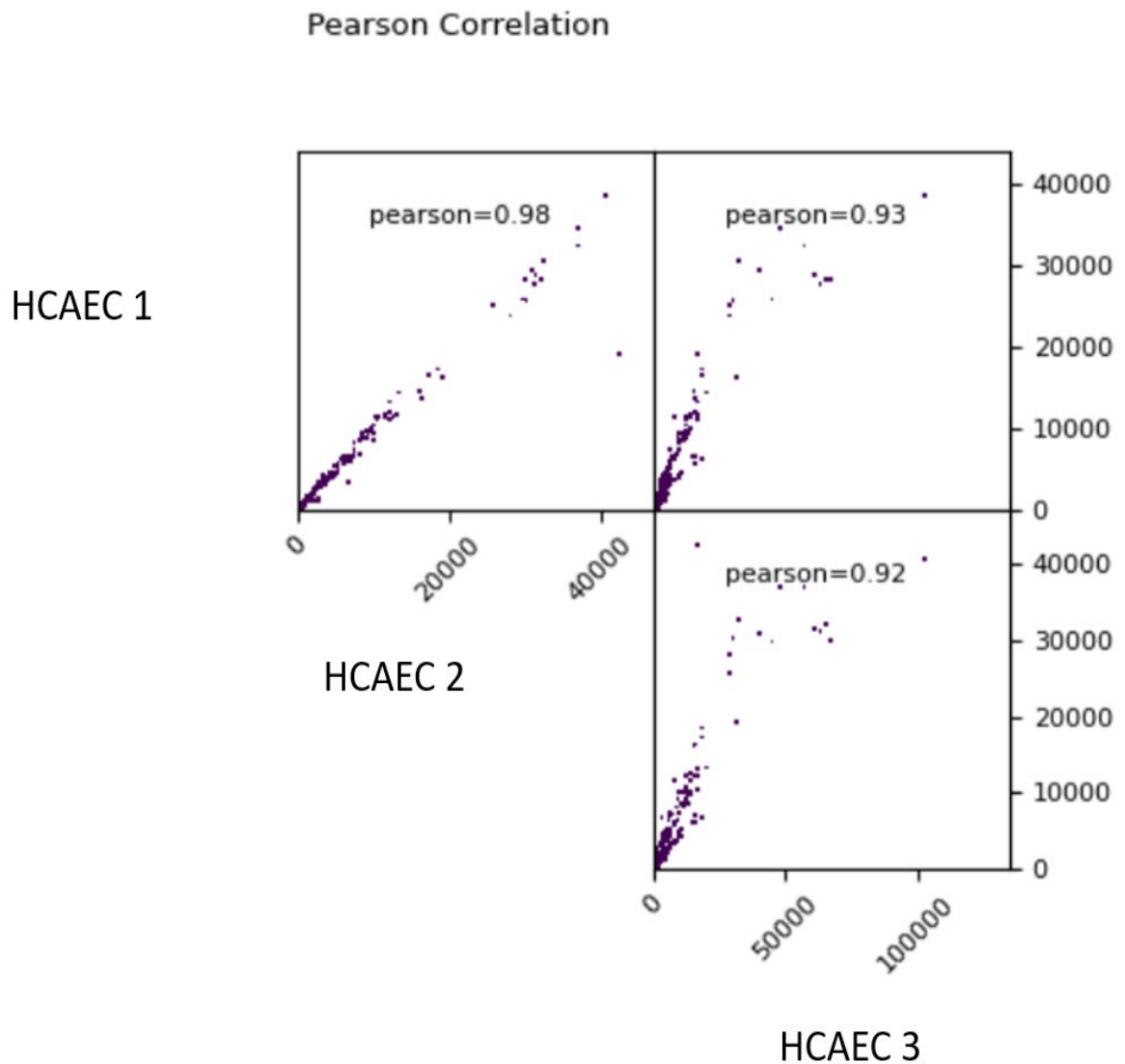
Figure 32. Pearson correlation coefficient of read coverages in HCAEC ATAC-seq samples. The y-axis and x-axis represent the number of fragments in HCAEC ATAC-seq samples. Each dot represents one genomic region.

Figure 33. Pearson correlation coefficient of red coverages in HCASMC ATAC-seq samples. The y-axis and x-axis represent the number of fragments in HCASMC ATAC-seq samples. Each dot represents one genomic region.

Pearson correlation coefficients from HCAEC and HCASMC ATAC-seq samples were examined together. For both HCAEC and HCASMC replicates clustered together but the two cell types clustered apart reflecting distinct differences in accessible chromatin (**Figure 34**).
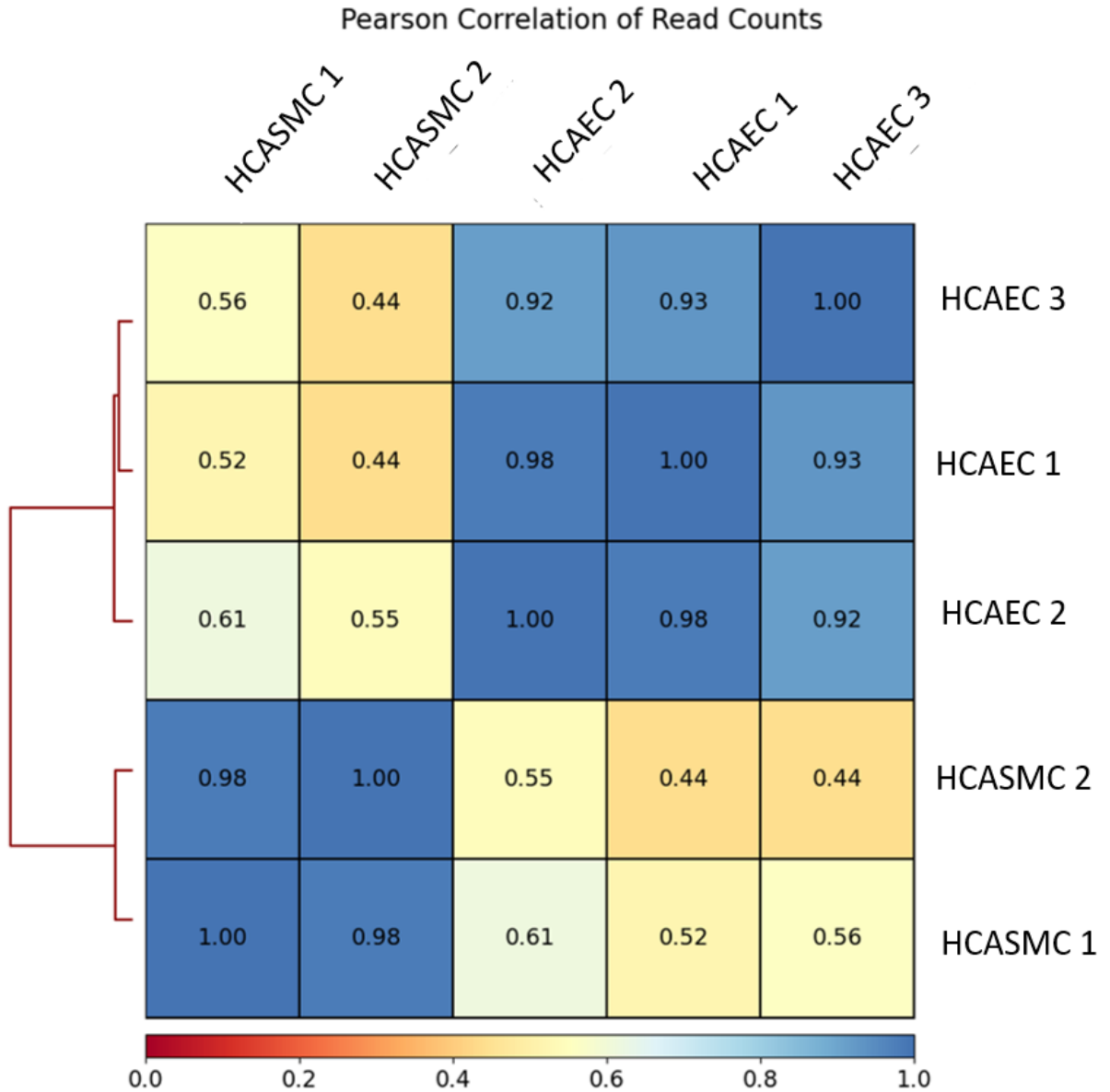
Figure 34. Pearson correlation coefficients of read coverages in HCAEC and HCASMC ATAC-seq samples. The y-axis represents the HCAEC and HCASMC ATAC-seq samples, while x-axis represents the correlation coefficient.

5.3.3   Comparison of ATAC-seq profiles in HCAEC and HCASMC

The FRiP score for HCAEC ATAC-seq libraries ranged from 0.08 to 0.14, below the recommended 0.2 standard, indicating some background noise (**Table 10**). The FRiP score for HCASMC ATAC-seq libraries was 0.35, meeting the recommended 0.2 standard (**Table 11**). Given the strong correlation between replicates, I combined the ATAC-seq data from each set of replicate samples to obtain a comprehensive set of peaks per cell type. This resulted in 86,811 and 209,743 ATAC-seq peaks for HCAEC and HCASMC, respectively.

Table 10. Fraction of HCAEC ATAC-seq reads in peaks.

Sample ID	Number of peaks	Reads in peaks	Total reads used to call peaks	FRiP score
HCAEC-1	61956	4596550	55594237	0.08
HCAEC-2	68693	5274199	53113993	0.1
HCAEC-3	73621	8034858	57906565	0.14
HCAEC -merged	86811	21514460	166614795	0.13

Table 11. Fraction of HCASMC ATAC-seq reads in peaks.

Sample ID	Number of peaks	Reads in peaks	Total reads used to call peaks	FRiP score
HCASMC-1	206048	12599372	35925512	0.35
HCASMC-2	160875	8074115	22886840	0.35
HCASMC-merged	209743	22917455	58812352	0.39

I then examined the correlation between these ATAC-seq peaks in HCAEC and HCASMC cells with ENCODE FAIRE-seq and DNase-seq endothelial (HUVEC) accessible chromatin datasets to determine differences between the methodologies and cell types. As expected, the HCAEC ATAC-seq peak profile shared a large proportion of accessible chromatin regions with ENCODE HUVEC DNaseI-seq (66.6%) and FAIRE-seq (67.6%) profiles, while the overlap of HCASMC ATAC-seq profiles was significantly lower with both HUVEC DNaseI-seq (31.7%) and FAIRE-seq (32.7%) (**Table 12**). Differences both subtle (e.g., varying peak height) and big (presence/absence of peaks) were observed upon inspection of individual loci, as shown for three example genes, including Platelet Derived Growth Factor Subunit B (*PDGFB*), SMAD Family Member 3 (*SMAD3*), and Transforming Growth Factor Beta 1 (*TGFB1*) (**Figure 35 A, and C**).

Table 12. Overlap of ATAC-seq peaks in HCAEC and HCASMC with ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks.

Datasets	HCAEC ATAC-seq		Datasets	HCASMC ATAC-seq	
	N	%		N	%
HCAEC ATAC-seq peaks/HUVEC FAIRE-seq peaks	58658	67.6	HCASMC ATAC-seq peaks/HUVEC FAIRE-seq peaks	68518	32.7
HCAEC ATAC-seq peaks/HUVEC DNaseI-seq peaks	57832	66.6	HCASMC ATAC-seq peaks/HUVEC DNaseI-seq peaks	66426	31.7
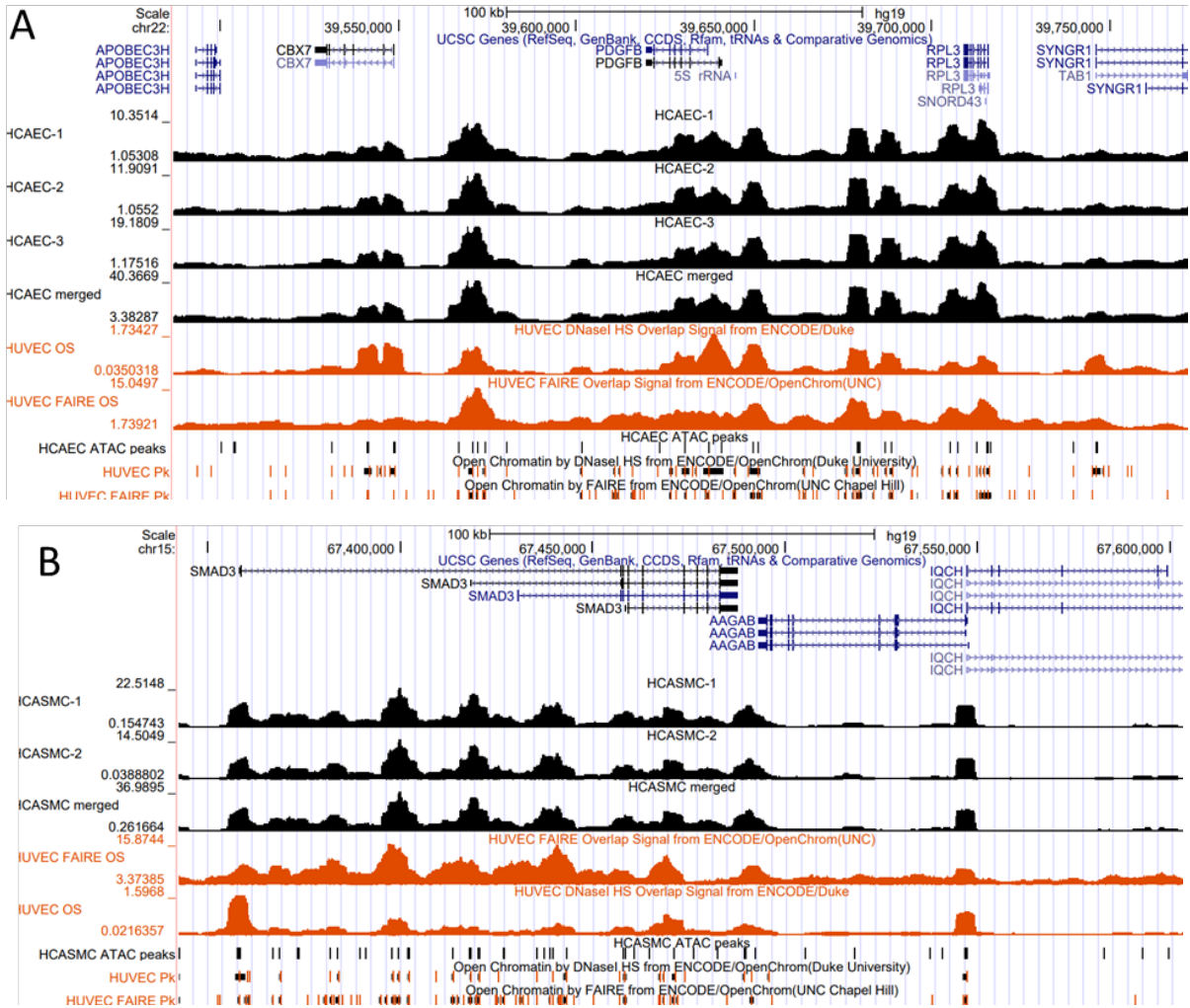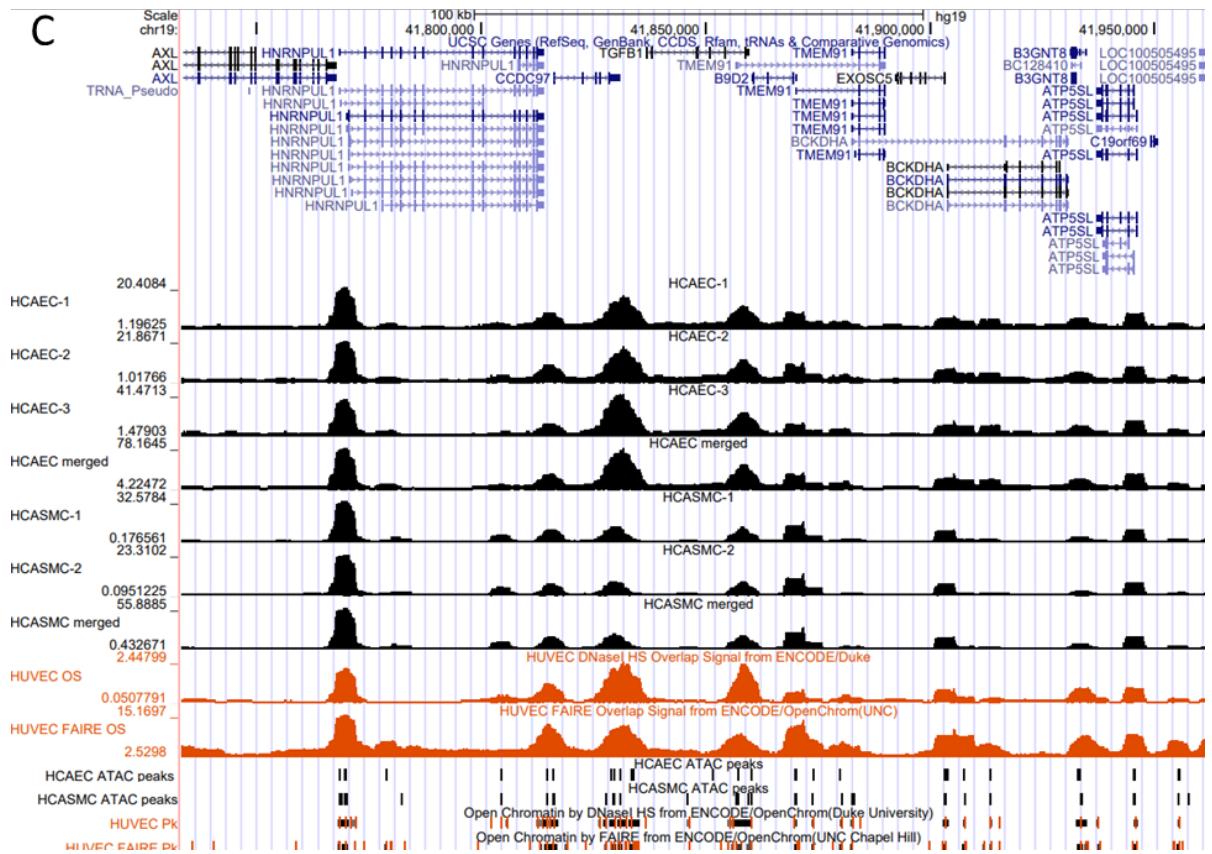
Figure 35. Overlap of ATAC-seq peaks in HCAEC and HCASMC with ENCODE HUVEC DNaseI-seq/FAIRE-seq peaks. The black bars on the two HUVEC DNaseI-seq/FAIRE-seq peaks indicate the span of the signal for the peak. a) maps of open chromatin regions in HCAEC at the *PDGFB* locus. The figure above shows the read alignments (HCAEC-1, HCAEC-2, HCAEC-3, and HCAEC-merged) and open chromatin regions (HCAEC ATAC peaks) in black. HCAEC ATAC-seq data and ENCODE data share a large portion of accessible chromatin annotations. b) maps of open chromatin in HCASMC at the *SMAD3* locus. The figure above shows the read alignments (HCASMC-1, HCASMC-2, and HCASMC-merged) and open chromatin regions (HCASMC ATAC peaks) in black. HCASMC ATAC-seq data and ENCODE data share significantly lower portion of accessible chromatin annotations. c) maps of open chromatin in HCAEC and HCASMC at the *TGFB1* locus. The read alignments (HCAEC-1, HCAEC-2, HCAEC-3, HCAEC-merged; HCASMC-1, HCASMC-2, HCASMC-merged) and open chromatin regions (HCAEC ATAC peaks; HCASMC ATAC peaks) in black. There are differences in the HCAEC and HCASMC ATAC-seq peaks.

5.3.4 Distribution of chromatin state at open chromatin regions in HCAEC ATAC-seq peaks

Despite a large proportion of overlap between ATAC-seq peaks in HCAEC and DNaseI-seq/FAIRE-seq peaks in HUVEC, there were differences between the datasets. To explore if any specific features were affected by the three different methods used to map accessible chromatin, ENCODE chromatin state annotations were examined. Chromatin states associated with active promoters and strong enhancers were enriched in HCAEC ATAC-seq peaks compared to HUVEC DNaseI-seq/FAIRE-seq peaks, while chromatin states associated with heterochromatin-low signal were enriched in HUVEC FAIRE-seq peaks compared to HCAEC ATAC-seq peaks and HUVEC DNaseI-seq. In addition, chromatin states associated with polycomb-repressed were enriched DNaseI-seq peaks compared to HCAEC ATAC-seq peaks and FAIRE-seq peaks (**Figure 36**).
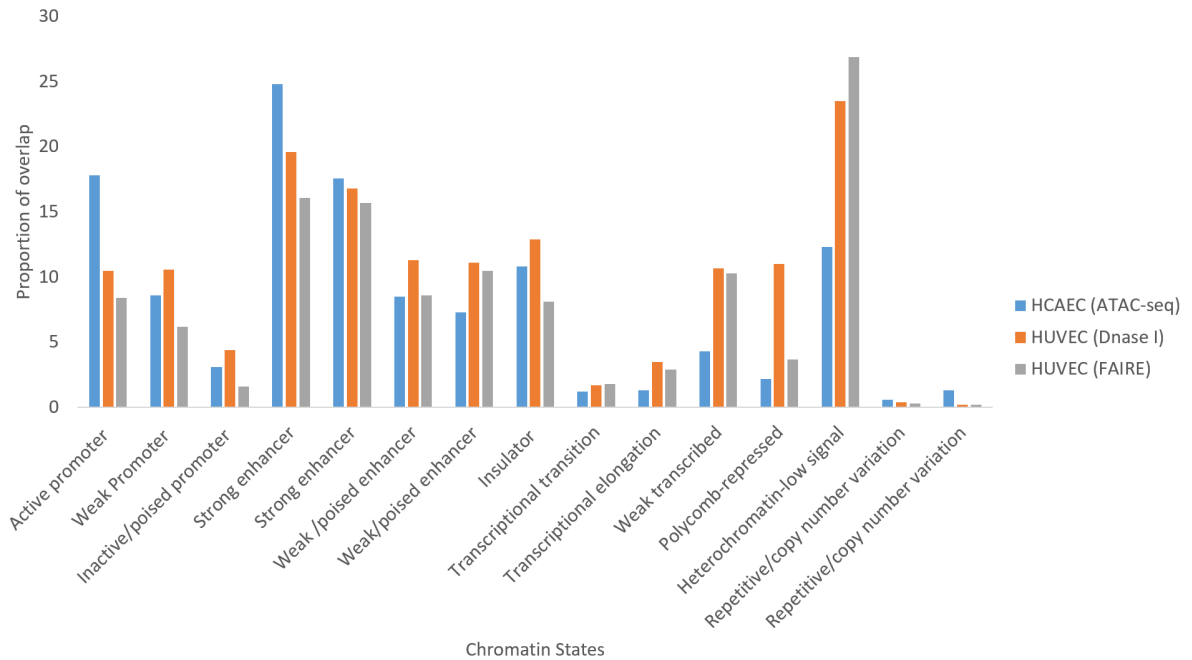
Figure 36. Chromatin state distribution between ATAC-seq/DNaseI-seq/FAIRE-seq/methods. Chromatin states associated with active promoter and strong enhancer were enriched in HCAEC ATAC-seq peaks compared to HUVEC DNaseI-seq/FAIRE-seq peaks. Chromatin states associated with heterochromatin-low signal were enriched in HUVEC FAIRE-seq peaks compared to HCAEC ATAC-seq peaks and HUVEC DNaseI-seq peaks.

5.3.5 Distribution of histone modifications at open chromatin region in HCAEC ATAC-seq peaks

To further assess differences and similarities between the three methods, ENCODE histone modifications were examined. Overall, a similar pattern of distribution of histone modifications was found in the overlapping HCAEC ATAC-seq peaks with HUVEC DNaseI-seq/FAIRE-seq peaks, which was also the case when examining unique HCAEC peaks i.e., absent HUVEC DNaseI-seq/FAIRE-seq peaks. However, certain histone modifications were over-represented in HCAEC ATAC-seq. Histone marks associated with active chromatin, including H3K4me1, H3K4me3, and H3K27ac were enriched in the overlapping HCAEC ATAC-seq peaks compared to HUVECs DNase-seq/FAIRE-seq peaks (**Figure 37 A**). In contrast, histone modifications associated with heterochromatin, including H3K27me3 were enriched in the overlapping HCAEC ATAC-seq peaks compared to HUVECs DNase-seq/FAIRE-seq peaks, while H3K9me3 were enriched in the unique HCAEC ATAC-seq peaks compared to HUVECs DNase-seq/FAIRE-seq peaks, respectively (**Figure 37 B**).
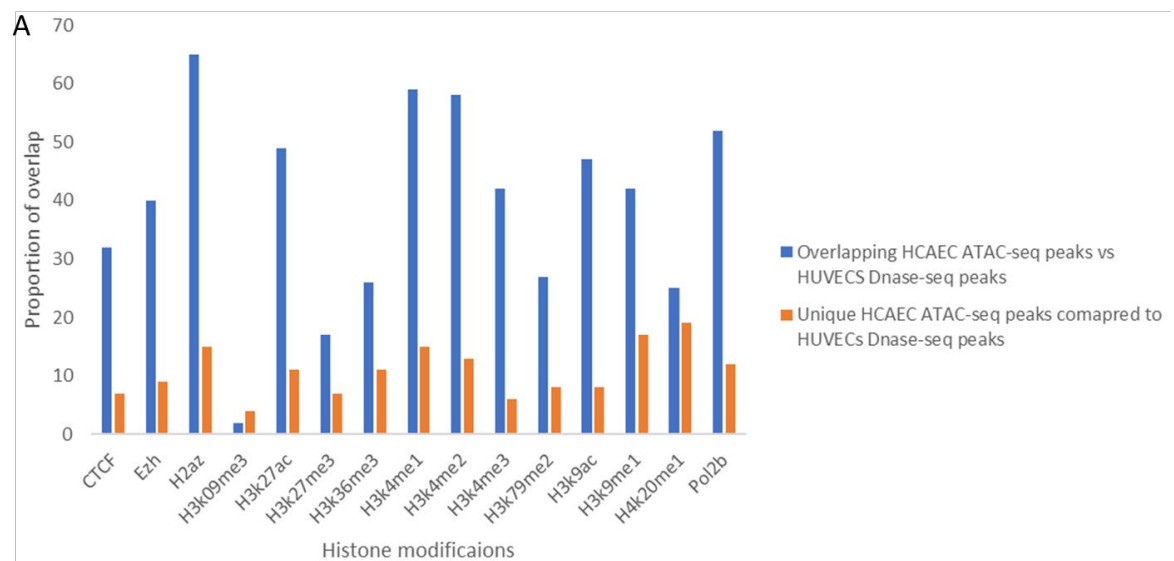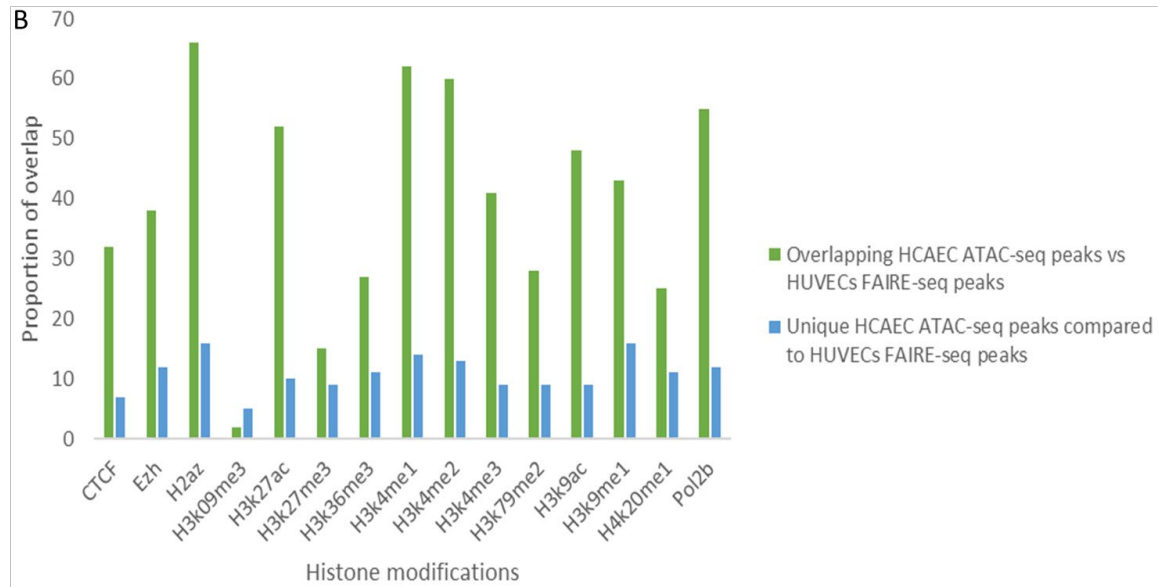
Figure 37. Histone modifications distribution between ATAC-seq/DNaseI-seq/FAIRE-seq/methods. Histone marks associated with active chromatin (H3K4me1, H3K4me3, and H3K27ac) were enriched in the overlapping HCAEC ATAC-seq peaks compared to HUVECs DNase-seq/FAIRE-seq peaks. Histone marks associated with inactive chromatin (H3K27me3 and H3K9me3) were enriched in the overlapping HCAEC ATAC-seq peaks compared to HUVECs DNase-seq/FAIRE-seq peaks, while H3K9me3 were enriched in the unique HCAEC ATAC-seq peaks compared to HUVECs DNase-seq/FAIRE-seq peaks, respectively.

5.3.6 Genomic landscape of ATAC-seq peaks in HCAEC and HCASMC

ATAC-seq peaks in HCAEC and HCASMC were correlated relative to the nearest genomic feature(s). In HCAEC, the largest proportion of peaks were found in promoters (33.9%), followed by those in distal intergenic (28.8%) and other intronic regions (20.3%). In HCASMC, the largest proportion of peaks were found in distal intergenic (39.0%), followed by those in intronic (27.2%), and promoter regions (15.4%) (**Figure 38 A**). Overall, the identified accessible chromatin regions were enriched around the TSS (**Figure 38 B and C**), which is consistent with the knowledge of such regions containing *cis*-regulatory elements.

116
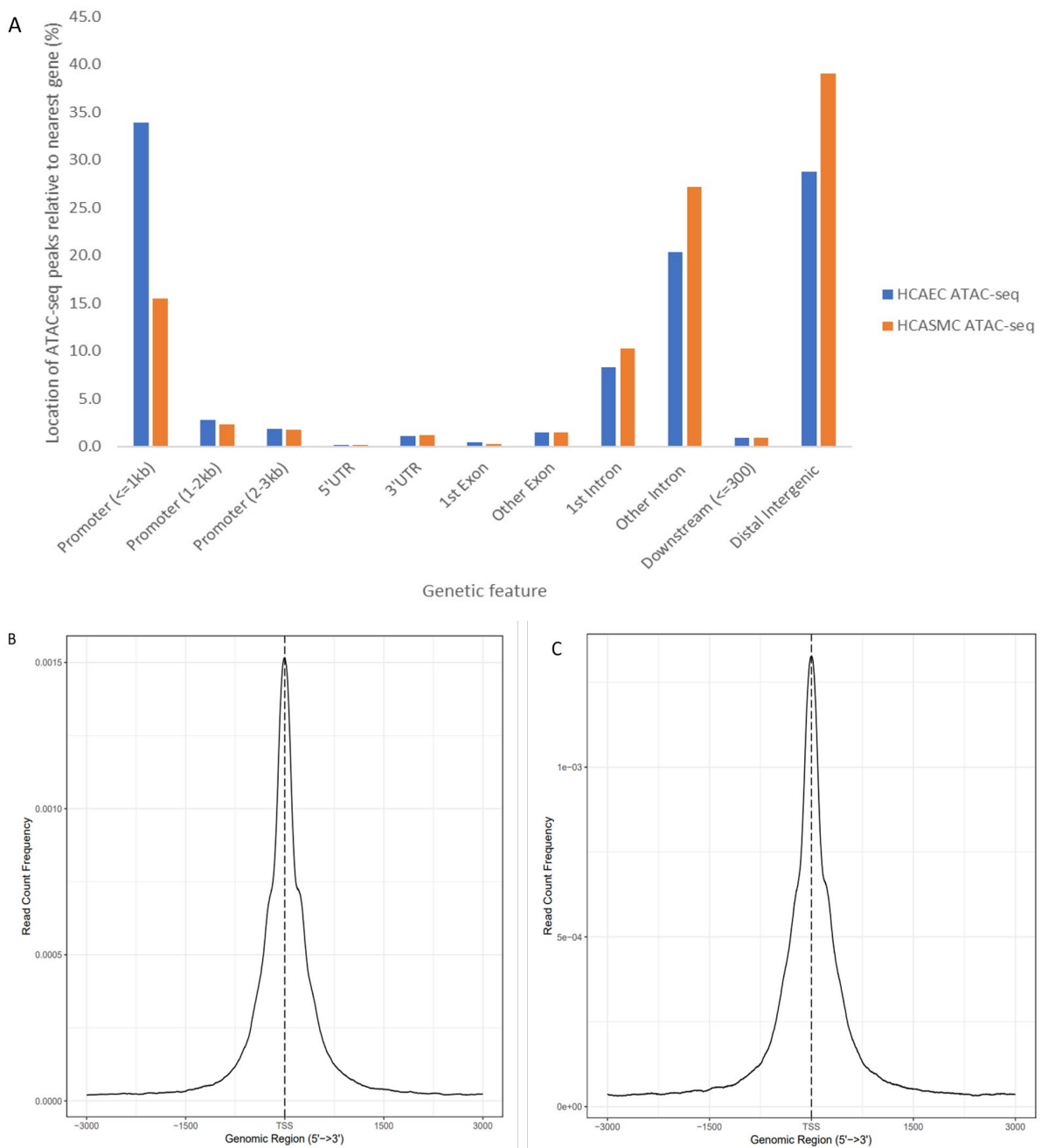
Figure 38. Distribution of accessible regions in HCAEC and HCASMC identified using ATAC-seq. (A) Downstream refers to where the centre of accessible chromatin is located less than 300 kb from the TSS. Distal intergenic regions refer to accessible region located more than 3 kb upstream from the TSS and more than 1 kb downstream from transcriptional termination sites (TTS). (B-C) Distribution of ATAC-seq peaks around the TSS for HCAEC and HCASMC, respectively. The centre of ATAC-seq peak was used to produce the distribution plots.

The nearest genes to accessible chromatin peaks were annotated using Chipseeker, which identified 6515 and 6971 genes surrounding ATAC-seq peaks in HCAEC and HCASMC, respectively. Several KEGG pathways showing enrichment in these gene sets were shared between HCAEC and HCASMC, including Rap1, MAPK, Hippo, Endocytosis, Oxytocin, Splingolipid signalling pathways. In contrast, Ras and Cellular senescence signalling pathways were specific to HCAEC, while cAMP, Vascular smooth muscle contraction, and Prolactin signalling pathways were specific to HCASMC (**Figure 39 A and B)**.
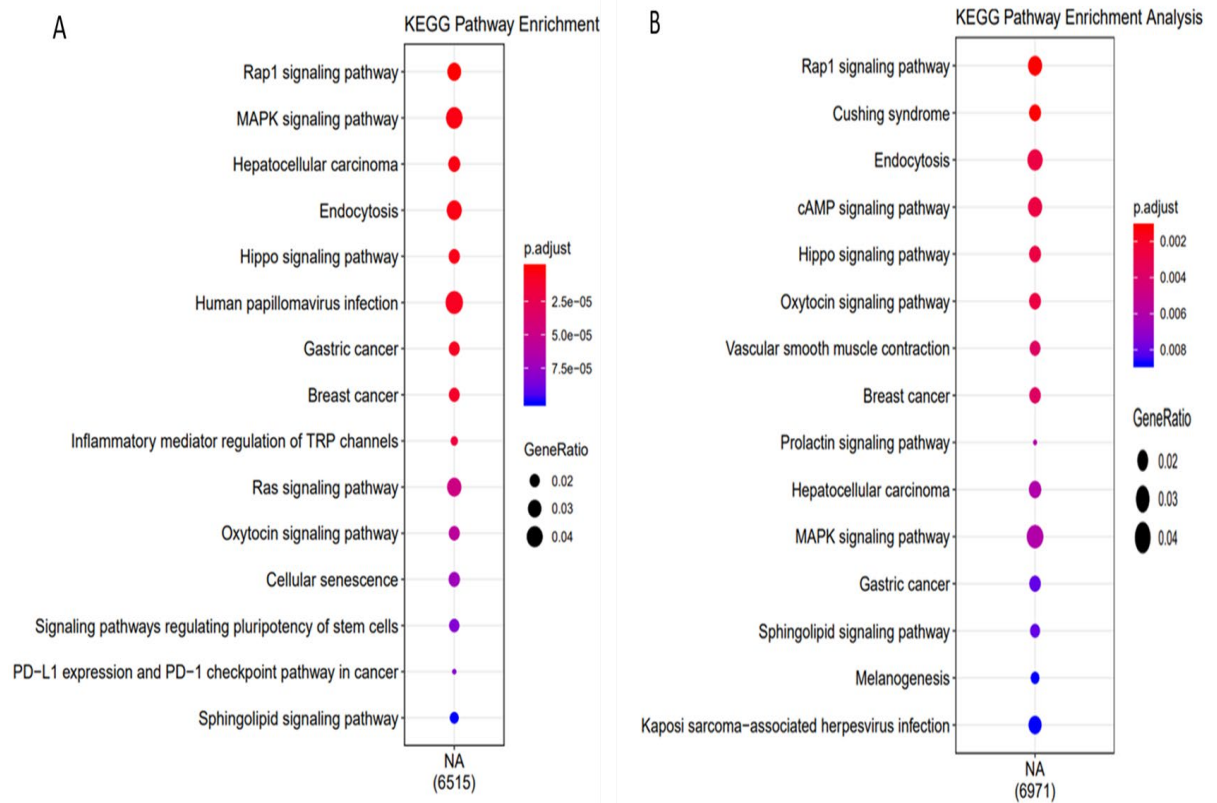
Figure 39. KEGG pathway enrichment analysis of genes surrounding vascular accessible chromatin in HCAEC and HCASMC, respectively. A-B) KEGG pathway enrichment analysis of genes surrounding ATAC-seq peaks in HCAEC and HCASMC, respectively. The colour of the dot represents the adjusted *p*-value and the size of the dot represents the gene ratio. The gene ratio represents the total number of genes associated with this pathway.

5.3.7 Enrichment of transcription factor binding motifs within ATAC-seq peaks in HCAEC and HCASMC

To complement the KEGG pathway analysis from genes near accessible chromatin peaks described above, I performed enrichment analysis for transcription factor binding motifs across ATAC-seq peaks in HCAEC and HCASMC using HOMER to identify regulatory motifs enriched in vascular accessible chromatin. There were a total of 358 motifs enriched in

HCAEC and 397 motifs enriched in HCASMC ATAC-seq peaks at a *p*-value $<1 \times 10^{-50}$ compared to the prevalence in the human reference genome (hg19). The most signficantly enriched motifs in HCAEC, include CTCF, BORIS, ATF3, JUNB, and Fra1 (**Figure 40**; **Table 13**). The most significantly eriched motifs in HCASMC, include BATF, Atf3, Fra1, AP-1, and JunB (**Figure 41**; **Table 14**). Motifs shared between the two cell types included ATF3, Fra1, and JunB. In contrast, CTCF and BORIS were spefic to HCAEC, while AP-1 was specific to HCASMC. CTCF plays a key role in maintaining higher order chromatin structure by facilitating the interactions between the regulatory sequences [284]. AP-1 plays an important role in a wide range of cellular processes, including proliferation, differentiation, and apoptosis [351]. Studies on gene regulation in endothelial cells demonstrated that the AP-1 transcription factors are implicated in angiogenesis and vascular development [352].
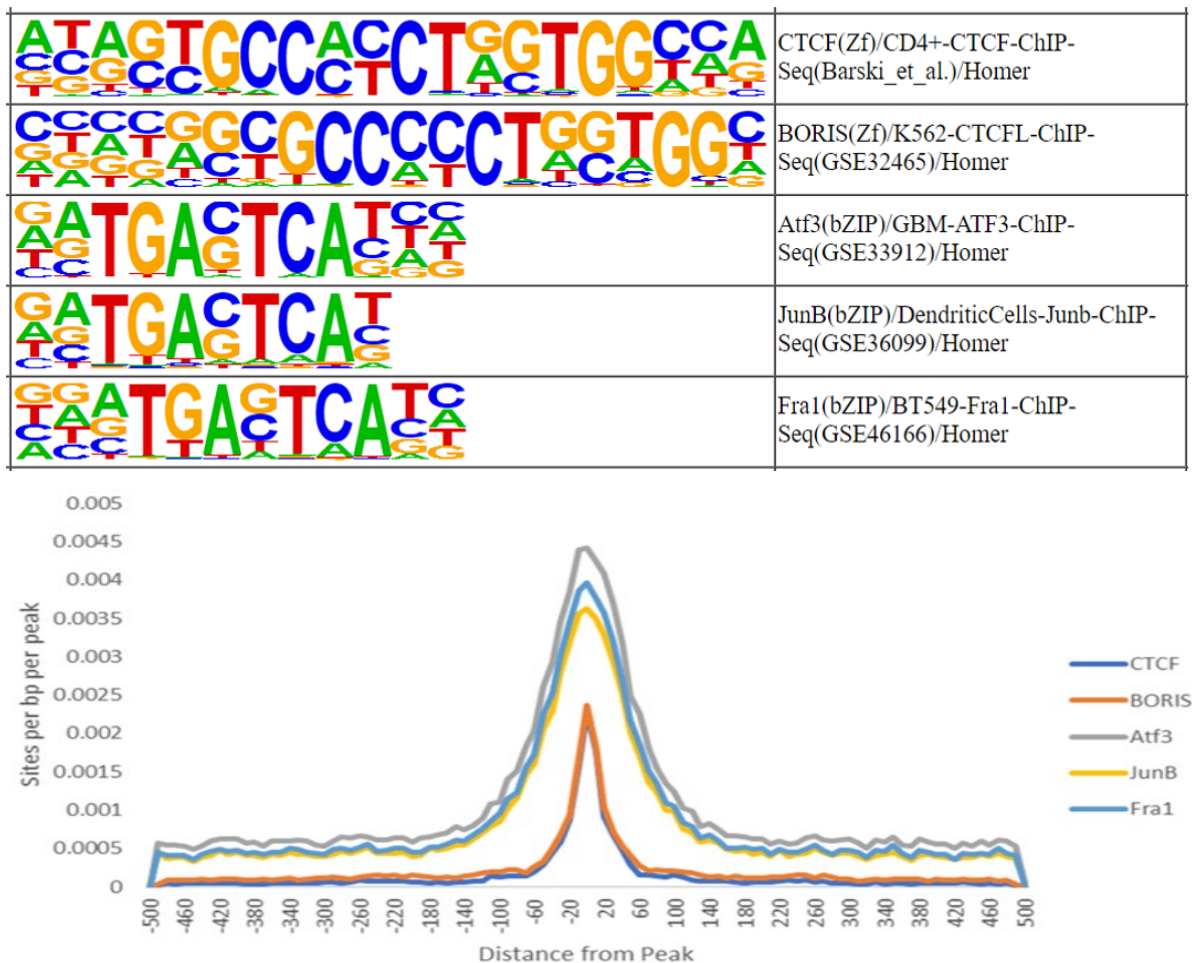
Figure 40. Histogram distribution for top five enriched motifs identified using known motif enrichment analysis in HCAEC ATAC-seq peaks.

Table 13. Top 5 enriched known motifs found in HCAEC ATAC-seq peaks.

Motif	*P*-value	Accessible chromatin with TF motif (%)	Inaccessible chromatin with TF motif (%)
CTCF	1x10-4961	6.9	0.2
BORIS	1x10-3444	7.1	0.5
Atf3	1x10-2640	8.3	1.2
JunB	1x10-2632	7.8	1.0
Fra1	1x10-2629	7.8	1.0

Motif logo	Annotation
TATGAGTCAT	BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer
GATGAGTCATCC	Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer
GGATGAGTCATC	Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer
ATGACTCATC	AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer
GATGACTCAT	JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer

Figure 41. Histogram distribution for top five enriched motifs identified using known motif enrichment analysis in HCASMC ATAC-seq peaks.

Table 14. Top 5 enriched known motifs found in HCASMC ATAC-seq peaks.

Motif	*P*-value	Accessible chromatin with TF motif (%)	Inaccessible chromatin with TF motif (%)
BATF	1x10-10821	13.6	1.8
Atf3	1x10-10743	13.5	1.8
Fra1	1x10-10686	12.3	1.4
AP-1	1x10-10509	14.1	2.1
JunB	1x10-10242	12.3	1.5

5.3.8 Genomic landscape of the vascular-specific and unique ATAC-seq peaks in HCAEC and HCASMC, respectively

I examined the intersection of HCAEC and HCASMC ATAC-seq datasets with the DNaseI Hypersensitivity Clusters in 125 cell types from the ENCODE (V3). This dataset includes several closely related cell types and many non-related cell types to ECs and SMCs. This analysis was performed to identify accessible chromatin regions unique to HCAEC and HCASMC datasets, respectively. I found 12,233 unique HCAEC accessible chromatin peaks (total ATAC-seq peaks excluding all overlapping endothelial ENCODE DNaseI-seq peaks) and 21,310 unique HCASMC accessible chromatin peaks (total ATAC-seq peaks excluding all overlapping smooth muscle ENCODE DNaseI-seq peaks). As several ENCODE cell types are closely related to those in this study, I also sought to identify all vascular-specific peaks by repeating the previous analysis but excluding related vascular cell types from the 125 ENCODE dataset. I found 17,570 vascular-specific accessible chromatin peaks (total ATAC-seq peaks excluding all overlapping non-vascular endothelial ENCODE DNaseI-seq peaks) and 23,433 vascular-specific accessible chromatin peaks (total ATAC-seq peaks excluding all overlapping non-vascular smooth muscle ENCODE DNaseI-seq peaks) (**Figure 42**) [353].
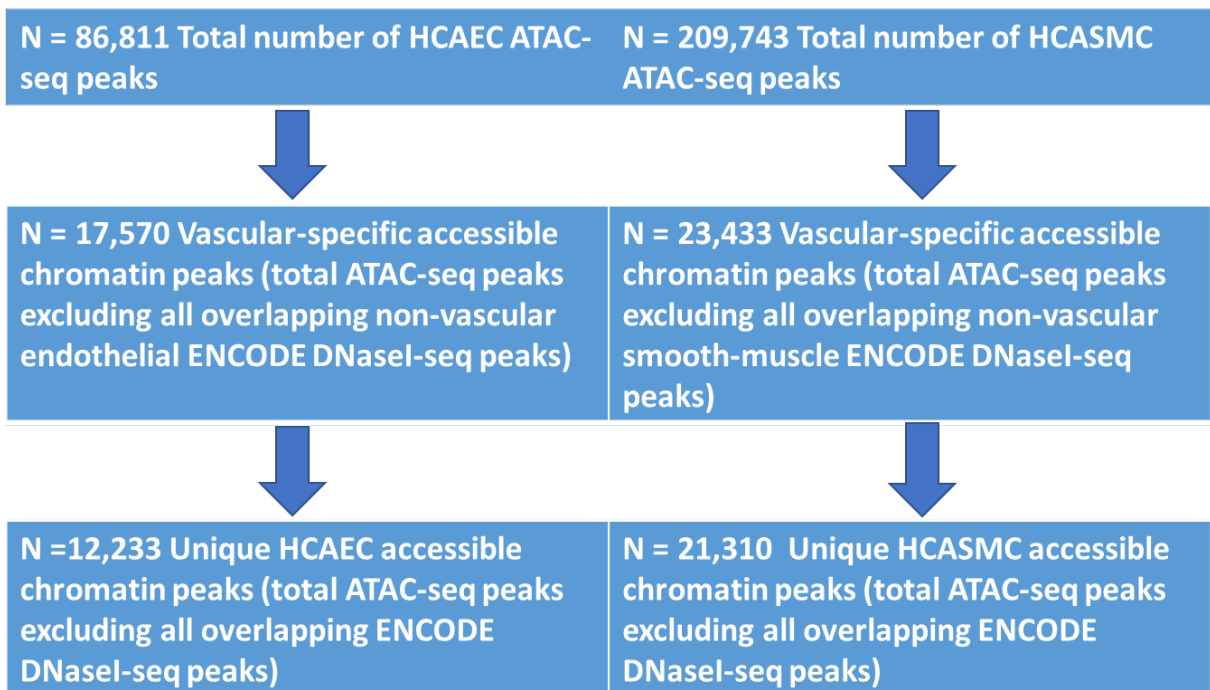
Figure 42. Total, vascular-specific and unique accessible chromatin peaks in HCAEC and HCASMC, respectively at each stage of the analysis as well as cell types were eliminated from the 125 ENCODE set.

ATAC-seq peaks within vascular-specific accessible chromatin and unique accessible chromatin peaks in HCAEC were correlated relative to the nearest genomic feature(s). In the vascular-specific accessible chromatin peaks a greater proportion was located outside of promoters: distal intergenic (44.0%), other intron (31.6%) and intron 1 (11.2%). In the unique HCAEC accessible chromatin peaks this was even more evident: distal intergenic (47.6%), other intron (32.4%) and intron 1 (10.3%) (**Figure 43)**. Different pattern of distribution was observed in the total accessible chromatin peaks, where the largest proportion was found in promoters, as detailed in Chapter 5, section 5.3.6.
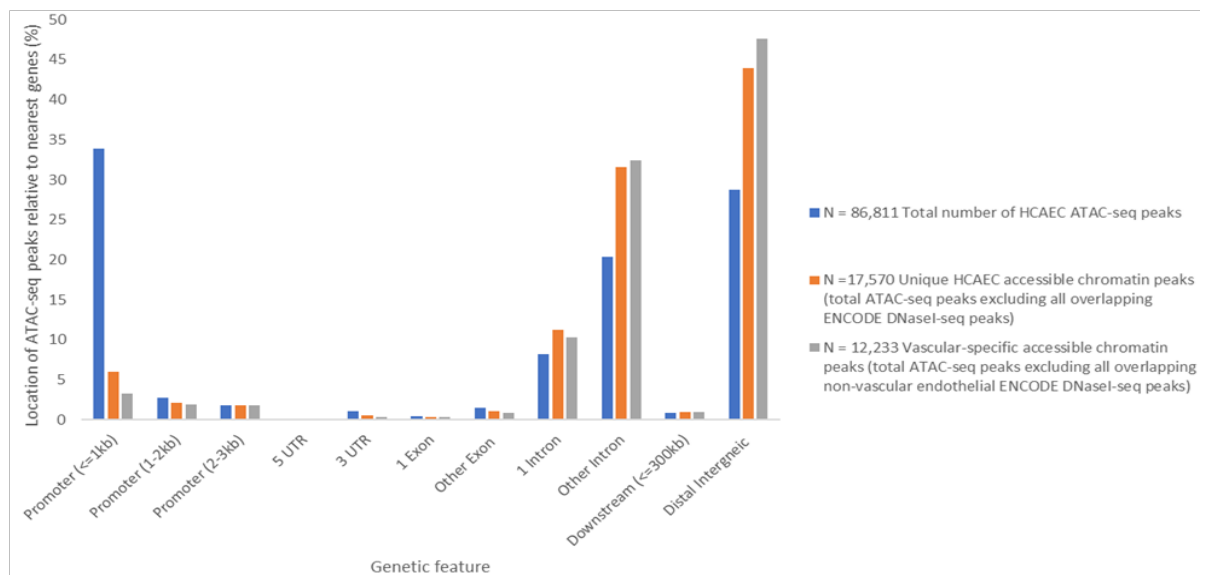
Figure 43. Distribution of accessible regions in the total, vascular-specific, and unique HCAEC identified using ATAC-seq. Downstream refers to centre of accessible regions located less than 300 kb from the TSS. Distal intergenic regions refer to accessible region located more than 3 kb upstream from the TSS and more than 1 kb downstream from TTS.

ATAC-seq peaks in vascular-specific accessible chromatin peaks and unique HCASMC accessible chromatin peaks were similarly correlated to their position relative to the nearest genomic feature(s). In the vascular-specific accessible chromatin peaks a greater proportion were again found outside of promoters: distal intergenic (51.1%), other intron (30.8%), and intron 1 (9.2%). In the unique accessible chromatin peaks this was even more evident: intergenic (51.5%), other intron (30.7%), and intron 1 (9.0%) (**Figure 44**). Different pattern of distribution was observed in the total accessible chromatin peaks, where the largest proportion was found in promoters as detailed in Chapter 5, section 5.3.6.
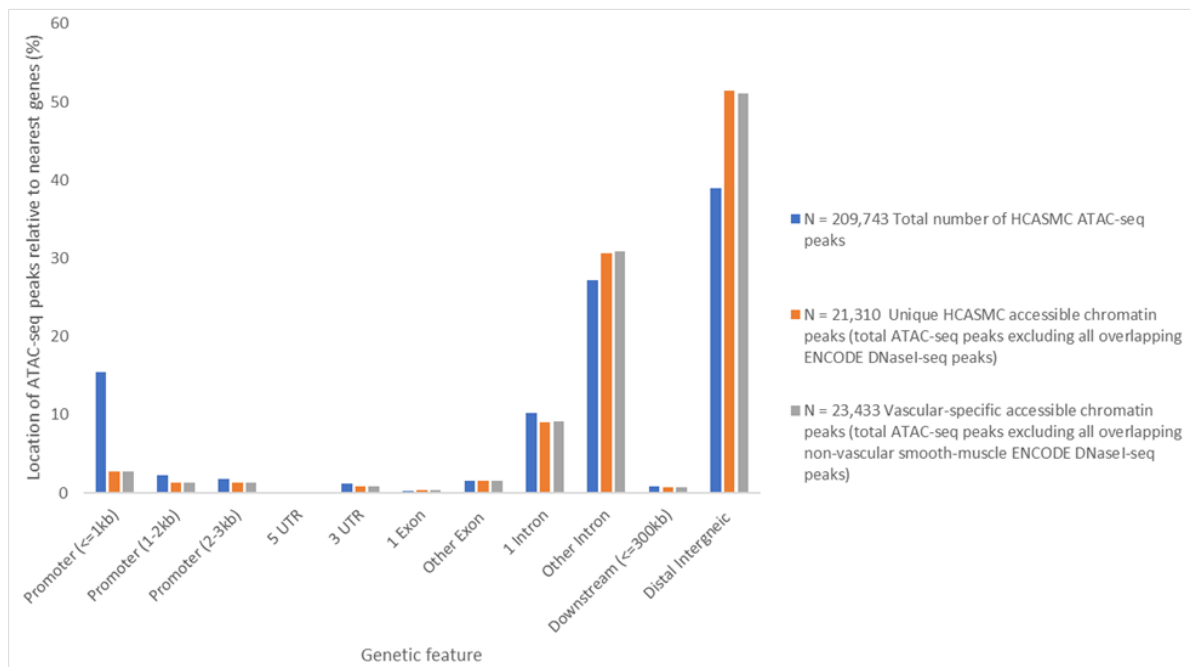
Figure 44. Distribution of accessible regions in the total, vascular-specific, and unique HCASMC identified using ATAC-seq. Downstream refers to centre of accessible regions located less than 300 kb from the TSS. Distal intergenic regions refer to accessible region located more than 3 kb upstream from the TSS and more than 1 kb downstream from TTS.

Overall, the identified accessible chromatin regions were enriched outside the TSS in the unique accessible chromatin peaks in HCAEC and HCASMC, while the identified accessible chromatin regions were enriched around the TSS in the total accessible chromatin peaks in HCAEC and HCASMC as detailed in Chapter 5, section 5.3.6 (**Figure 45 A and B**).

A
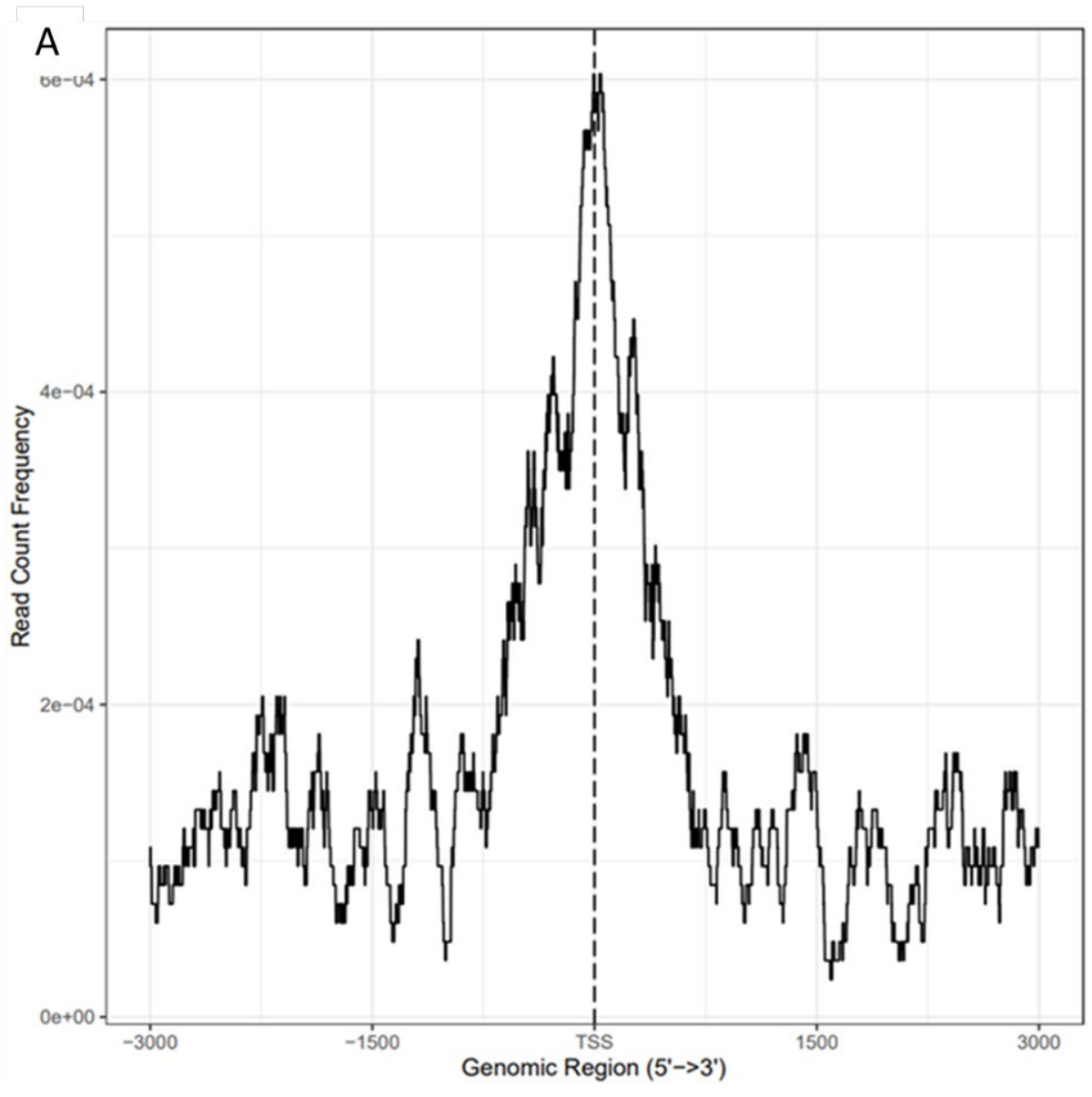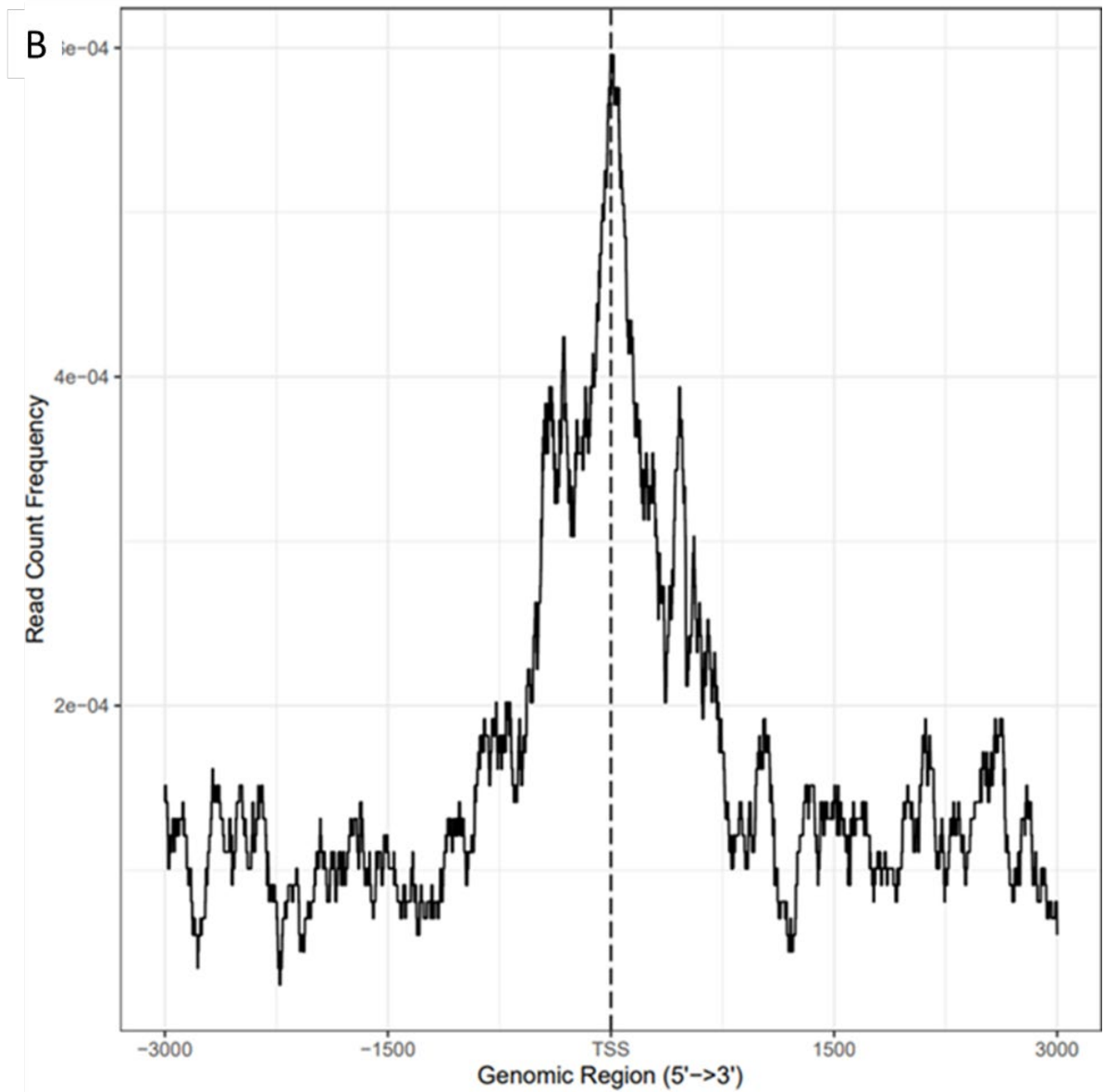
Figure 45. Distribution of unique accessible chromatin peaks in HCAEC and HCASMC using ATAC-seq. A-B) Distribution of ATAC-seq peaks outside the TSS for the unique accessible chromatin peaks in HCAEC and HCASMC, respectively. The centre of ATAC-seq peak was used to produce the distribution plots.

The nearest genes to accessible chromatin peaks were annotated using Chipseeker, which identified 2927 genes surrounding HCAEC ATAC-seq peaks in vascular-specific accessible chromatin peaks. Several KEGG pathways showing enrichment in these gene sets were shared between vascular-specific and total HCAEC, including Rap1, Hippo, and cAMP, Phospholipadase D, Oxytocin signalling pathways. In contrast, TGF-beta and Wnt signalling pathways were specific to the vascular-specific HCAEC (**Figure 46**).
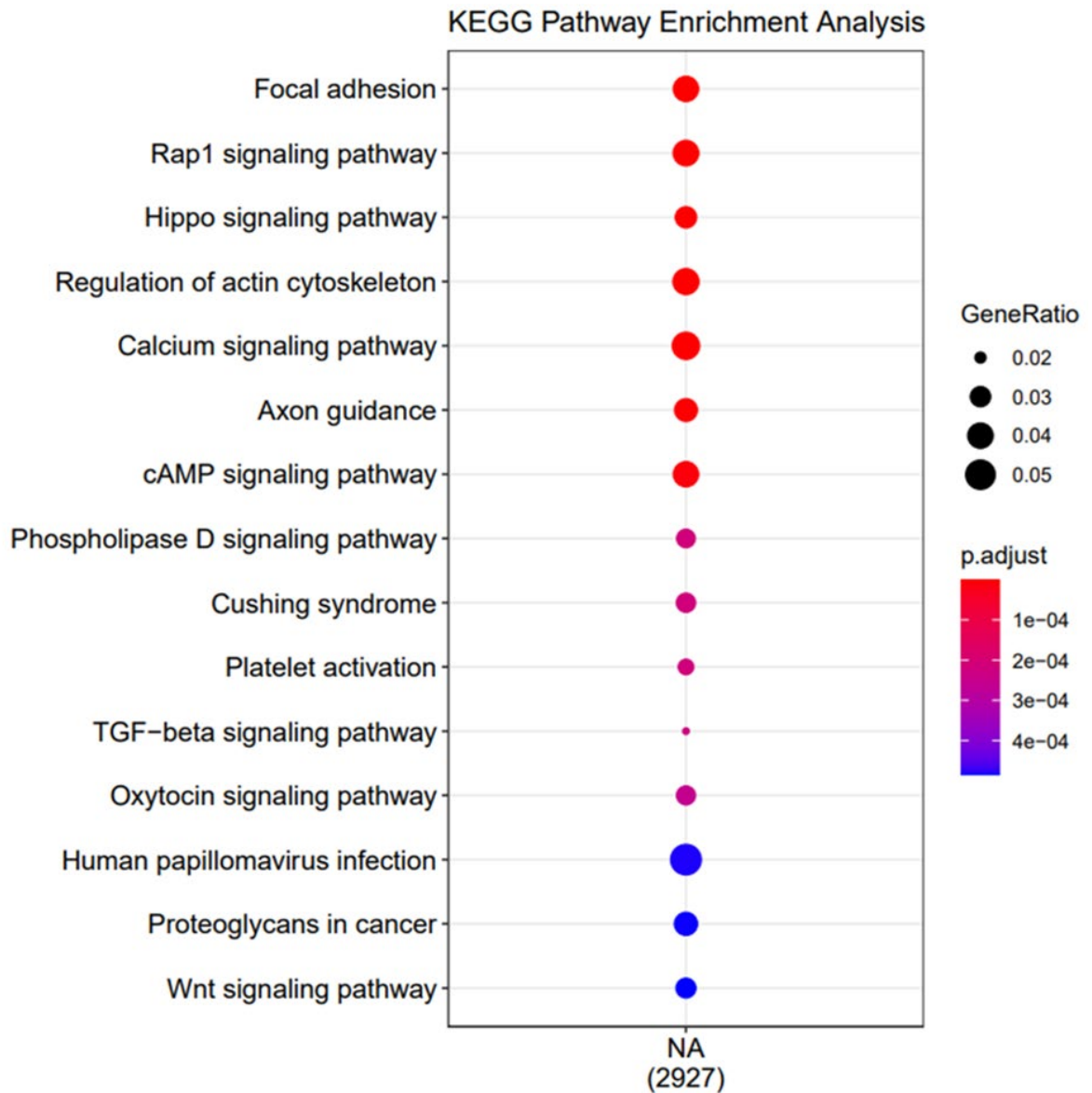
Figure 46. KEGG pathway enrichment analysis of genes surrounding vascular-specific accessible chromatin in HCAEC. The colour of the dot represents the adjusted *p*-value and the size of the dot represents the gene ratio. The gene ratio represents the total number of genes associated with this pathway.

5.3.9 Enrichment of transcription factor binding motifs within the vascular-specific and unique ATAC-seq peaks in HCAEC

To complement the KEGG pathway analysis from genes near accessible chromatin peaks described above, I performed enrichment analysis using the unique and vascular-specific accessible chromatin peaks in HCAEC using HOMER to identify regulatory motifs enriched in vascular accessible chromatin. There were a total of 181 motifs identified in vascular-specific and 52 motifs identified in unique accessible chromatin peaks in HCAEC at p-value $\ll 1\times10^{-50}$ compared to the human reference genome (hg19). The most significantly enriched motifs in vascular-specific accessible chromatin peaks in HCAEC, include those of BATF, Fra1, JunB, AP-1, and Fra2 (**Figure 48**; **Table 16**). Motifs shared between the vascular-specific and total accessible chromatin peaks in HCAEC, included Fra1 and JunB. In contrast, BATF, AP-1 and Fra2 were specific to the vascular-specific, while CTCF, BORIS and ATF were specific to the total HCAEC. The most significantly enriched motifs in unique accessible chromatin peaks in HCAEC, include those of CTCF, BORIS, Fra2, BATF, and Fra1 (**Figure 47**; **Table 15**). Motifs shared between the unique and total accessible chromatin peaks in HCAEC, included CTCF, BORIS, and Fra1. In contrast, Fra2 and BATF were specific to the unique, while ATF3 and JUNB were specific to the total accessible chromatin peaks in HCAEC.

Motif	Name
	CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer
	BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer
	Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer
	BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer
	Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer

Figure 47. Histogram distribution for top five enriched motifs identified using known motif enrichment analysis in unique HCAEC accessible chromatin peaks.

Table 15. Top 5 enriched known motifs found in unique HCAEC ATAC-seq peaks.

Motif	*P*-value	Accessible chromatin with TF motif (%)	Inaccessible chromatin with TF motif (%)
CTCF	1x10-39	2.7	0.2
BORIS	1x10-33	3.0	0.4
Fra2	1x10-14	3.3	1.1
BATF	1x10-14	3.8	1.4
Fra1	1x10-14	3.4	1.2

Motif	Name
	BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer
	Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer
	JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer
	AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer
	Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer

Figure 48. Histogram distribution for top five enriched motifs identified using known motif enrichment analysis in vascular-specific HCAEC accessible chromatin peaks.

Table 16. Top 5 enriched known motifs found in vascular-specific HCAEC ATAC-seq peaks.

Motif	*P*-value	Accessible chromatin with TF motif (%)	Inaccessible chromatin with TF motif (%)
BATF	1x10-355	11.1	1.5
Fra1	1x10-349	10.0	1.2
JunB	1x10-345	10.2	1.3
AP-1	1x10-337	11.6	1.8
Fra2	1x10-331	9.1	1.0

5.4   Discussion

I identified 86,811 and 209,743 ATAC-seq peaks in HCAEC and HCASMC, respectively, reflecting the level of chromatin accessibility in these two vascular cell types. The difference in the number of peaks found between HCAEC and HCASMC could be due to several reasons. ATAC-seq is often robust with minor variations of cell number. Over-digestion of chromatin

and larger fraction of reads that map to inaccessible regions of the genome could be due to using too few cells, while under-digestion and high molecular weight fragments are due to using too many cells. Thus, the same number of cells were used for all samples [135].

However, these results were consistent with previous studies using ATAC-seq to examine the genomic landscape of vascular endothelial and smooth muscle cells. A study by Lalonde *et al.*, [354], examined human aortic endothelial cells using ATAC-seq treated and untreated with TNFα and identified 95,491 ATAC-seq peaks [354]. A study by Miller *et al.*, [355], examined human coronary artery smooth muscle cells using ATAC-seq and an average of ~150,000 open chromatin peaks per sample.

Comparison of open chromatin profiles between HCAECs and HUVECs cells showed that although there are strongly correlated, they display regulatory differences that are likely to reflect separate cellular processes. Comparison of ATAC-seq (HCAEC) to DNaseI-seq and FAIRE-seq (HUVEC, ENCODE) data suggests that these tools may be complementarity in mapping open chromatin regions to fully explore the regulatory landscape.

I studied the presence of HUVEC epigenetic alterations in HCAEC ATAC-seq peaks compared to HUVEC DNaseI-seq/FAIRE-seq to establish whether differences in the tools to measure accessible chromatin may be explained by these. Differences in FAIRE-seq and DNase-seq may be due to the specific complexes bound at each open chromatin site, which could affect the ability of formaldehyde in FAIRE-seq to crosslink, or DNaseI to cut. FAIRE-seq sites tend to occur in distal regions, while DNase sites tend to occur at transcription start sites [356]. FAIRE-seq cannot detect some nucleosome-depleted regions that tightly bound by nonhistone proteins. However, FAIRE-seq can more readily detect chromatin structures away from promoters compared to DNaseI-seq.

The majority of shared accessible chromatin ATAC-seq peaks in HCAEC and HCASMC were in promoter regions, while vascular-specific accessible chromatin peaks were predominantly in non-promoter and enhancer regions. This is likely due to enhancer elements having a greater role in tissue-specific gene expression compared to promoters. The identification of novel enhancers and their associated gene expression will be useful to study the role of tissue-specific enhancers in human biology and disease [357].

ATAC-seq peaks mark regions with regulatory potential, typically close to the genes they regulate, I performed KEGG pathways gene enrichment analysis to explore vascular pathways

enriched in this data. This analysis revealed a number of similar pathways from genes near HCAEC and HCASMC ATAC-seq peaks, including Rap1, MAPK, Hippo, endocytosis, oxytocin, sphingolipid signalling pathways.

Rap1 promotes angiogenic signalling by several factors, including FGF2, VEGF and S1P. This results in activation of Erk, Akt, Rac1 and Afadin, regulating endothelial cell proliferation, migration, and tubule formation [358].

MAPK plays an important role in several processes, including differentiation, proliferation, migration, senescence, and apoptosis of a large array of cell types. Activation of the MAPK signalling pathway has been shown to upregulate the expression of VEGF, which acts to increase tube formation in functional vessels, thus inducing angiogenesis [359].

Hippo-YAP/TAZ plays an important role in retinal blood vasculature and differentiation of endothelial progenitor cells to matured endothelial cells. YAP is inactivated by phosphorylation and redistributed in a cell contact-dependent matter in HUVECs. Knockout of YAP in HUVECs showed decrease in the total network and endothelial sprouts, respectively. Hippo-YAP/TAZ plays an important role in vascular sprouting, vascular barrier formation, and vascular remodeling [360].

Endocytosis plays an important role in several processes ranging from nutrient uptake to intracellular signal transduction. It also plays an important role in VEGF receptor signalling, angiogenesis and differential regulation in angiogenic compared to quiescent endothelial cells [361].

Oxytocin stimulates migration and invasion in HUVECs, thus indicating a possible role for the peptide in the regulation of angiogenesis. Oxytocin activates phosphatidylinositol-3-kinase (PI-3-K)/AKT/endothelial eNOS pathway. Oxytocin phosphorylates proline-rich tyrosine kinase-2 (Pyk-2) and Src kinase in a PLC- and calcium-dependent manner. Oxytocin induces HUVEC capillary outgrowth via Pyk-2 phosphorylation, which activates Src which in turn activates the PI-3-K/AKT pathway [362].

Sphingolipids regulate vascular permeability, augments endothelial integrity, and promote vascular leak. The mechanisms by which sphingolipids regulate endothelial barrier function occur via multiple different pathways and disruption of these pathways have been implicated in several diseases [363].

In addition, there are a number of enriched pathways from loci specific to HCAEC and HCASMC accessible chromatin. Ras and cellular senescence pathways are specific to HCAEC. Ras regulates angiogenesis and vascular permeability by activating of downstream effectors [364]. Cellular senescence can occur *in vivo* or *in vitro* in response to various stressors, leading to suppression of cell proliferation [361]. cAMP, vascular smooth muscle contraction, and prolactin pathways are specific to HCASMC. cAMP-adenosine pathway may contribute to the regulation of vascular biology. Adenosine has multiple actions, mediated primarily by A 1 and A 2 adenosine receptors, that may serve important roles in vascular biology, particularly with regard to reducing the risk and consequences of vaso-occlusive diseases [365]. Vascular smooth muscle contraction plays in important role in vascular dysfunction and exaggerated vasoconstriction in vascular disease [366]. Prolactin regulates the development of mammary gland, contributes to a broad range of pathologies, including angiogenesis via the release of pro-angiogenic factors by leukocytes and epithelial cells [367].

Moreover, there are a number of additional pathways from genes surrounding vascular-specific accessible chromatin, including TGF-beta and Wnt singaling pathways. TGF-β family members are multifunctional cytokines that express their effects on cells, including endothelial cells via type I and type II serine/threonine kinase receptors and intracellular SMAD transcription factors. Hence, TGF-β plays an essential role in vascular biology and disease [368]. Wnt signalling plays a role in vascular morphogenesis in the embryo and in organ-specific endothelial differentiation. Wnt signalling regulates fundamental aspects of development, including cell fate specification, proliferation, and survival, and may use different receptors and signalling pathways. Both loss- and gain-of-function experiments of members of the Wnt signalling pathway were found to cause marked alterations of vascular development and endothelial cell specification. Furthermore, altered Wnt signalling in the endothelium may contribute to pathological conditions such as retinopathies, pulmonary arterial hypertension, and stroke [369].

I examined TF motif enrichment in order to understand the regulation of gene expression in HCAEC and HCASMC ATAC-seq datasets. The most significantly enriched motifs in HCAEC ATAC-seq peaks, include CTCF, BORIS, ATF3, JUNB, and Fra1. CTCF is an important mediator and genomic insulator of long-range genomic interactions. The role of CTCF specifically in development of the vascular system has not been investigated [370], however CTCF has been shown to bind to the VEGFA promoter, preventing surrounding

enhancers from activating its expression [371, 372]. CTCF gene duplication gives rise to brother of the regulator of imprinting sites (BORIS) [373-378]. CTCF and BORIS share an almost identical DNA binding domain that recognise the same DNA sequences [373-378]. CTCF and BORIS possess distinct functions and act in mutually exclusive manner. CTCF is ubiquitously expressed, while BORIS is expressed in germ cells [373].

AP-1 regulates a wide range of cellular processes, including proliferation, differentiation, and apoptosis [379, 380]. AP-1 is composed of members from the JUN, FOS, ATF, and MAF protein families [379, 380] and members share common, basic leuine-zipper domains [381]. Studies on gene regulation in endothelial cells demonstrated that the AP-1 transcription factors are implicated in angiogenesis and vascular development [352]. In particular, it has been revealed that JunB is transiently induced in endothelial cells at the angiogenic frontier where is controls tip cell specification during vascular development [352]. Moreover, JunB plays a role in tissue-specific vascular maturation processes during neurovascular interaction in mouse embryonic skin and retina vasculatures. This data corroborates the idea that AP-1 may act as an angiogenic transcription factor that induces endothelial cell migration in HCAECs [352].

In comparison to HCAECs, the most significantly enriched motifs in HCASMC ATAC-seq peaks included BATF, ATF3, FRA1, AP-1 and JunB. Since this work was performed, another study looking at HCASMC ATAC-seq profiles was performed by Miller *et al.,* [355]. HCASMC were treated under serum-free control, TGF-β, PDGF-BB or PDGF-DD conditions and profiled using deep ATAC-seq. They performed *de novo* motif enrichment analysis of each stimulated condition using serum-free control as a background to identify relevant TFBSs. The most significantly enriched motifs in TGF-β-treated HCASMCs were the AP-1 family members followed by TGF-β-related TFs [355].

Interestingly, Atf3, JunB, and Fra1 are the three common motifs in HCAEC and HCASMC ATAC-seq peaks. A study by Bysani *et al.,* [142] examined ATAC-seq open chromatin regions with genomic binding of islets-specific TFs using ChIP-seq data. They used HOMER to identify enriched putative TF motifs in the open chromatin regions of islets. The most significantly enriched motifs in the open chromatin regions of non-diabetic islet donors include those of FRA1, ATF3, and AP-1. The FRA1 motif was the most significantly enriched among islet ATAC-seq peaks. They found that the motif for FRA1 was enriched in ATAC-seq peaks in alpha- and beta-cells [142]. FRA1 is involved in MAPK signalling and oxidative stress

pathways. However, there is limited knowledge of its role in islets and future studies evaluating its binding and function would be interesting [284]. However, HOMER is a differential motif discovery algorithm, common repeats are usually in both the target and background sequences. Transcription factor binds to a certain class of repeats, which may cause several large stretches of similar sequence to be processed, which may result in bias [292].

This map of open chromatin regions from ATAC-seq in HCAEC and HCASMC adds to existing vascular accessible chromatin data from HUVEC cells derived from FAIRE-seq and DNaseI-seq. It will provide a resource to interrogate specific genomic regions for regulatory potential and variants for potential functionality.

6. Impact of VEGFA-stimulation on the open chromatin landscape of vascular endothelial cells

6.1 Introduction

VEGFA VEGFB, VEGFC, VEGFD, and VEGFE, as well as PlGF -1 and -2 belong to a group of structurally related dimeric proteins [382]. VEGFA ligands induce angiogenic signals through several receptors. VEGFR-1 (FLT1) binds VEGFA, VEGFB, and PlGF -1/2 and promotes monocyte migration. VEGFR-2 (KDR) binds VEGFA, VEGFC, VEGFD, and VEGFE and promotes endothelial cell proliferation, migration and survival [382]. VEGFR-3 binds to VEGFC and VEGFD and promotes growth of lymphatic vessels. VEGFA has two coreceptors, including NRP-1/2. NRP-1/2 bind VEGFA ligands and enhance their affinity to other VEGFRs [383].

VEGFA binds VEGFR-2 leading to activation of tyrosine kinases, including Tyr951, Tyr996, Tyr1054, Tyr1175, and Tyr1214 [384]. In turn, this leads to activation of signal-transduction molecules MAPK, Src, Ras, Akt, PI3K, and PLC-γ. Tyr1175 and PLC-γ phosphorylation leads to Ca2+ release, which results in PKC activation. PKC activation stimulates the Raf/MEK/ERK pathway, which leads to cell proliferation. PKC activation leads to activation nitric oxide synthase activity, which subsequently results in vascular permeability (**Figure 49**) [385].
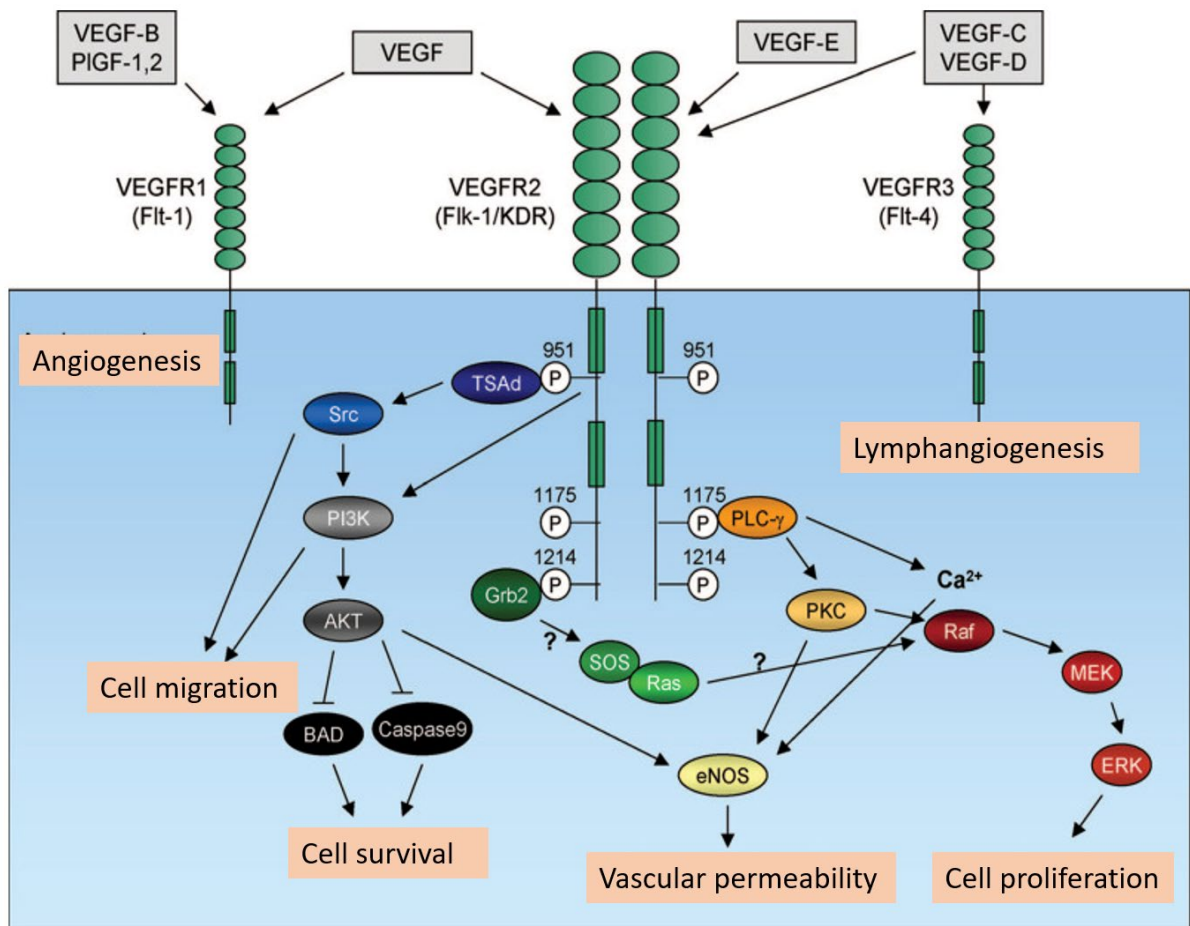
Figure 49. VEGF signal transduction. VEGFA, VEGFB, VEGFC, VEGFD, VEGFE and PlGF bind three VEGFR tyrosine kinases, which leads to dimerisation, autophosphorylation of receptors, and activation of downstream signal-transduction pathways. VEGFR2 activates signal-transduction molecules MAPK, Src, Ras, Akt, PI3K, and PLC-γ, and regulates cell proliferation, migration, survival, and vascular permeability (adapted from Nilsson *et al.,* 2006).

PI3K plays an essential role in cell survival, migration, and proliferation. Some studies implicate the role of Src kinases, β-catenin, and VE-cadherin in activation of PI3K [386, 387]. PI3K is activated through VEGFR-2 leading to accumulation of phosphatidylinositol-3,4,5-triphosphate. This results in Akt/PKB phosphorylation, which inhibits proapoptotic proteins, BAD, and caspase-9.

Src kinases, including Src, Fyn, and Yes, mediate VEGFA induced PI3K activation and regulate fiber organisation. Grb2 kinds pTyr1214 on VEGFR-2 resulting in the guanine-nucleotide–exchange Sos, which activates Ras pathway [384]. Ras pathway plays an important role in cell proliferation [265].

VEGFA is the most well-studied ligand [388]. Several isoforms are produced through alternative splicing, including $VEGFA_{121}$, $VEGFA_{121}b$, $VEGFA_{145}$, $VEGFA_{145}b$, $VEGFA_{165}$, $VEGFA_{165}b$, $VEGFA_{183}$, $VEGFA_{189}$, and $VEGFA_{206}$ [245, 389]. $VEGFA_{xxx}$ family members have pro-angiogenic, while $VEGFA_{xxx}b$ have anti-angiogenic effects [390, 391]. Several of these $VEGFA_{xxx}b$ isoforms act as targets to treat angiogenic diseases [392-394].

The aim of this chapter was to characterise the genomic regulatory landscape of vascular endothelial cells under angiogenic conditions. The effect of VEGFA was examined on chromatin structure using HCAECs to determine whether novel regulatory elements were uncovered compared to their basal state.

6.2 Methods

6.2.1 RNA-seq data analysis

RNA-seq was performed VEGFA-stimulated and unstimulated and HCAEC. I used single samples without replicates. This was an exploratory analysis to examine the effects of VEGFA stimulation on HCAEC. 10 µg recombinant VEGFA for 1 hour was used to stimulate the HCAEC. RNA-seq raw data was processed using standard procedures detailed in Chapter 3. DNA from VEGFA-stimulated HCAEC was quantified using the Qubit and TapeStation as detailed in Chapter 3, section 3.8. multiBamSummary [293] was used to compute the read coverages of VEGFA-stimulated HCAEC samples in Chapter 3, section 3.20. The BEDTools [341] intersect function was used to identify all reads in peaks (VEGFA-stimulated HCAEC ATAC-seq reads in peaks) as detailed in Chapter 3, section 3.21. Annotation of the ATAC-seq peaks in VEGFA-stimulated and unique VEGFA-stimulated HCAEC was performed using ChIPseeker as detailed in Chapter 3, section 3.18 [291]. ATAC-seq peaks in VEGFA-stimulated and unique VEGFA-stimulated HCAEC were evaluated for biological themes among gene clusters, using compareCluster from ChIPseeker as detailed in Chapter 3, section 3.18 [291]. TF binding motif analysis of ATAC-seq peaks in VEGFA-stimulated and unique VEGFA-stimulated HCAEC datasets was performed using HOMER as detailed in Chapter 3, section 3.19 [292].

6.3 Results

6.3.1 Effects of VEGFA-stimulation in HCAEC on genome-wide gene expression

FastQC analysis did not identify any sub-standard sample (Appendix, **Figure 79** and **Figure 80**) show a representative examples of the FastQC reports generated. RNA-seq was performed

to examine the effect of VEGFA-stimulation on HCAEC gene expression. The uniquely mapped reads from VEGFA-stimulated and unstimulated and HCAEC samples had high mapping quality (>70% uniquely mapped reads), which were processed for downstream analysis (**Table 17**).

Table 17. Summary mapping statistics of VEGFA-stimulated and unstimulated and HCAEC.

HCAEC	Uniquely mapped reads %	76.2
	% of reads mapped to multiple loci	22.6
	% of reads unmapped: too short	1.0
VEGFA-stimulated HCAEC	Uniquely mapped reads %	76.4
	% of reads mapped to multiple loci	23.0
	% of reads unmapped: too short	0.3

Htseq-count was used to count how many reads mapped to each gene using Ensembl gene location information. There was a total of 1,978,829 reads with no features, which could not be assigned to any gene, 840,128 ambiguous reads, which were mapped to a location containing several genes and were hence excluded from further analysis, and 6,365,273 reads with more than one alignment for the VEGFA-stimulated HCAEC. In comparison, for the unstimulated HCAEC there was a total of 1,777,657 reads with no features, 780,497 ambiguous reads, which were mapped to a location containing several genes, and 6,043,979 reads with more than one alignment (**Table 18**). In total, 148 genes were up-regulated (1.5 fold) and 565 down-regulated genes (0.75 fold) (data not shown) upon VEGFA-stimulation.

Table 18. Count aligned reads per gene with Htseq-count.

	HCAEC	VEGFA-stimulated HCAEC
No feature	1978829	1777657
Ambiguous	840128	780497
Too low aQual	0	0
Not aligned	0	0
Alignment not unique	6365273	6043979

To confirm whether VEGFA-stimulation resulted in expected changes of HCAEC gene expression several biologically relevant genes were examined, including genes in the early growth response (*EGR*) family and nuclear receptor (*NR4A*) subfamily. **Figure 50** show there was an increase in expression of all three EGR-family genes, including *EGR4*, *EGR2* and *EGR3* and all three NR4A-family genes, including *NR4A1*, *NR4A2* and *NR4A3*.
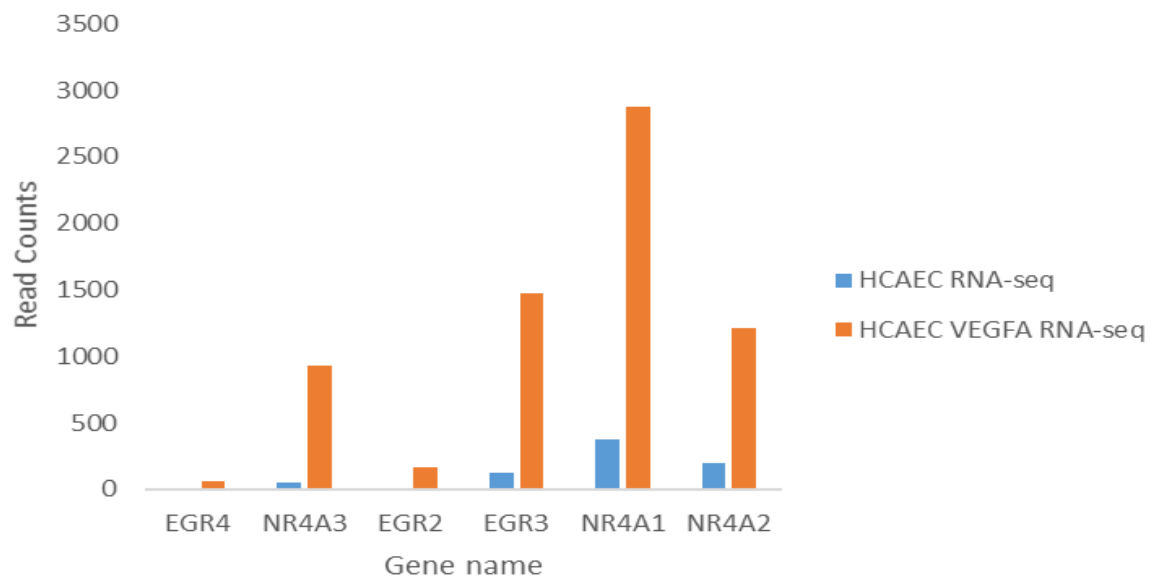
Figure 50. Comparison of *EGR* and *NR4A* gene expression from in VEGFA-stimulated and unstimulated HCAEC. *EGR3* and *NR4A1* are strongly upregulated upon VEGFA-stimulation.

To examine upregulation of VEGF-related pathways, I performed pathway analysis using Reactome database using the 148 (1.5-fold) up-regulated genes. This demonstrated enrichment of processes involved in VEGF-VEGFR interactions (p = 5.8 x 10^{-3}; FDR = 0.13) (**Figure 51**). Appendix, **Table 33** shows detailed Reactome pathway analysis of the 148 up-regulated genes upon VEGFA-stimulation.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found
FOXO-mediated transcription of cell death genes	3	23	0.002	4.29E-3	1.12E-1	4
Transcriptional regulation of white adipocyte differentiation	6	109	0.007	4.58E-3	1.14E-1	7
VEGF ligand-receptor interactions	2	8	0.001	5.77E-3	1.3E-1	3
VEGF binds to VEGFR leading to receptor dimerization	2	8	0.001	5.77E-3	1.3E-1	2
Response of EIF2AK4 (GCN2) to amino acid deficiency	6	115	0.008	5.9E-3	1.3E-1	11

Figure 51. Reactome pathway analysis of the 148 up-regulated genes (1.5 fold) upon VEGFA-stimulation. This figure shows closely related vascular pathways, including VEGFA ligand-receptor interactions and VEGFA binds VEGFR leading to receptor dimerisation.

6.3.2   ATAC-seq

ATAC-seq in the three VEGFA-stimulated HCAEC samples was performed as detailed in Chapter 3, section 3.5. **Table 19** shows the amount of purified ATAC-seq DNA per sample. As before, VEGFA-stimulated HCAEC samples resulted in fragments with no size periodicity (**Figure 52**).

Table 19. DNA concentration of ATAC-seq libraries determined by Qubit.

Sample ID	Concentration [pg/ul]
HCAEC-1-VEGFA	3.4
HCAEC-2-VEGFA	104
HCAEC-3-VEGFA	104
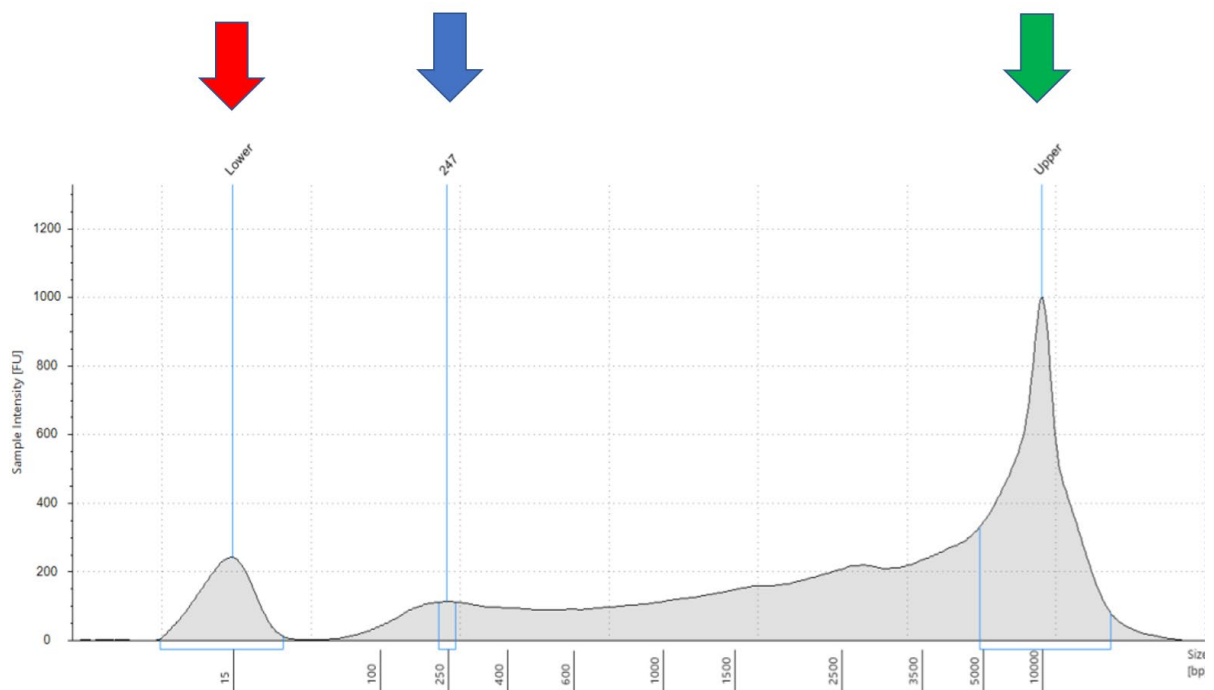
Figure 52. Electropherogram of fragment size distribution of ATAC-seq libraries determined by TapeStation. The X-axis represents the product size (bp) and the Y-axis represents the arbitrary fluorescence intensity (FU). Red and green arrows represent lower and upper size markers, respectively. The figure shows poor fragment size distribution of VEGFA-stimulated HCAEC sample. Blue arrow at 247 bp indicate fragment size of peak, which may represent mononucleosome fragments.

FastQC analysis did not identify any sub-standard sample (Appendix, **Figure 77**) shows a representative example of one of the FastQC report generated.

6.3.3 Correlation analysis of VEGFA-stimulated HCAEC ATAC-seq samples

Pearson correlation coefficient from the three VEGFA-stimulated HCAEC samples ranged from 0.89 to 0.96, demonstrating the desired strong correlation between samples (**Figure 53**).
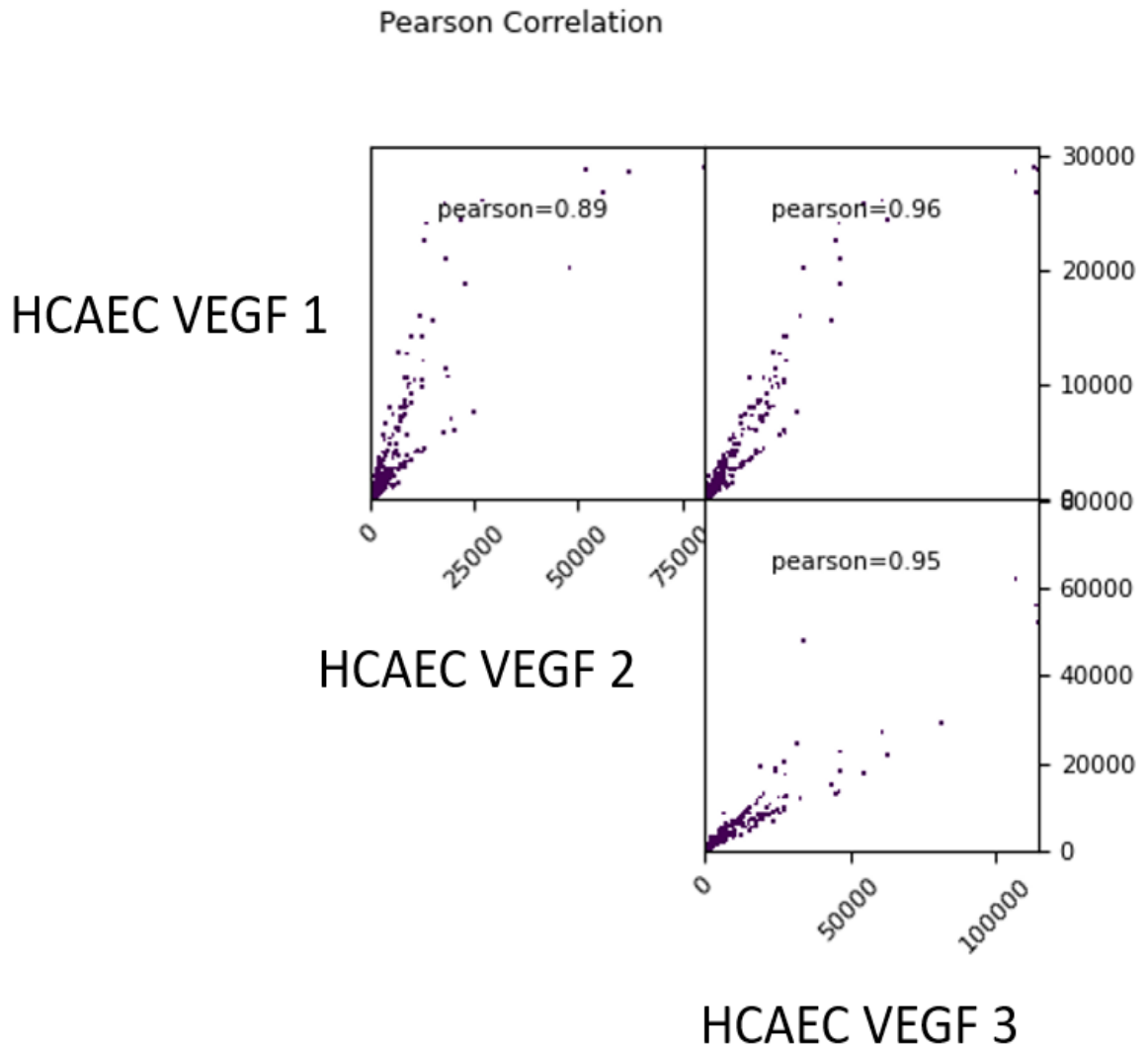
Figure 53. Pearson correlation coefficient of read coverages in VEGFA-stimulated HCAEC ATAC-seq samples. The y-axis and x-axis represent the number of fragments in VEGFA-stimulated HCAEC ATAC-seq samples. Each dot represents one genomic region. The gene ratio represents the total number of genes associated with this pathway.

6.3.4 ATAC-seq peaks in VEGFA-stimulated HCAECs

The FRiP score for VEGFA-stimulated HCAEC ATAC-seq libraries ranged from 0.12 to 0.16, meeting the recommended 0.2 standard (**Table 20**). Given the strong correlation between replicates, I combined the ATAC-seq data from each set of replicate samples to obtain a comprehensive set of peaks per cell type. ATAC-seq peaks were called by combining the data from each set of replicate samples resulting in 108,736 ATAC-seq peaks for VEGFA-stimulated HCAEC. In addition, there are 39,935 unique ATAC-seq peaks of VEGFA-stimulated HCAEC.

Table 20. Fraction of VEGFA-stimulated ATAC-seq reads in peaks.

Sample ID	Number of peaks	Reads in HCAEC VEGFA Peaks	Total reads used to call peaks	FRiP score
HCAEC-1-VEGFA	72,352	5944100	47927199	0.12
HCAEC-2-VEGFA	95,952	14015406	62640505	0.22
HCAEC-3-VEGFA	98,931	17727349	108979469	0.16
HCAEC-VEGFA-merged	108,736	42757536	219547173	0.19

6.3.5 Comparison of the genomic landscape of ATAC-seq peaks in the VEGFA-stimulated and unstimulated HCAEC

ATAC-seq profiling was performed for ATAC-seq peaks in the VEGFA-stimulated and unstimulated HCAEC to compare their distribution relative to genetic features based on the nearest gene in vascular accessible chromatin. Of the accessible regions in the VEGFA-stimulated HCAEC identified using ATAC-seq, the largest proportion of peaks mapped in distal intergenic (30.3%), promoter (29.4%) and other intronic regions (21.9%). There is a similar distribution in the unstimulated HCAEC as detailed in Chapter 5, section 5.3.6 (**Figure 54 A**). The accessible regions identified were again enriched around the TSS (**Figure 54 B**).

A
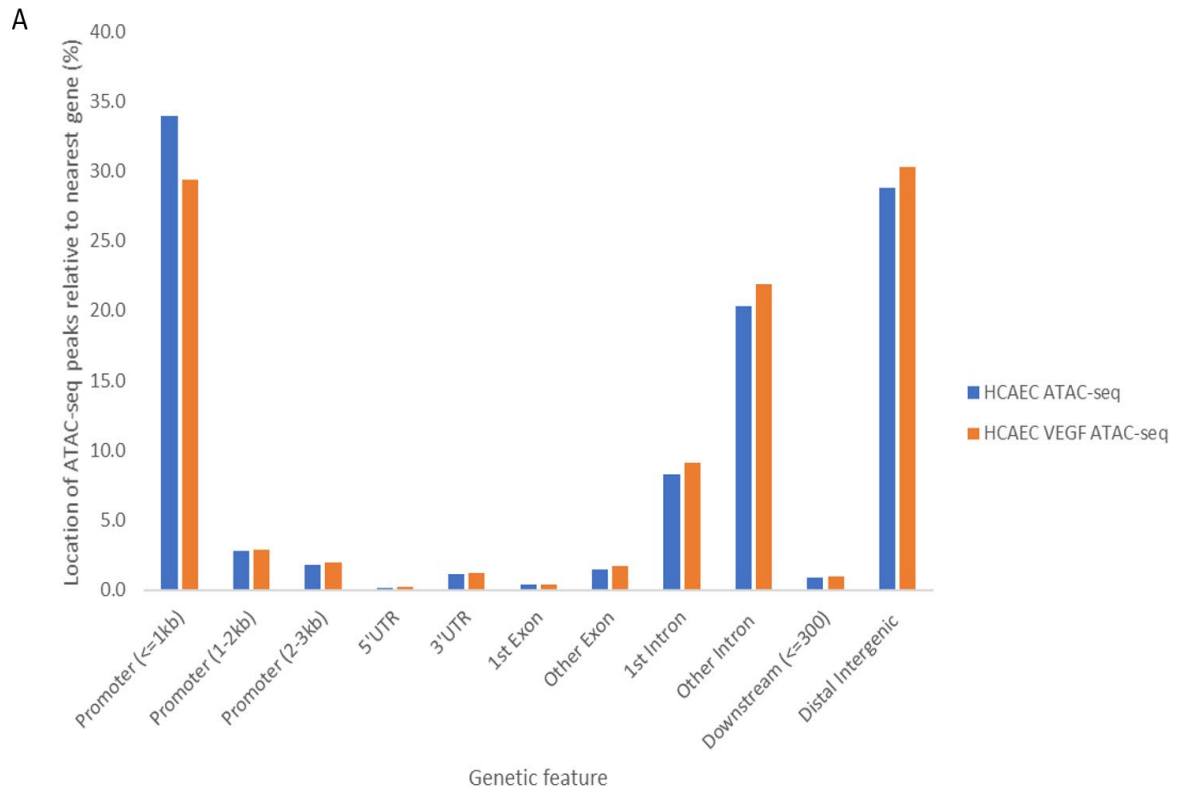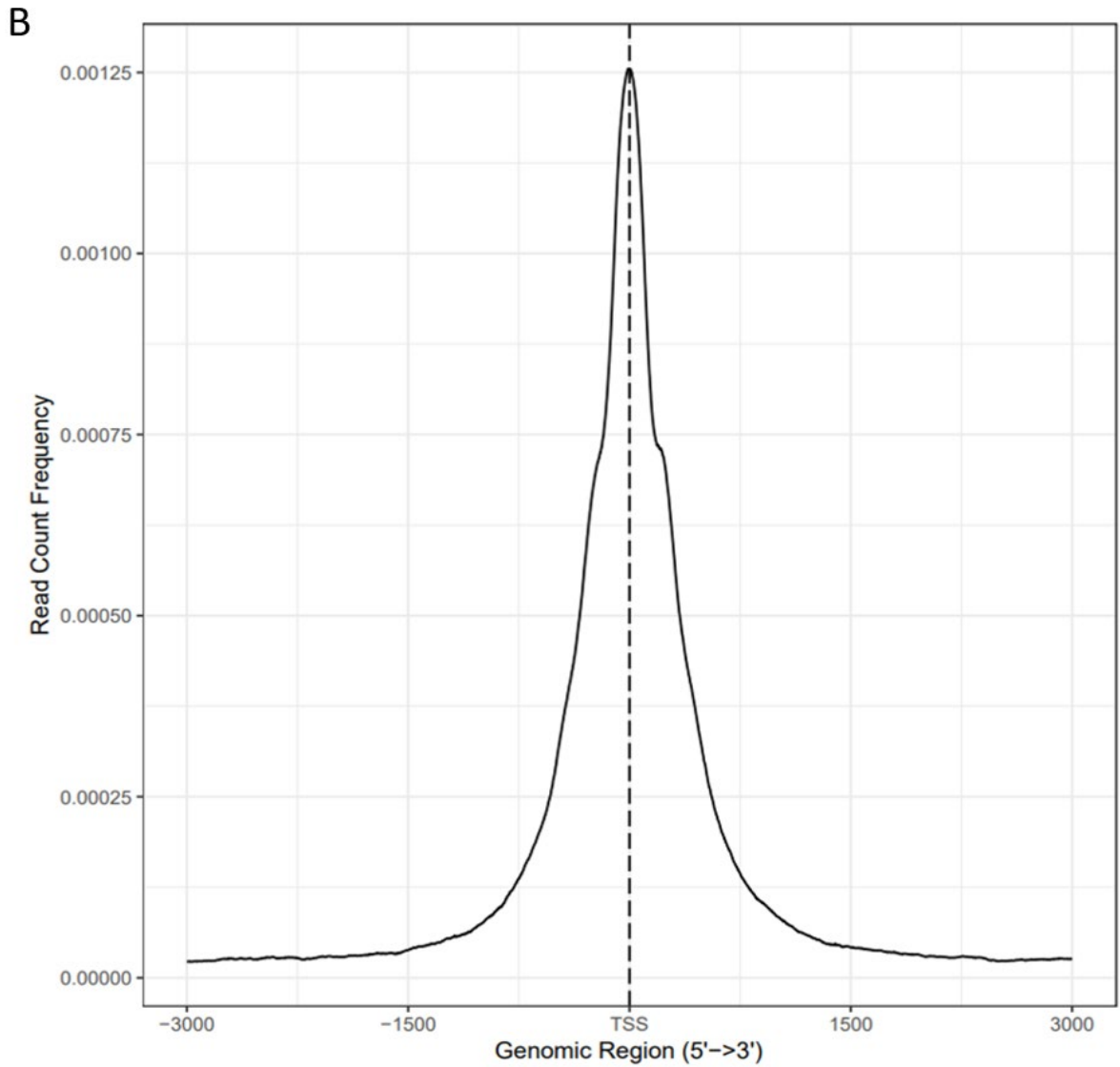
Figure 54. Distribution of accessible regions in the VEGFA-stimulated and unstimulated HCAEC identified using ATAC-seq. (A) Downstream refers to where the centre of accessible chromatin is located less than 300 kb from the TSS. Distal intergenic regions refer to accessible region located more than 3 kb away from the TSS and more than 1 kb from TTS (B) Distribution of ATAC-seq peaks around the TSS for the VEGFA-stimulated HCAEC. The centre of ATAC-seq peak was used to produce the distribution plots.

The nearest genes to accessible chromatin were annotated using Chipseeker, which identified 6,714 genes surrounding ATAC-seq peaks in the VEGFA-stimulated HCAEC. Several KEGG pathways showing enrichment in these gene sets were shared between VEGFA-stimulated and unstimulated HCAEC, including MAPK, Hippo, Rap1, Ras, Endocytosis, and Oxytocin signalling pathways. In contrast, Apelin signalling pathway was specific to the VEGFA stimulated HCAEC, while cellular senescence and sphingolipid signalling pathways were specific to the unstimulated HCAEC (**Figure 55**).
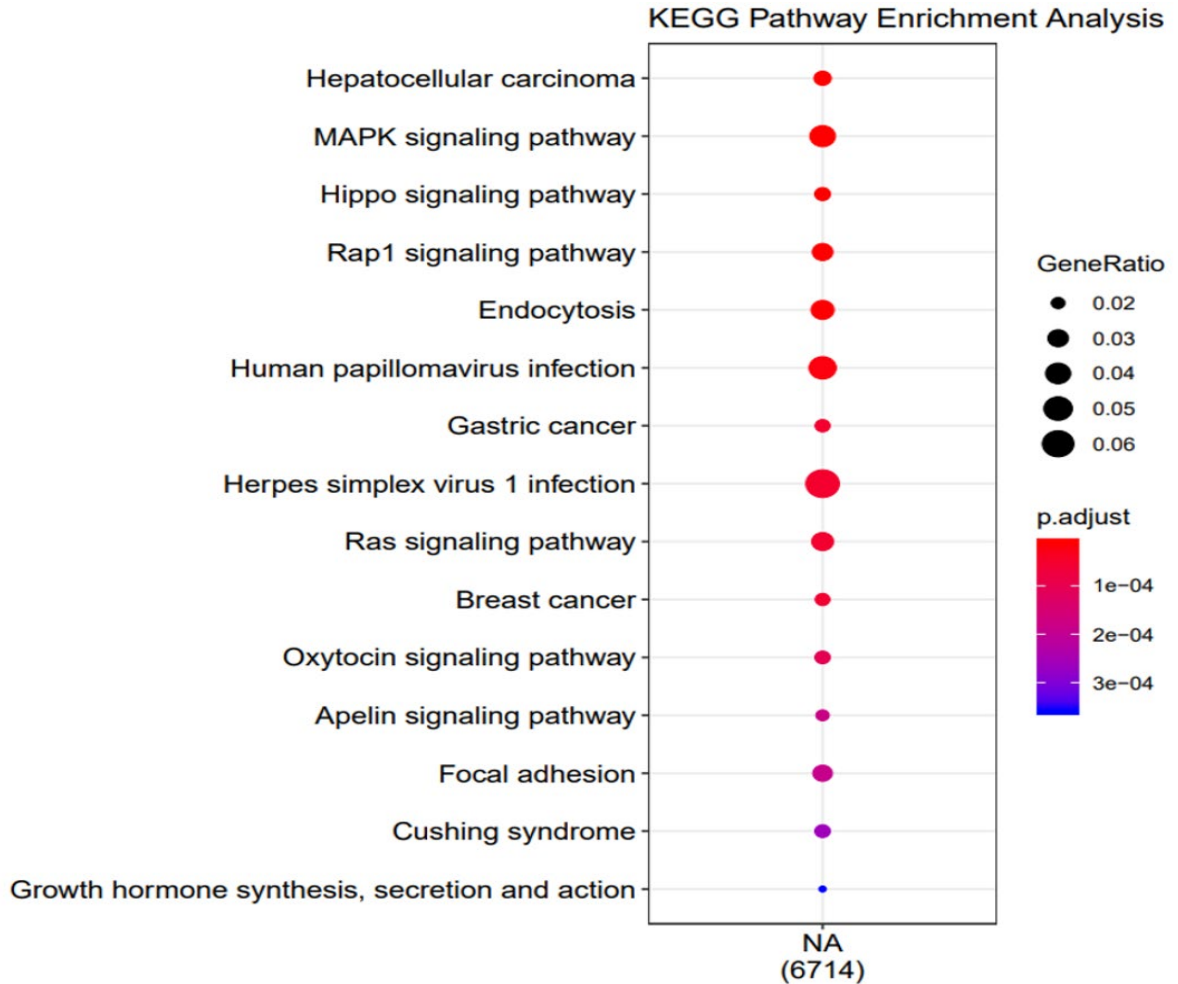
Figure 55. KEGG pathway enrichment analysis of genes surrounding vascular accessible chromatin in the VEGFA-stimulated HCAEC versus unstimulated HCAEC. The colour of the dot represents the adjusted *p*-value and the size of the dot represents the gene ratio. The gene ratio represents the total number of genes associated with this pathway.

6.3.6 Enrichment of transcription factor binding motifs in the VEGFA-stimulated HCAEC

To complement the KEGG pathway analysis from genes near accessible chromatin peaks described above, I performed enrichment analysis for transcription factor binding motifs across

146

ATAC-seq peaks in the VEGFA-stimulated HCAEC ATAC-seq peaks was performed using HOMER to identify regulatory motifs enriched in vascular accessible chromatin. There were a total of to 365 motifs at a p-value $<1\times10^{-50}$ compared to the human reference genome (hg19). The most significantly enriched motifs in the VEGFA-stimulated HCAEC, include those of CTCF, BORIS, ATF3, Fra1, and BATF (**Figure 56**; **Table 21**). Motifs shared between the VEGFA-stimulated and unstimualted HCAEC, included CTCF, BORIS, ATF3, and Fra1. In contrast, BATF was specific to the VEGFA-stimualted HCAEC, while JUNB was specific to the unstimulated HCAEC.
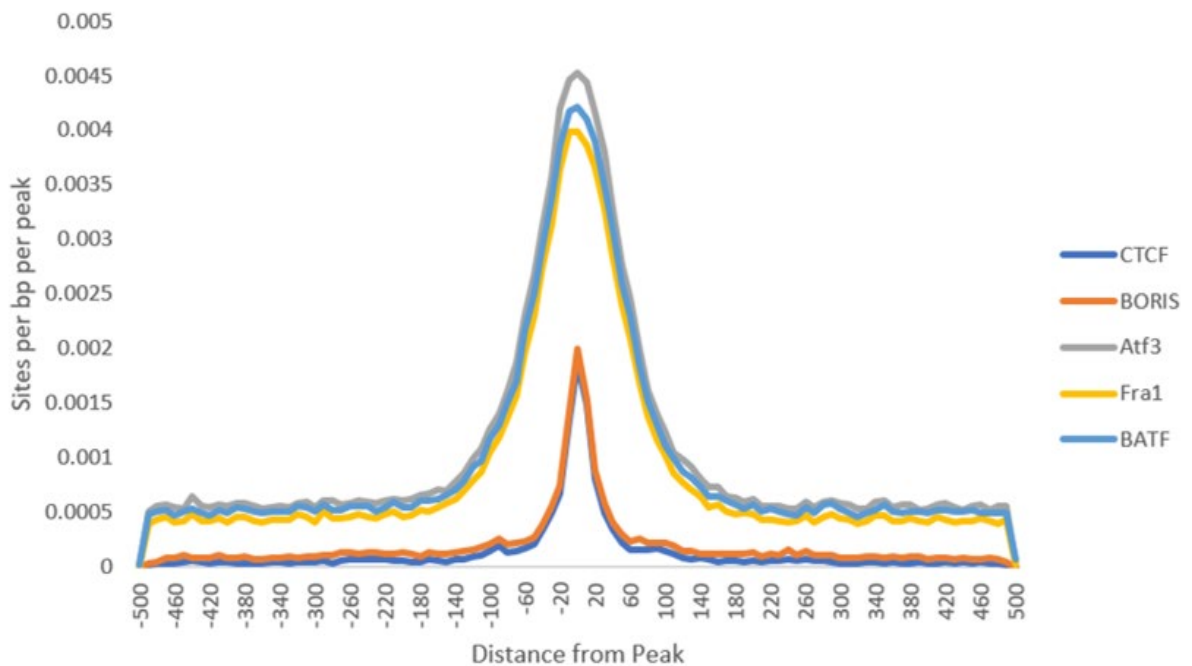
Figure 56. Histogram distribution for top five enriched motifs identified using known motif enrichment analysis in the VEGFA-stimulated HCAEC.

Table 21. Top 5 enriched known motifs found in VEGFA-stimulated HCAEC ATAC-seq peaks.

Motif	*P*-value	Accessible chromatin with motif (%)	Inaccessible chromatin with motif (%)
CTCF	1x10-4548	6.0	0.2
BORIS	1x10-3413	6.2	0.5
Atf3	1x10-3283	8.9	1.3
Fra1	1x10-3274	8.1	1.1
BATF	1x10-3200	8.8	1.3

6.3.7 Comparison of the genomic landscape of ATAC-seq peaks in the unique VEGFA-stimulated and unstimulated HCAEC

ATAC-seq profiling was performed in the unique VEGFA-stimulated compared to the unstimulated HCAEC, to examine their distribution relative to genic features based on the nearest gene. Of the accessible regions in the unique VEGFA-stimulated HCAEC, the largest proportion of peaks mapped in distal intergenic (37.6%), other intronic (28.8%) and promoter regions (11.6%), There is a different distribution in the unstimulated HCAEC, where the largest proportion of peaks were found promoters as detailed in Chapter 5, section 5.3.6 (**Figure 57 A and B**).
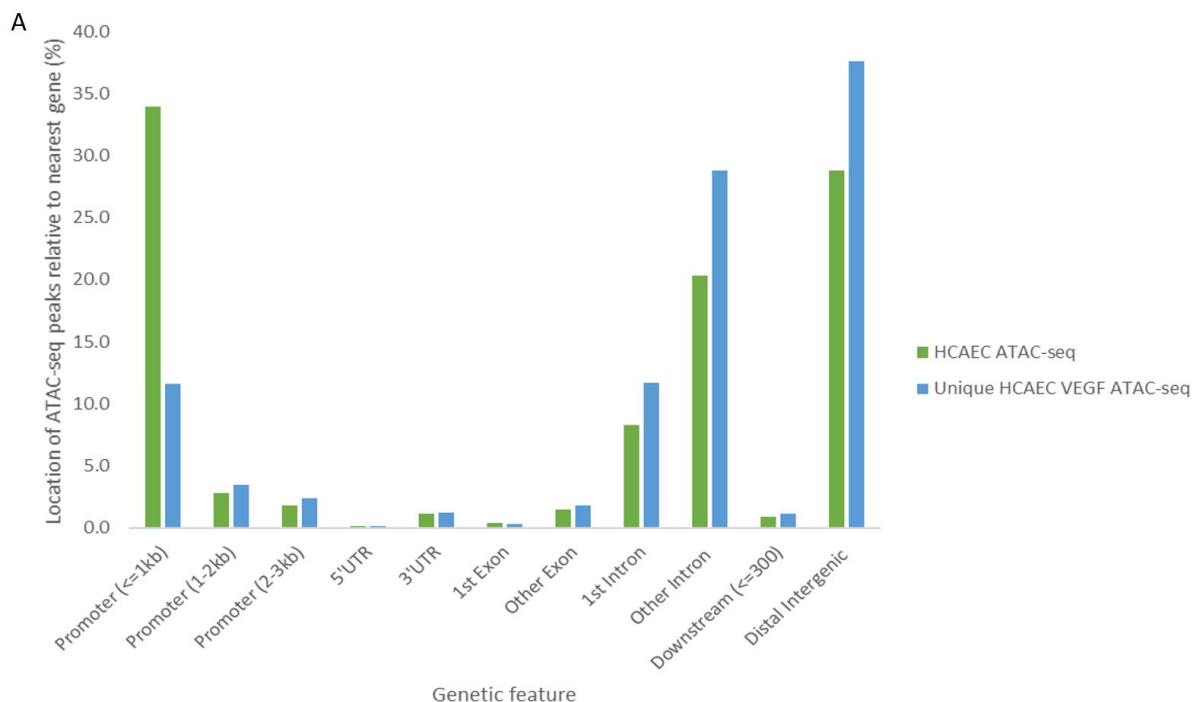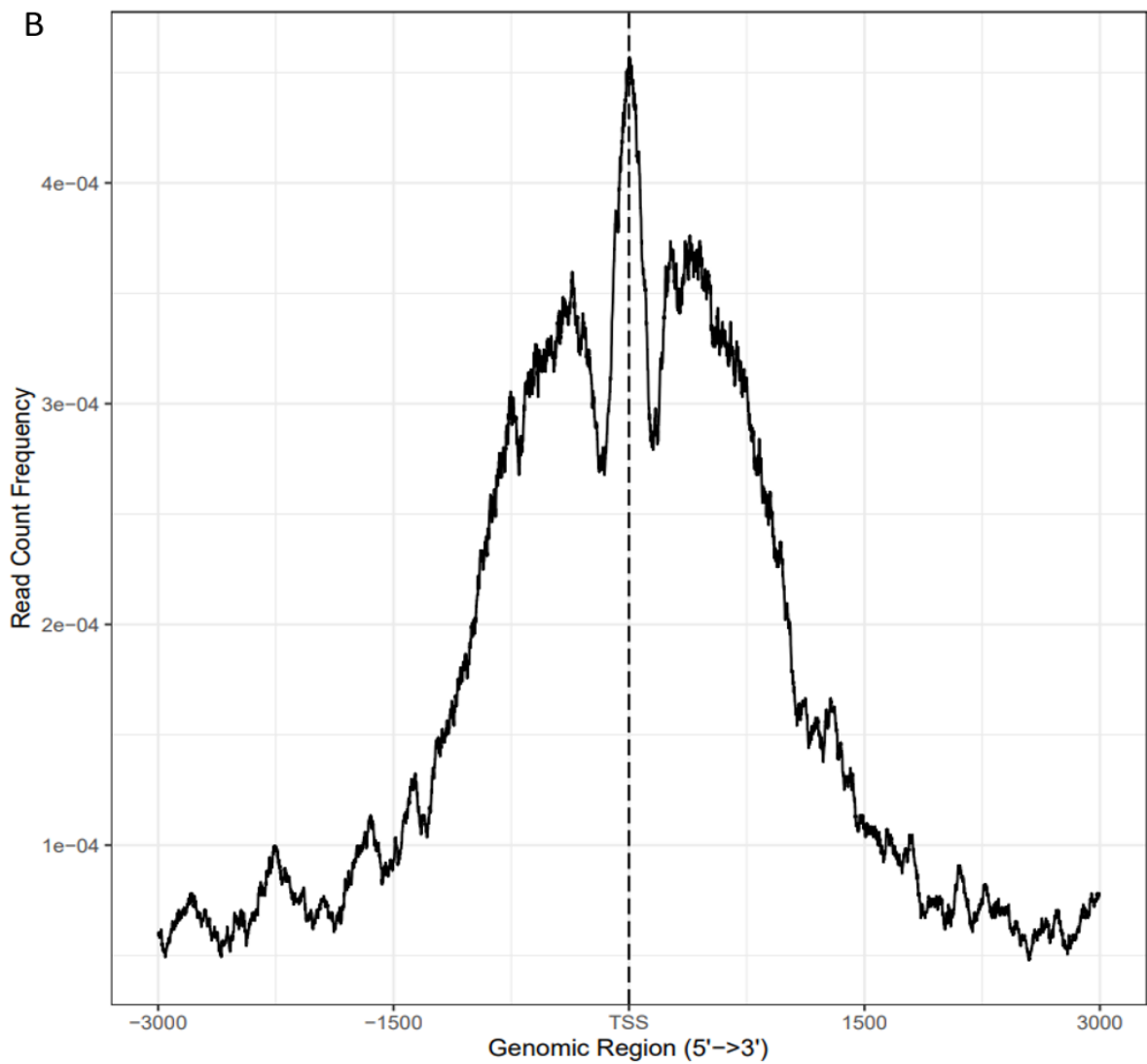
Figure 57. Distribution of accessible regions identified from unique VEGFA-stimulated and the unstimulated HCAEC using ATAC-seq. (A) Downstream refers to centre of accessible regions located less than 300 kb from the TSS. Distal intergenic regions refer to accessible region located more than 3 kb away from the TSS and more than 1 kb from TTS (B) Distribution of unique VEGFA-stimulated HCAEC ATAC-seq peaks relative to TSS.

The nearest genes to accessible chromatin were annotated using Chipseeker, which identified 4,901 genes surrounding unique VEGFA-stimulated ATAC-seq peaks. Several KEGG pathways showing enrichment in these gene sets were shared between the unique VEGFA-stimulated and unstimulated HCAEC, including Rap1, MAPK, Endocytosis, Hippo, and Ras signalling pathways. In contrast, Phospholipase D signalling pathway was specific to the unique VEGFA-stimulated HCAEC, while oxytocin, cellular senescence, and sphingolipid signalling pathways were specific to unstimulated HCAEC (**Figure 58**).
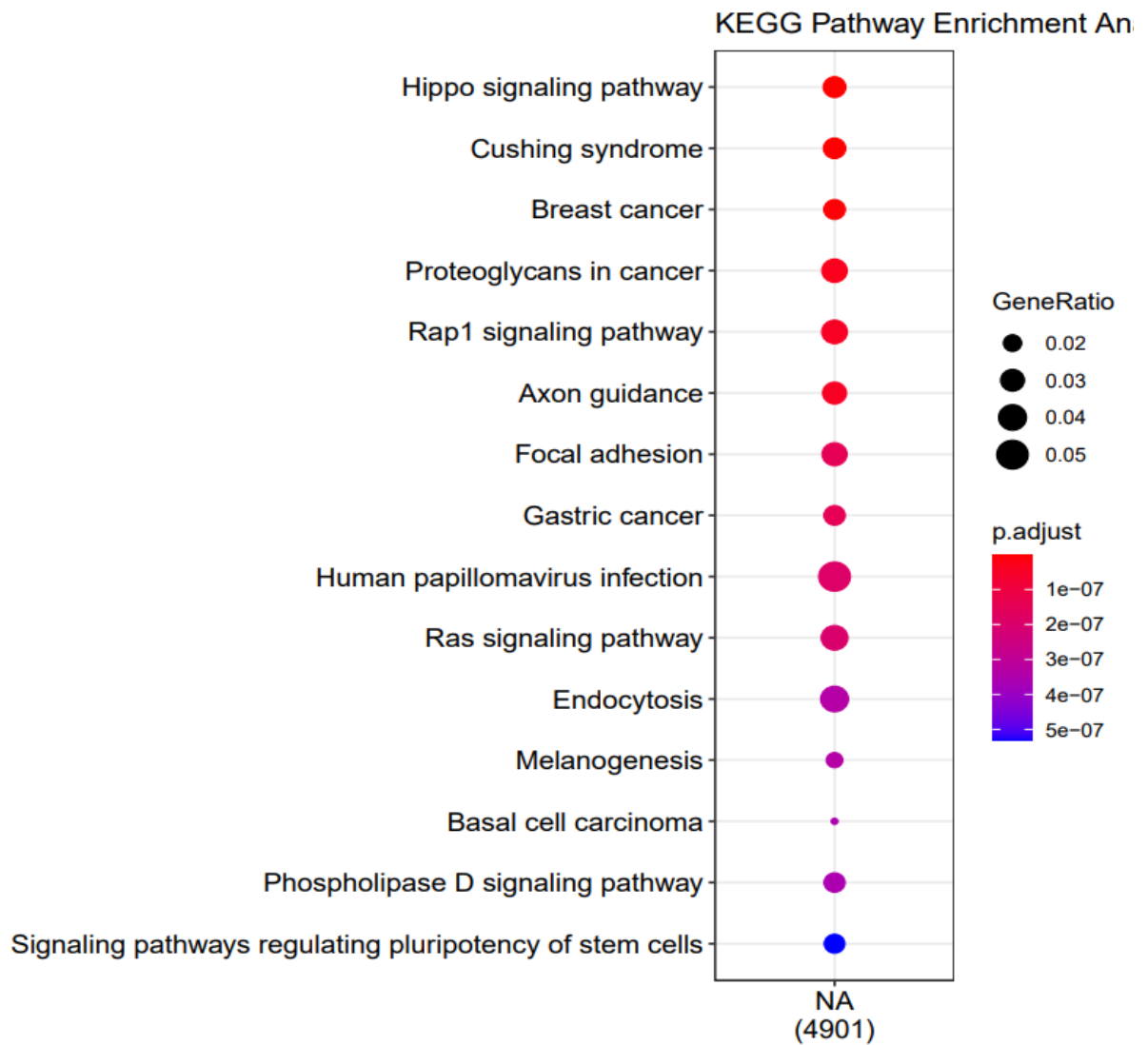
Figure 58. KEGG pathway enrichment analysis of genes surrounding vascular accessible chromatin in the unique VEGFA-stimulated HCAEC versus unstimulated HCAEC. The colour of the dot represents the adjusted *p*-value and the size of the dot represents the gene ratio. The gene ratio represents the total number of genes associated with this pathway.

6.3.8 Enrichment of transcription factor binding motifs in the unique VEGFA-stimulated HCAEC

To complement the KEGG pathway analysis from genes near accessible chromatin peaks described above, I performed enrichment analysis for transcription factor binding motifs across ATAC-seq peaks in the unique VEGFA-stimulated HCAEC using HOMER to identify regulatory motifs enriched in vascular accessible chromatin. There were a total of 287 motifs in the unqiue VEGFA-stimulated HCAEC at *p*-value $\ll 1 \times 10^{-50}$ compared to the human reference genome (hg19). The most significantly enriched motifs in the unqiue VEGFA-

stimulated HCAEC, include those of Fra1, Atf3, AP-1, Fra2, and BATF (**Figure 59**; **Table 22**). Motifs shared between the unique VEGFA-stimulated and unstimulated HCAEC, included ATF3 and Fra1. In contrast, AP-1, Fra2, and BATF were specific to the unique VEGFA-stimulated HCAEC, while CTCF, BORIS and JUNB were specific to the unstimulated HCAEC.

	Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer
	Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer
	AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer
	Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer
	BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer
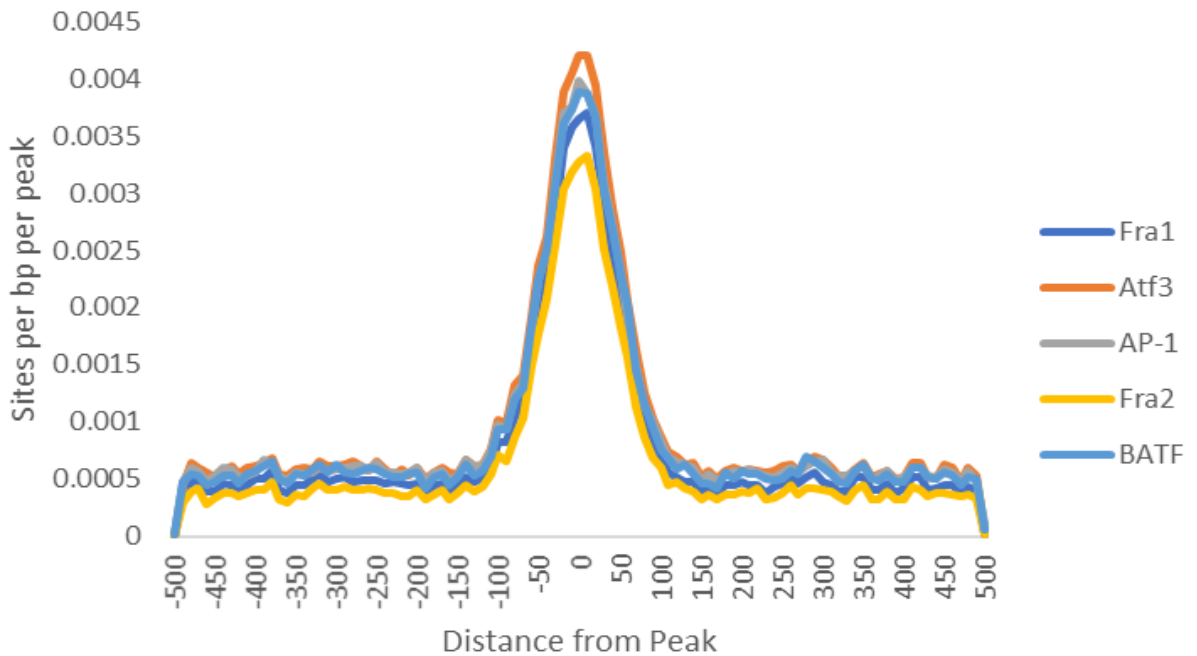
Figure 59. Histogram distribution for top five enriched motifs identified using known motif enrichment analysis in the unique VEGFA-stimulated HCAEC ATAC-seq peaks.

152

Table 22. Top 5 enriched known motifs found in the unique VEGFA-stimulated HCAEC ATAC-seq peaks.

Motif	*P*-value	Accessible chromatin with motif (%)	Inaccessible chromatin with motif (%)
Fra1	1x10-1125	8.4	1.0
Atf3	1x10-1101	9.2	1.3
AP-1	1x10-1068	9.8	1.5
Fra2	1x10-1068	7.8	0.9
BATF	1x10-1066	9.1	1.3

6.4 Discussion

Previous studies have shown large-scale chromatin remodeling on stimulation of individual cell types [395-398]. VEGFA activates both VEGFR-1 and VEGFR-2, promoting angiogenesis, vascular permeability, cell migration, and gene expression. I performed a comprehensive analysis of stimulation-dependent chromatin and gene expression changes in HCAEC using VEGFA-stimulation [399].

One major effect of stimulation was the observed widespread upregulation of gene expression. I identified 148 genes (1.5 fold) to be upregulated upon VEGFA-stimulation. A study by Liu *et al.,*[400] identified several genes that were strongly upregulated by VEGFA in an oligonucleotide microarray screen of endothelial cells, including members of the EGR and NR4A families, which agrees with my data. These genes play key role in the orchestration of long-term biological functions essential for the angiogenic response to VEGFA. The EGR family comprises four members, *EGR1*, *EGR2*, *EGR3* and *EGR4*, all characterised by a highly conserved DNA-binding domain composed of three Cys2-His2 zinc-finger motifs recognising a nine-base-pair segment of DNA (GCGGGGGCG) [401]. *EGR1*, *EGR2* and *EGR3* are transcriptional activators, whereas *EGR4* is a transcriptional repressor [402]. The NR4A family comprises *NR4A1*, *NR4A2* and *NR4A3*, which are transcription factors with roles in differentiation, proliferation, and apoptosis in various cell types. Overexpression of *NR4A1* in HUVECs causes cell cycle arrest and inhibition of *NR4A1* protein expression suppresses VEGFA-induced angiogenesis [403, 404].

EGR3 was one of the genes most strongly upregulated by VEGFA in a screen of oligonucleotide microarray screen of endothelial cells, which agrees with my data [405]. However, VEGFA-induced EGR transcriptional activation and both EGR mRNA upregulation and transcriptional activation were blocked by specific inhibitors of KDR, PKC and ERK pathway, indicating that the same signalling pathway mediated these responses. *EGR3*

silencing in HUVECs inhibited VEGFA-induced *EGR3* upregulation, cell proliferation, cell migration and tubulogenesis, identifying *EGR3* as one of the few transcription factors known to regulate VEGFA-dependent angiogenic responses. *EGR3* knockdown also strongly increased the basal level of apoptosis in HUVECs, but interestingly had no effect on the antiapoptotic response to VEGFA [406-409].

The study by Liu *et al*., [400] also examined the effects on endothelial functions of knockdown of *NR4A1*, a transcriptional regulator belonging to the NR4A family of nuclear receptors, also strongly upregulated by VEGFA [410]. Previous studies found that overexpression of *NR4A1* induced angiogenesis and *NR4A1* antisense inhibited VEGFA-driven angiogenesis [403, 404]. *NR4A1* siRNA was effective is suppressing VEGFA-induced *NR4A1* mRNA expression and decreased VEGFA-dependent cell proliferation, migration and tubulogenesis. *NR4A1* siRNA had little effect on basal or VEGFA-dependent cell survival.

PKC is a key mediator of VEGFA/VEGFAR2 signalling, and recent findings indicate that protein kinase D (PKD) is a major downstream effector of PKC-dependent pathways in response to VEGFA [405, 411]. Inhibition of PKD1 and PKD2 protein expression significantly decreased *EGR3* gene expression indicates a role for PKC-dependent activation of PKD1 and PKD2 in VEGFA-induced *EGR3* gene expression. In addition, knockdown of PKD2, but not PKD1, also significantly attenuated *NR4A1* expression. PKD1 has previously been reported to undergo translocation to the nucleus in fibroblastic cells, suggesting that PKD1 may be a proximate mediator of transcriptional upregulation, a possibility that warrants further study [412].

The majority of accessible chromatin in the VEGFA-stimulated and unstimulated HCAEC were in located in promoter regions, while the accessible chromatin in the unique VEGFA-stimulated HCAEC were located in non-promoter and enhancer regions. This is likely due to enhancer elements having a greater role in tissue-specific gene expression compared to promoters. The identification of novel enhancers and their associated gene expression will be useful to study the role of tissue-specific enhancers in human biology and disease [357].

Another major effect of VEGFA-stimulation was the marked increase in the number of accessible chromatin sites. I found 108,736 peaks in the VEGFA stimulated dataset compared to 86,811 peaks in unstimulated HCAEC. Furthermore, there are several different pathways that were enriched in the VEGFA-stimulated HCAEC compared to the unstimulated HACEC.

For instance, apelin signalling pathway is specific to the VEGFA-stimulated HCAEC. Apelin and its receptor angiotensin II protein J (APJ) are expressed in endothelial and smooth muscle cells. *In vitro* and *in vivo* studies have demonstrated that, apelin is upregulated in response to hypoxia in peripheral and cardiac tissues. Apelin/APJ deficiency is preventative against oxidative stress-linked atherosclerosis, and the apelin/APJ system works as a mediator of oxidative stress in vascular tissue. Activation of apelin, like VEGF, promotes the formation of new blood vessels and the proliferation of endothelial cells, umbilical endothelial cells and gastric cells [413].

In addition, I found 39,935 peaks in the unique VEGFA stimulated HCAEC compared to 86,811 peaks in the unstimulated HCAEC. Furthermore, there are several different pathways that were enriched in the unique VEGFA-stimulated HCAEC compared to the unstimulated HACEC. For instance, phospholipase D (PLD) signalling pathway is specific to the unique VEGFA-stimulated HCAEC. The role of PLD in VEGF signalling during angiogenesis is unclear. However, silencing of PLD2 by siRNA blocks VEGF-mediated signalling in immortalised HUVECs. PLD2 silencing inhibits VEGF-induced endothelial cell survival, proliferation, migration, and tube formation. Moreover, while PLD2-knockout mice exhibited normal development, while loss of PLD2 inhibits VEGF-mediated *ex vivo* angiogenesis [414].

Enrichment analysis of transcription factor binding motifs in the VEGFA-stimulated HCAEC compared to the unstimulated HCAEC was performed to identify regulatory motifs enriched in vascular accessible chromatin under the two examined conditions. The most significantly enriched motifs in the VEGFA-stimulated HCAEC include CTCF, BORIS, ATF3, FRA1, and BATF, whereas those from the unique VEGFA-stimulated dataset included Fra1, Atf3, AP-1, Fra2, and BATF. This analysis further implicates members of the AP-1 family as regulators of gene expression in endothelial cells, particularly following VEGFA-stimulation.

A study by Jia *et al.,* [415] showed that VEGFA induces immediate-early genes AP-1 family gene and expression of JunB under various conditions [415]. Inhibition of AP-1 DNA binding activity by adenovirus expressing a potent dominant negative form of c-Fos significantly attenuated VEGFA-induced HUVEC migration and proliferation and cyclin D1 expression. Knockdown of JunB with adenovirus expressing JunB shRNA reduces VEGFA-induced JunB expression and attenuated HUVEC migration. However, the JunB-expressing virus has no effect on VEGFA-induced cyfclin D1 protein expression and proliferation. These results

suggest that VEGFA-induced endothelial migration is mediated primarily by induction of JunB whereas the promotion of endothelial proliferation by VEGFA is mediated by JunB-independent AP-1 family members [415].

The VEGF-family constitutes a group of structurally and functionally related growth factors that play an important role in many physiological functions of endothelial cells. Given the pivotal role VEGF plays in a number of pathophysiological processes including formation of the vascular system during embryonic development, regulation of capillary growth, and the maintenance of the vasculature [416], mapping the regulatory landscape of VEGFA-stimulated cells can provide new tools to examine these processes, offer potential therapeutic targets and facilitate the characterisation of disease risk alleles.

7. Functional annotation of coronary artery disease and blood pressure GWAS index variants using vascular open chromatin maps

7.1 Introduction

The recent exponential growth in the number of robust and replicable genetic associations for many complex traits including disease is promising to improve our understanding of the underlying biological processes. The majority of the GWAS-associated variants map to non-coding parts of the genome and are likely to be involved in cis or trans gene expression regulation activity.

Typically, GWAS do not yield the causal variant and, in many loci, it is challenging to even pinpoint the causal gene. The challenge to identify the causal variant is more daunting in loci with extensive LD in the region around the sentinel SNP. Furthermore, regulatory elements are often tissue-restricted and there is need to test variants in the appropriate cell types and environmental conditions. Advancing from GWAS signals to functional annotation has been facilitated by the growing number of genome databases that enable the biological interpretation of GWAS signals [338, 339]. However, further experimental data sets are required to provide comprehensive coverage across all cell types.

Large-scale GWAS have identified many loci associated with CAD at genome-wide significance ($p < 5 \times 10^{-8}$) and a much larger number of putative susceptibility loci based on a stringent FDR cutoff [227, 239-241]. The latest analysis by the CARDIoGRAMplusC4D consortium using 181,522 CAD cases and 984,168 controls detected 241 associations at genome-wide significance [303]. In addition, this study identified a further 656 conditionally independent associations at a significance level (p-value $< 2.52 \times 10^{-5}$) approximating a 1% FDR.

High blood pressure is a major risk factor for CAD with 7.8 million deaths and 148 million disabilities worldwide in 2015 [417]. Blood pressure is a complex genetic trait influenced by the interlay of environmental and genetic factors [418, 419]. Previous genetic studies have identified and validated 274 loci, explaining ~3% of the trait variance [420-425].

The latest GWAS meta-analysis of blood pressure (BP) traits including systolic (SBP), diastolic (DBP) and pulse pressure (PP) was undertaken in over 1 million people of European ancestry. The discovery phase included UKBB [426, 427] and data from the International

Consortium of Blood Pressure-Genome Wide Association Studies (ICBP, N = 757,601) whereas replication data were from the US Million Veterans Program (MVP, N = 220,520) [428] and the Estonian Genome Centre University of Tartu (EGCUT, N = 28,742) Biobank [429]. The discovery phase identified 1,062 SNPs at $p < 1 \times 10^{-6}$ with concordant effect direction between UKBB and ICBP (two-stage analysis). Following replication, the combined meta-analysis identified 1,062 SNPs reaching genome-wide significant ($p < 5 \times 10^{-8}$) and having $p < 0.01$ in the replication data with concordant direction of effect between discovery and replication.

Blood vessel walls are comprised of ECs and VSMCs, which play an important role in vascular biology. ECs [430] and VSMCs [346] play an important role in the pathological progression of atherogenesis and BP traits but the modes and molecular mechanisms causing these are less well understood. Therefore, ECs and VSMCs are suitable cell types to investigate the regulatory potential of genetic variants for the associated with CAD and BP traits. I constructed maps of accessible chromatin from HCAEC and HCASMC as described in Chapters 5 and 6.

I used this data including those derived from HCAEC after VEGFA-stimulation to examine the latest set of association signals for CAD and BP with the aim to prioritise potentially causal variants for future functional studies.

7.2   Methods

7.2.1   Intersecting CAD-and BP-associated variants with ATAC-seq peaks

In the CAD GWAS undertaken by CARDIoGRAMplusC4D, the FDR following the meta-analysis was assessed using the 'qvalue' R package for all 20.1M variants. The p-value cut-off for a q-value of 1% was 2.52×10^{-5} and there were 47,622 variants reaching that threshold. Joint conditional analysis was performed using GCTA to identify 897 approximately independent association signals. The ATAC-seq peaks from HCAEC and HCASMC as were intersected with the 897 1% FDR sentinel CAD risk variants and all their proxies in the 1000 Genomes data set ($r^2 > 0.8$) totaling 21,461 variants [228]. Similarly, the ATAC-seq peaks in VEGFA-stimulated HCAEC were intersected with these 21,461 CAD variants [228]. The ATAC-seq peaks in HCAEC and HCASMC were intersected with the 901 sentinel BP risk variants [428] and all their proxies in the 1000 Genomes data set ($r^2 > 0.8$) totaling 14,168 variants. Similarly,

the ATAC-seq peaks in VEGFA-stimulated HCAEC were intersected with these 14,168 BP variants.

7.2.2   Prioritisation of non-coding variants in genome-wide association signals

RegulomeDB (http://regulomedb.org/) was used to annotate variants within regulatory elements, including different regulatory features such as TFBSs, chromatin states of different cell types and eQTL data. This information was combined to calculate a score for variant prioritisation. Category 1 (a-f) indicate additional annotations from the most confident to the least confident. Category 2 (a–c) indicate direct evidence of binding. Category 3 (a–b) is considered less confident in affecting binding due to a more incomplete set of evidence. Categories 4–6 lack evidence of the variant disrupting the site of binding.

HaploReg (http://compbio.mit.edu/HaploReg) was used to annotate the SNPs along with sequence conservation, regulatory protein binding, reference epigenomes, expression quantitative trait loci and regulatory motifs. Sequence conservation is defined by mammalian evolutionarily constrained elements. Regulatory protein binding sites from different cell types and treatments were obtained from the ENCODE Project ChIP-Seq data. Reference epigenomes from the Roadmap Epigenomics project for the following data sets were included: ChromHMM states corresponding to enhancer or promoter elements, from the 15-state core model and 25-state model incorporating imputed data; histone modification ChIP-seq peaks using the gappedPeak algorithm for H3K27ac, H3K9ac, H3K9me1 and H3K9me3; and DNase hypersensitivity data peaks using the narrowPeak algorithm. eQTLs were obtained from the GTEx pilot analysis v6, the GEUVADIS project and 12 other studies in order to annotate variants with their putative regulatory target genes and the tissue(s) in which genotype has been associated with gene expression level. A wide range of QTLs, including eQTLs and other molecular QTLs such as metabolite QTLs, were also extracted from the GRASP database, build 2.0.0.0. Regulatory motifs are defined by a library of position weight matrices from commercial, literature and motif-finding analysis of the ENCODE project was used to score the effect of variants on regulatory motifs using the position weight matrix-scanning process described previously.

GTEx (https://www.gtexportal.org/home/) was used to determine whether selected variants were significant eQTL by combining genetic variation with gene expression in artery-coronary, artery-tibial, and artery-aorta tissues.

7.2.3   Differential motif enrichment analysis

Bedtools getfasta [289] function was used to extract sequences from a FASTA files for each of the intervals defined in a BED file. CentriMo [310] was used to perform differential motif enrichment analysis was performed for CAD and BP SNPs. Firstly, we extracted sequences from a FASTA file corresponding to the CAD and BP SNPs by extracting the 25 bp region around the center of each SNPs. Several analyses for the CAD association data were performed, including 1% FDR CAD SNPs (control sequences) compared to HCAEC SNPs (primary sequences), 1%FDR CAD SNPs (control sequences) compared to HCASCM SNPs (primary sequences), unstimulated HCAEC (control sequences) compared to VEGFA-stimulated HCAEC (primary sequences). Similar analyses for the BP association data were performed, including BP SNPs (control sequences) compared to HCASMC SNPs (primary sequences), unstimulated HCAEC (control sequences) compared to VEGFA-stimulated HCAEC (primary sequences).

7.3   Results

7.3.1   CAD-associated variants in vascular accessible chromatin

As previously described, I mapped accessible chromatin regions in HCAEC and HCASMC cells using ATAC-seq identifying 86,811 and 209,743 peaks, respectively. In addition, I found 108,736 ATAC-seq peaks in VEGFA-stimulated HACEC. These were intersected with the 21,461 CAD risk variants corresponding to the 897 lead risk variants at 1% FDR and their proxies (r2 >0.8) from the latest CARDIoGRAMplusC4D study [303]. I identified a total of 499 CAD risk variants (Appendix, **Table 34**) located in vascular accessible chromatin peaks. In unstimulated HCAEC, I found 288 CAD-associated SNPs (in 161 CAD loci) to overlap with 86,811 ATAC-seq peaks, while in the VEGFA-stimulated cells, I found 403 CAD-associated SNPs (in 204 CAD loci) to overlap with 108,736 ATAC-seq peaks. In total, 437 CAD variants (in 216 CAD loci) were found to map in ATAC-seq peaks when combining unstimulated and VEGFA-stimulated HCAEC data. Of those, 149 variants were specific to the ATAC-seq peaks from VEGFA-stimulated cells (**Figure 60**). In HCASMC, I found 292 CAD-associated SNPs (in 167 CAD loci) to overlap with 209,743 ATAC-seq peaks.
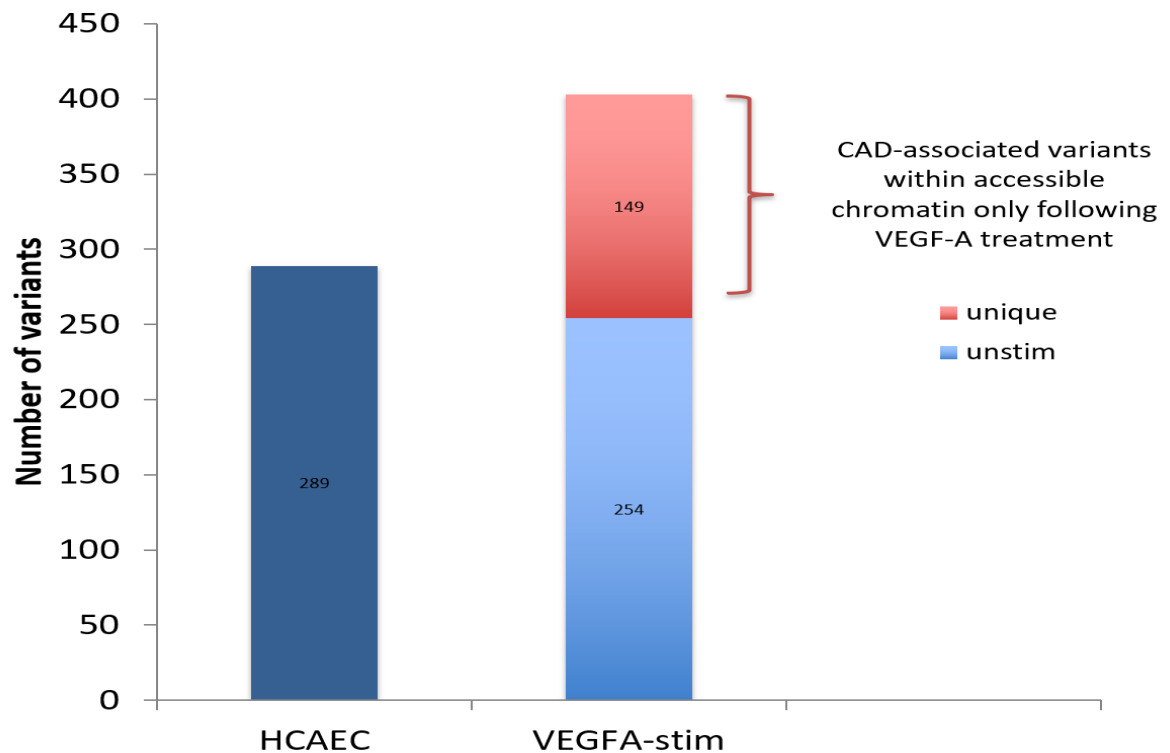
Figure 60. CAD-associated variants in unstimulated and VEGFA-stimulated HCAEC. The 149 CAD variants following VEGFA-stimulation, which were examined for their regulatory effects.

Of the 288 CAD-associated variants in unstimulated HCAEC ATAC-seq peaks, 3 were insertions and deletions and 5 were not found in the 1000 Genomes Phase 1 data. Of the 403 CAD-associated variants in VEGFA-stimulated HCAEC ATAC-seq peaks, 8 were insertions and deletions and 4 were not found in the 1000 Genomes Phase 1 data. The remaining 280 and 391 CAD SNPs overlapping ATAC-seq peaks in this study were compared to the overall set of 21,461 CAD-associated SNPs for enrichment in overlap with signatures of regulatory elements reported in Haploreg. There was clear enrichment in regulatory potential for the CAD SNPs overlapping HCAEC ATAC-seq peaks with the exception of predicted motif changed annotation, which reflects the abundance of such sites in the genome. Of the 292 CAD-associated SNPs in HCASMC ATAC-seq peaks, 10 were insertions and deletions and 5 were not found in the 1000 Genomes Phase 1 data. The remaining 277 CAD SNPs overlapping ATAC-seq peaks in this study were compared to the overall set of 21,461 CAD-associated SNPs for enrichment in overlap with signatures of regulatory elements reported in Haploreg. As with HCAECs, there was clear enrichment in regulatory potential for the CAD SNPs overlapping HCASMC ATAC-seq peaks with the exception of 'motif changed' (**Figure 61**).
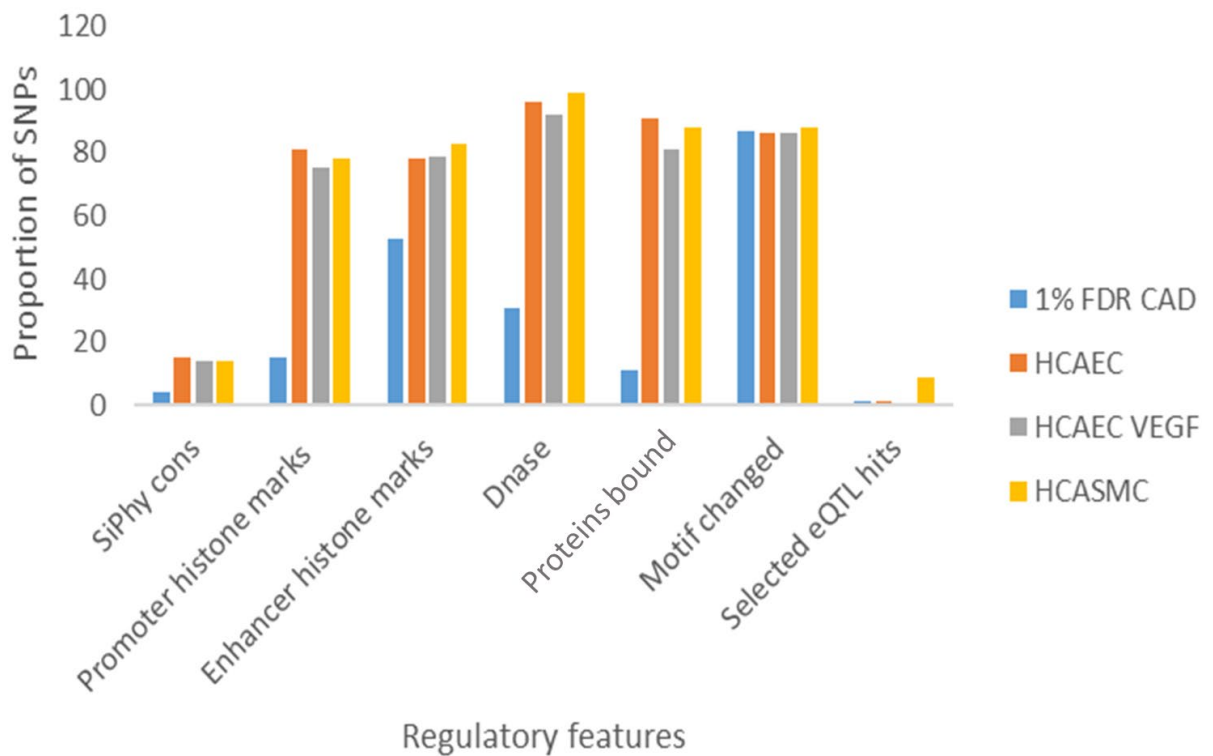
Figure 61. Distribution of the 1% FDR CAD SNPs compared to HCAEC CAD SNPs, VEGFA-stimulated HCAEC CAD SNPs and HCASMC CAD SNPs within regulatory annotations using Haploreg. The regulatory features include conserved regions, promoter, and enhancer histone marks, DNaseI, proteins bound, motifs changed and selected eQTL hits.

7.3.2 Prioritising CAD-associated variants in vascular accessible chromatin

I then examined the 499 CAD-associated SNPs located in ATAC-seq peaks for possible regulatory functions using RegulomeDB. RegulomeDB classifies variants based on their likelihood to affect binding and expression of a target gene giving a score of 1 (most likely) to 7 (least likely). For variants with strong regulatory potential (score 1-2), RegulomeDB reports a more refined classification score based on supporting evidence (**Table 23**). CAD variants that possessed strong regulatory potential with a score ≤2; include 114 HCAEC SNPs; 135 VEGFA-stimulated HCAECs and 115 HCASMCs. In total, 29 SNPs had a high priority score of 1 (a-f) of which 12 were shared across all three cell types and conditions tested (**Table 24**).

Table 23. Distribution of CAD-associated variants within vascular accessible chromatin using RegulomeDB.

Category	Description	HCAEC	HCAEC VEGF	HCASMC
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak	5	5	5
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak	4	9	6
1c	eQTL + TF binding + matched TF motif + DNase peak	1	1	0
1d	eQTL + TF binding + any motif + DNase peak	0	1	1
1e	eQTL + TF binding + matched TF motif	0	0	0
1f	eQTL + TF binding/DNase peak	13	11	6
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak	38	40	41
2b	TF binding + any motif + DNase footprint + DNase peak	51	66	55
2c	TF binding + matched TF motif + DNase peak	2	2	1
3a	TF binding + any motif + DNase peak	17	23	21
3b	TF binding + matched TF motif	0	0	0
4	TF binding + DNase peak	141	199	142
5	TF binding or DNase peak	8	25	3
6	Motif hit	3	7	0
7	No evidence	2	6	1

eQTL = expression quantitative trait loci, TF = transcription factor, DNase = deoxyribonulease

Table 24. Annotation of CAD-associated variants within vascular accessible chromatin with RegulomeDB category of 1 (a-f).

dbSNP IDs	Nearest genes	HCAEC	HCAEC VEGF	HCASMC	Rank	eQTL	Proteins bound*	Motifs altered*
rs1874883	ANAPC13				1a	ANAPC13	HIC1	HIC1
rs34826769	PAPD7				1a		NRF1	NRF1
rs9388486	CENPW				1a		ATF1,ATF2,ATF3,ATF4,CREB1	ATF1,ATF2,ATF3,ATF4,CREB1
rs2302701	SSH1				1a		EGR1	EGR1
rs6728904	RFTN2				1b	HSDP1	42	5
rs56316522	KDM3B				1b		366	EWSR1,FLI1
rs1534310	CNPY4				1b		474	NFKB1,ARHGEF7,CD40
rs1332327	LIPA				1b		188	TFAP4
rs7950280	PPFIA1				1b	PPFIA1	13	EWSR1,FLI1
rs175040	EIF2B2				1b	EIF2B2,MLH3	495	KLF12,SP4
rs1894401	FES				1b	FES	45	PAX5
rs2070737	RSPH6A				1b	SYMPK	149	KLF7,PAX5,SREBF1
rs7252126	RSPH6A				1b	SYMPK	176	CEBPA,DDIT3
rs10824083	ADK				1c		CTCF	CTCF
rs6127200	TGM2				1d		112	TEAD3
rs6442108	CDC25A				1f	ZNF589	79	
rs6442109	CDC25A				1f	ZNF589	87	
rs7793613	COG5				1f	GPR22	6	
rs10819474	IER5L				1f	PPP2R4	149	
rs10824084	ADK				1f		339	
rs28558789	CD151				1f		17	
rs1786141	VPS11				1f	HMBS	456	
rs1784460	VPS11				1f	HMBS	600	
rs179744	AP4S1				1f		398	
rs11636934	LOXL1				1f		61	
rs5002487	SREBF1				1f	SREBF1	ZNF600, POLR2A	
rs8077485	NPEPPS				1f		66	
rs4794019	IGF2BP1				1f	ATP5G1	POLR2A,JUN,GATA2,FOS	
rs2241718	TGFB1				1f	CIC	358	

*Numbers show proteins bound where over 4 proteins bind at the locus

*Numbers show motifs altered where over 4 motifs alerted at the locus

ATAC-seq data indicated enrichment of AP-1 motifs in vascular accessible chromatin as detailed in Chapter 5. To examine the potential role of AP-1 transcription factors in CAD loci, ENCODE HUVEC AP-1 binding and predicted effects on AP-1 motif were examined in the unstimulated and VEGFA-stimulated HCAEC ATAC-seq peaks. There were 156 CAD-associated SNPs that had either ENCODE HUVEC AP-1 binding or predicted AP-1 motif changed (**Table 25**).

Table 25. Enrichment of AP-1 sites in vascular accessible chromatin using Haploreg.

rsID	Nearest genes	HCAEC ATAC-seq	HCAEC VEGFA ATAC-seq	ENCODE HUVEC AP-1 binding	Predicted AP-1 motif change
rs10171574	CXCR4				
rs10183863	CXCR4				
rs1370526	DKFZp686O1327				
rs1370525	DKFZp686O1327				
rs4668380	MIR873				
rs149846585	NBEAL1				AP-1
rs3755237	NRP2				
rs3738484	MCL1				ATF3/ATF4
rs6127200	TGM2				AP-1
rs8141133	CABIN1				
rs2370576	ANAPC13				AP-1
rs1863911	ANAPC13				ATF3
rs62271373	TSC22D2				
rs10912901	KIAA0040				
rs6853156	REST				
rs80242894	LOC645513				ATF3
rs28634456	LOC645513				ATF3
rs78332141	LOC645513				ATF3
rs28714318	LOC645513				ATF3
rs28495013	LOC645513				ATF3
rs6823685	PALLD				
rs6854026	PALLD				
rs1032763	LOC285696				MAF
rs113728457	FER				MAF
rs79536589	FER				ATF3
rs75421844	FER				
rs7732639	FBN2				MAF
rs1958604	EBF1				
rs16878812	FKBP5				
rs9400476	TRAF3IP2-AS1				
rs4945487	MARCKS				
rs7770043	MARCKS				AP-1
rs73534327	MARCKS				MAF
rs9388486	CENPW				AP1/ATF3/6
rs6906478	EPB41L2				
rs118099258	GNA12				
rs7807755	RAC1				
rs6965212	LOC100129148				AP-1
rs181657044	MTUS1				
rs73551707	BMP1				

SNP	Gene				TF
rs6982502	TRIB1				
rs58914007	ST3GAL1				
rs59413766	ST3GAL1				
rs16904940	ST3GAL1				AP-1
rs1537373	CDKN2B-AS1				
rs7861724	DCAF10				AP-1
rs12555751	DCAF10				AP-1/ATF3
rs10814624	DCAF10				ATF3
rs75986742	DCAF10				ATF3
rs11789185	ENG				
rs138415650	IER5L				
rs10819474	IER5L				
rs3737180	ZEB1				
rs2433027	PLCE1				AP-1
rs71485762	OBFC1				
rs4627080	TMEM41B				
rs2593650	SLC17A6				AP-1
rs3016316	RIN1				
rs7104718	PPFIA1				AP-1/ATF3
rs2508619	KLHL35				
rs4252591	NA				ATF3
rs1784459	VPS11				AP-1
rs1786141	VPS11				MAF
rs7127978	UBASH3B				
rs7128198	UBASH3B				
rs7129071	UBASH3B				ATF3
rs12403025	SLC6A9				AP-1
rs2371494	NA				
rs12817989	FGD6				BATF
rs2302700	SSH1				AP-1/ATF3
rs2302701	SSH1				AP-1/ATF3
rs17086701	FLT1				
rs10689124	STARD13				
rs9569666	STARD13				AP-1
rs9569667	STARD13				
rs12858634	STARD13				
rs4942039	NAA16				
rs11206803	PPAP2B				
rs71445078	CARS2				AP-1
rs3986303	EXD1				AP-1
rs11636934	LOXL1				
rs2507	LOXL1				
rs4888409	CFDP1				
rs9913156	ARRB2				
rs2644383	NPEPPS				AP-1
rs1010322	ZNF652				
rs7246865	MYO9B				ATF6
rs35976034	MAP1S				
rs2241719	TGFB1				
rs2241718	TGFB1				
rs12462166	B9D2				AP-1/ATF3
rs2241712	B9D2				ATF3/ATF6
rs60572996	HEATR5B				
rs1549721	ZFP36L2				
rs2023292	ZFP36L2				
rs582384	PRKCE				

SNP	Gene				TF
rs6720415	AAK1	gray	orange		AP-1
rs4315609	NCRNA00189	gray		yellow	
rs12277907	NA	gray	orange		ATF3
rs2275117	ZMYND12	gray	orange		AP-1
rs57749886	RCOR1	gray	orange		AP-1
rs4794019	IGF2BP1	gray		yellow	
rs4968721	PECAM1	gray	orange	yellow	
rs61459202	TGFB1	gray	orange	yellow	AP-1/MAF
rs7254679	TGFB1	gray	orange	yellow	
rs10692845	APOB	gray		yellow	
rs11676780	LRRFIP1	gray			AP-1
rs16987150	TGM2	gray	orange	yellow	
rs17196927	TGM2	gray	orange	yellow	
rs17196913	TGM2	gray	orange	yellow	
rs11204859	TUFT1	gray	orange		MAF
rs6762186	FRMD4B	gray		yellow	
rs1727887	ARHGEF26	gray	orange		AP-1
rs11248061	DGKQ	gray	orange		ATF3
rs10902762	DGKQ	gray	orange		ATF3
rs13193957	MARCKS	gray	orange	yellow	
rs2502400	MARCKS	gray	orange	yellow	
rs17080102	PLEKHG1	gray	orange	yellow	
rs188227226	MTUS1	gray	orange		ATF3
rs3737178	ZEB1	gray	orange	yellow	
rs3737179	ZEB1	gray	orange	yellow	
rs11000897	ADK	gray	orange		AP-1
rs4934258	BMPR1A	gray	orange	yellow	
rs61871680	BTBD16	gray	orange	yellow	
rs1002707	SRRM1	gray			BATF
rs56348932	PPAP2B	gray	orange		AP-1
rs4888403	CFDP1	gray			MAF
rs113985803	BPTF	gray	orange		AP-1
rs71378928	WIPI1	gray			ATF3
rs3901289	BTBD16	gray	orange	yellow	
rs28372895	CD151	gray	orange		AP-1
rs8077485	NPEPPS	gray	orange		AP-1
rs7328306	FLT1		orange	yellow	
rs9569668	STARD13		orange	yellow	
rs17293632	SMAD3		orange	yellow	AP-1/BATF
rs9936936	CMIP		orange		AP-1
rs13031876	GEN1		orange		MAF
rs12105229	AOX2P		orange		AP-1
rs1781416	TUFT1		orange		ATF3
rs9606203	ARVCF			yellow	
rs62355881	NA			yellow	
rs62380877	KDM3B		orange		ATF3
rs251023	DIAPH1		orange		AP-1
rs192240803	CCNT1		orange		ATF3
rs185029099	CCNT1		orange		ATF3
rs184368932	CCNT1		orange		ATF3
rs17114036	PPAP2B		orange		AP-1
rs7256001	ZNF568		orange		MAF
rs117955557	EXOC3L2		orange		AP-1
rs2345383	TGM2			yellow	
rs112379863	KDM3B		orange		ATF3
rs631556	NAV1		orange	yellow	

rs6997760	DOCK5				
rs9569651	STARD13				
rs10911228	LAMC1				
rs78785901	SASH1				AP-1

7.3.3 Differential motif enrichment analysis of CAD-associated variants in vascular accessible chromatin

Differential motif enrichment analysis of CAD-associated variants in vascular accessible chromatin was performed to identify sequence motifs that were differentially enriched in the HCAEC prioritised CAD SNPs ($n = 288$) relative to all 1% FDR CAD SNPs, including proxies ($n = 21,461$). The most differentially enriched motifs were members of the EGR family of transcription regulatory factors (**Figure 62 A**). Similar analysis to identify sequence motifs that were differentially enriched in VEGFA-stimulated HCAEC prioritised CAD SNPs ($n = 403$) relative to unstimulated HCAEC prioritised CAD SNPs ($n = 288$) found again members of the EGR family of transcription regulatory factors as the most differentially enriched motifs (**Figure 62 B**).
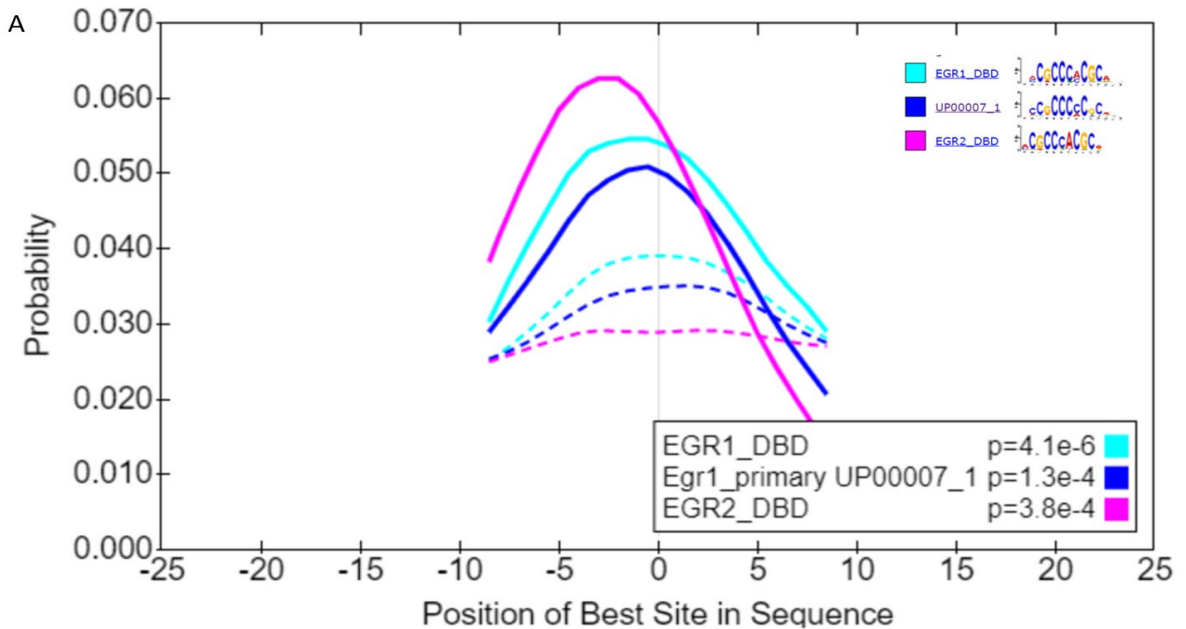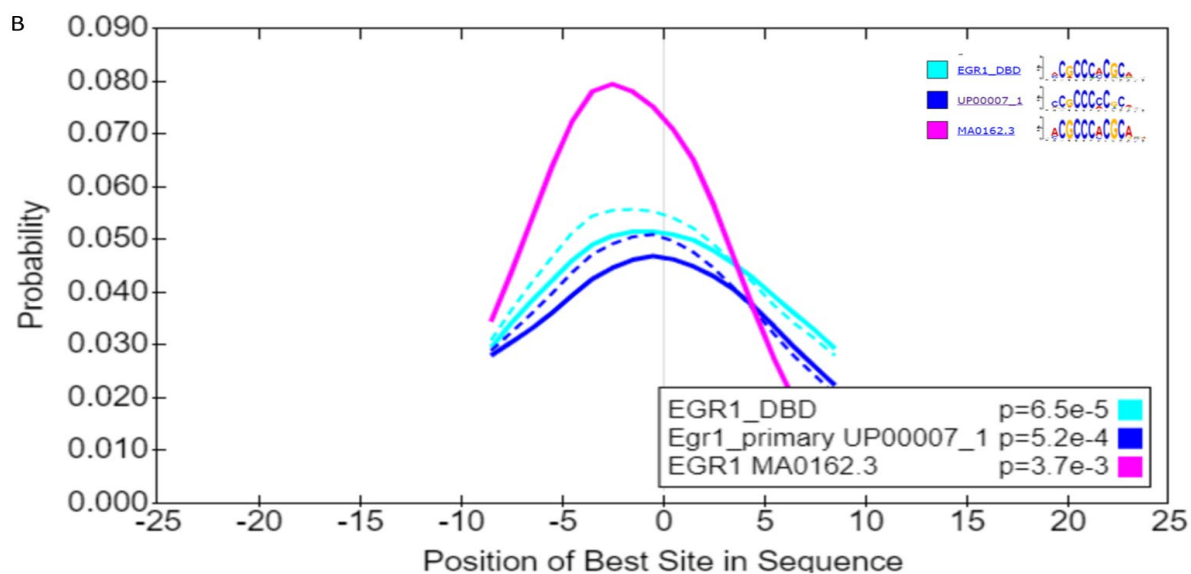
Figure 62. A) Differential motif enrichment analysis HCAEC CAD SNPs relative to 1% FDR CAD SNPs. This plot shows the central enrichment of the members of EGR family of transcription regulatory factors (solid line) and 1%FDR CAD SNPs (dotted line). B) Differential motif enrichment analysis in unstimulated compared to VEGFA-stimulated HCAEC SNPs. This plot shows the central enrichment of the members of the EGR family of transcription factors in VEGFA-stimulated HCAEC (solid line) and unstimulated HCAEC (dotted line).

Analysis to identify sequence motifs that are differentially enriched in HCASMC prioritised CAD SNPs (*n* = 292) relative to 1% FDR CAD SNPs (*n* = 21,461) found members of Hairy/enhancer-of-split related with YRPW motif protein 2 (HEY2) and EGR family of transcription regulatory factors as the most differentially enriched motifs (**Figure 63**).
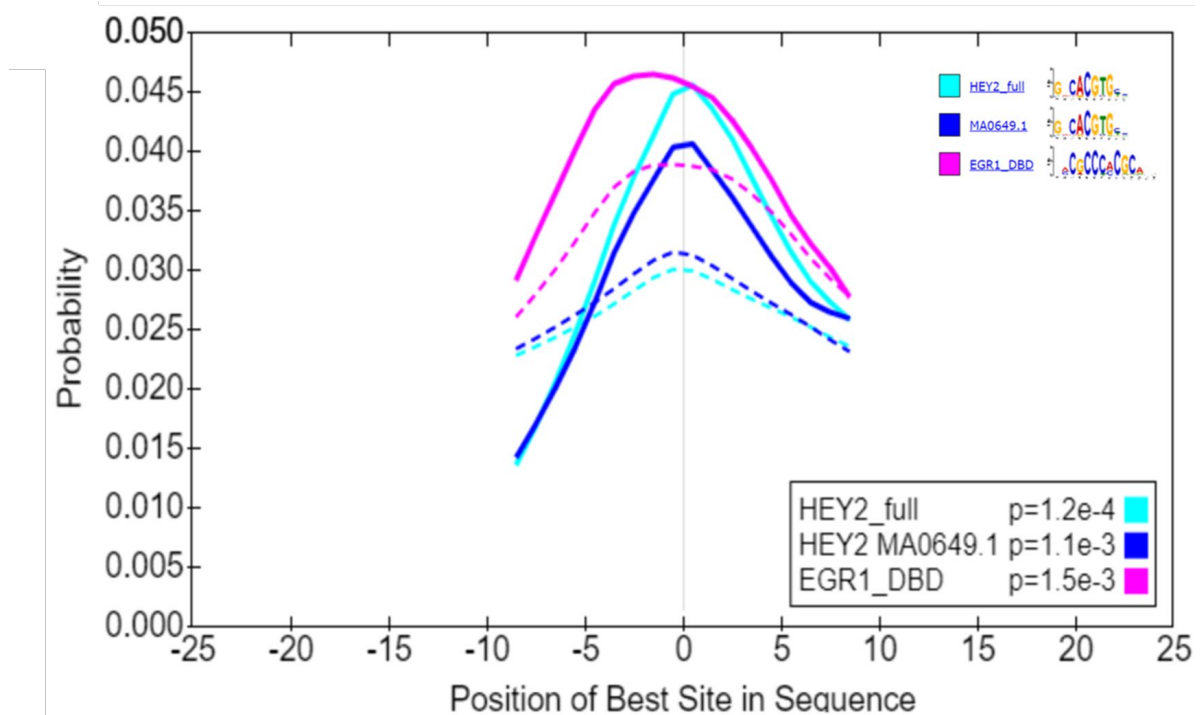
Figure 63. Differential motif enrichment analysis HCASMC CAD SNPs relative to 1% FDR CAD SNPs. This plot shows the central enrichment of members of HEY and EGR family of transcription regulatory factors in HCASMC (solid line) and 1% FDR CAD SNPs (dotted line).

7.3.4 Candidate functional CAD SNPs with regulatory potential upon VEGFA-stimulation

As VEGFA acts through VEGF receptors 1 and 2 and neuropilin 1 and 2, and given that one of the CAD signals is located at the *NRP2* locus, I examined this locus in more detail. The SNP rs3755237 is located at the *NRP2* gene and overlaps an ATAC-seq peak in VEGFA-stimulated HCAEC cells. In the unstimulated HCAEC cells there is no corresponding ATAC-seq peak. At the *NRP2* locus there are three additional CAD associated proxies, but none are located within accessible chromatin or within a coding region and therefore less likely to be functional (**Figure 64**).
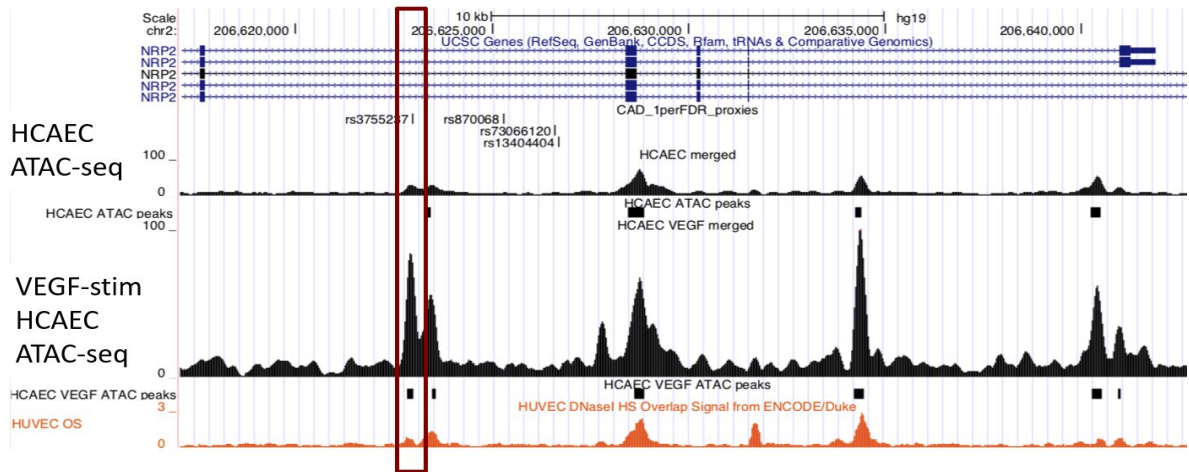
Figure 64. Open chromatin at the *NRP2* locus in VEGFA-stimulated HCAEC as displayed by the UCSC Genome Browser. The figure above shows the candidate SNP rs3755237 is located within ATAC-seq peak upon VEGFA-stimulation.

GTEx shows that the candidate SNP rs3755237 is an eQTL in three vascular tissues examined, tibial artery, coronary artery, and aorta artery. The minor allele results in decreased expression of *NRP2* (**Figure 65**).

$p = 9.8 \times 10^{-3}$ $p = 9 \times 10^{-7}$ $p = 3.1 \times 10^{-19}$
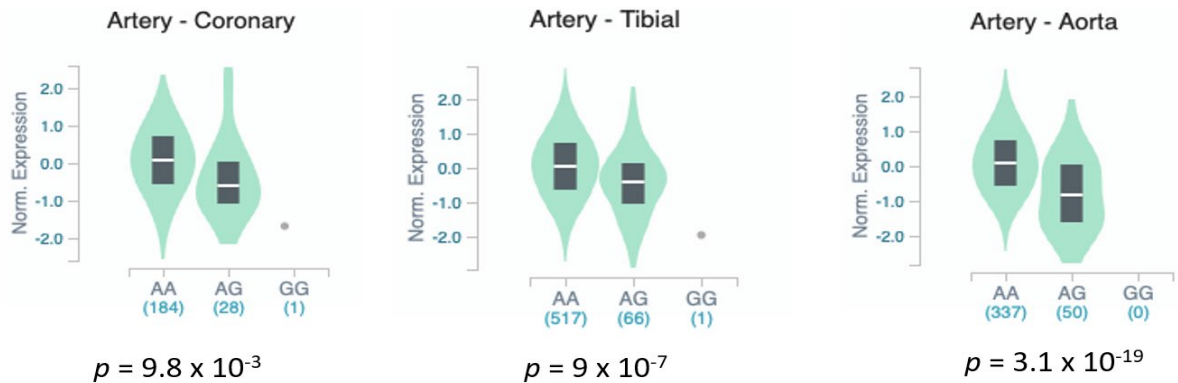
Figure 65. Allele-specific cis-eQTLs according to rs3755237 in coronary, tibial and aorta tissues in the GTEx database. Violin plots of allele-specific cis-eQTLs according to rs3755237 genotypes in coronary ($p = 9.8 \times 10^{-3}$), tibial ($p = 9 \times 10^{-7}$), and aorta ($p = 3.1 \times 10^{-19}$) tissues in the GTEx database. A and G alleles indicate the reference (minor) and alternative (major) allele types, respectively. The density distribution of the samples in each genotype is indicated by the teal region. The medium expression value of each genotype is indicated by the white line in the box plot. Higher expression is seen within the A allele.

GTEx shows that in aortic artery and tibial artery, the candidate SNP rs3755237 is the top eQTL at the *NRP2* locus (**Figure 66**).
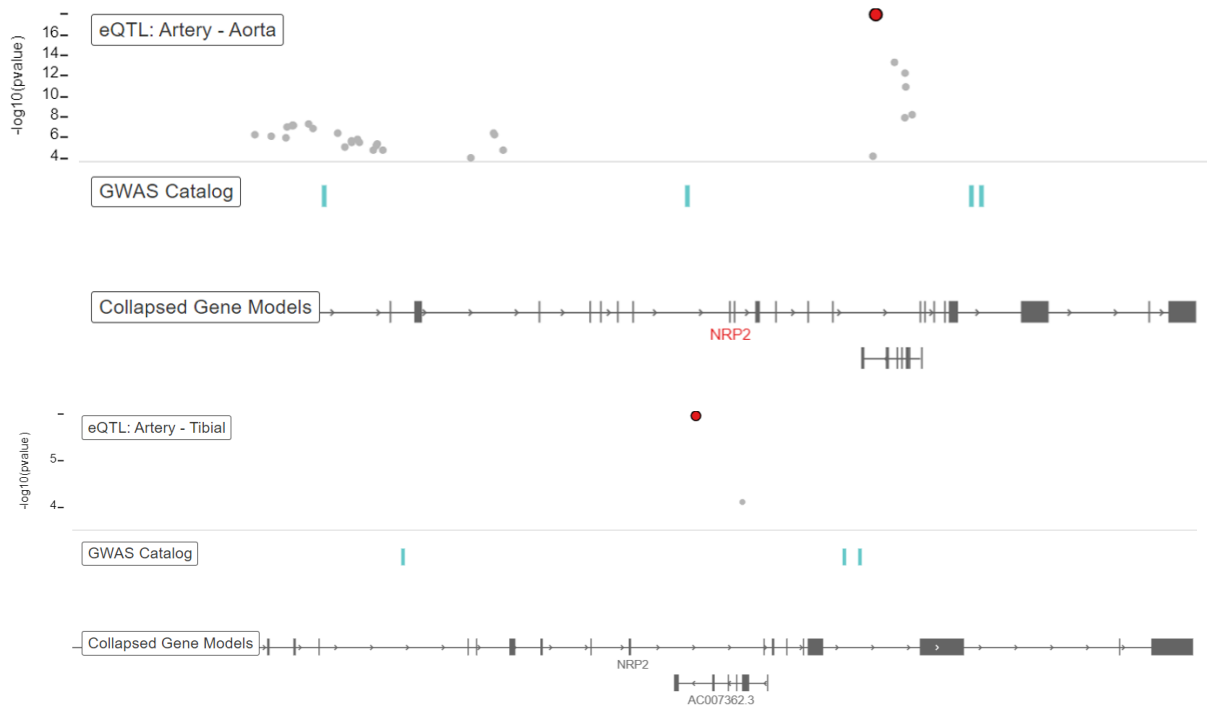
Figure 66. Top eQTL at the *NRP2* locus. A single-tissue eQTL track, where each dot represnts a siginficant cis-eQTL, the x-axis represents the genome position, and the y-axis represents the significance -log10 (*p*-value).The red dot refers to a highest significance cis-eQTL (+/- 1Mb window) at FDR<=0.05 for *NRP2* (rs3755237), and the grey dots refer to the significant eQTLs for all other genes in the genomic region.

From ENCODE project TF binding in HUVECs, cJun is bound at the site corresponding to the candidate SNP rs3755237 (**Figure 67**).
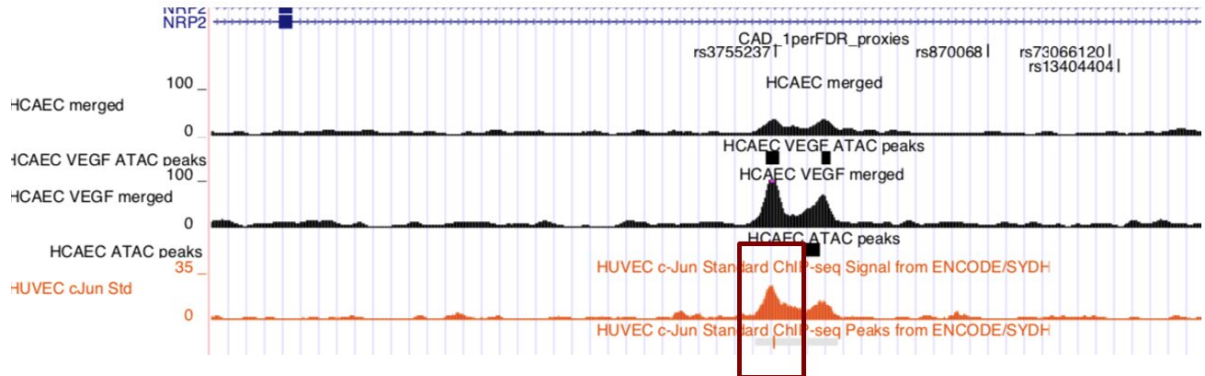
Figure 67. Accessible chromatin upon VEGFA-stimulation. This figure represents transcription factor binding sites of cJun in HUVEC as identified by ChIP-seq.

Looking at the predicted DNA binding motifs altered by the candidate SNP rs3755237, however, the TRANSFAC database suggests binding of Pou3f is destroyed by the alternative allele (**Figure 68**).

```
CCAGCTGAGGGCTGACTCATGTTTTAAGCATGGACTATGCAAAGAGACTACATCTCCTA  12.1
CCAGCTGAGGGCTGACTCATGTTTTAAGCGTGGACTATGCAAAGAGACTACATCTCCTA  0.2
```

Figure 68. Predicted Pou3f2 binding at rs3755237 A>G locus. The variant changes the alternative allele of Pou3f2, changing a match from A to G. The sequence marked 12.1 indicates the reference allele, while the sequence marked 0.2 indicates the alternative allele.

7.3.5 Prioritisation of functional CAD SNPs based on the FGWAS approach

In parallel to my work described above to prioritise candidate functional variants among the 499 CAD risk SNPs overlapping accessible chromatin in vascular cells, CARDIoGRAMplus4D had undertaken prioritisation work in the 241 CAD risk loci reaching genome-wide significance using the FGWAS approach [303] identifying 116 enriched regions. These 116 regions comprised 1,456 potential causal variants based on the corresponding 95% confidence credible sets. To combine the findings of the two studies, I intersected the 499 CAD associated variants identified in vascular accessible chromatin with the 1,456 SNPs from the CARDIoGRAMplusC4Dstudy. This led to the identification of 62 overlapping SNPs between the two sets of candidate variants (Appendix, **Table 35**). These included 9 of the 49 most statistically significant FGWAS regions (i.e. those including variant(s) with posterior probability of association (PPA)≥0.5) where the top prioritised variant is the same between the

two sets, a further 2 where the variant in accessible chromatin is within the 95% credible set but not the most significant one by FGWAS, and finally 1 where the variant in accessible chromatin is not within the 95% credible set of the FGWAS prioritised region (**Table 26**). Among the remaining 67 (out of 116) prioritised FGWAS regions not reaching statistical significance, 34 contained a prioritised by accessible chromatin variant(s) within the corresponding 95% credible set. Appendix, **Table 36** summarises the list of CAD risk SNPs that are located with accessible chromatin of vascular chromatin and have been further prioritised either by the approaches described in this Chapter and/or FGWAS.

Table 26. Overlapping SNPs between the 499 CAD associated variants identified in vascular accessible chromatin and the 1,456 SNPs from the CARDIoGRAMplusC4Dstudy.

FGWAS region	rsID	Nearest gene	PPA
47	rs11591147[a]	PCSK9	1.00
231	rs9337951 [a]	KIAA1462	1.00
387	rs2306363 [a]	SIPA1	0.91
429	rs10488763 [a]	FDX1	0.51
1055	rs116843064 [a]	ANGPTL4	1.00
1064	rs7246865 [a]	MYO9B	0.71
1151	rs582384 [a]	PRKCE	0.57
1219	rs114192718 [a]	SAP130	0.56
2164	rs2107595 [a]	HDAC9	0.98
1790	rs6823685[b]	PALLD	0.0316405
1790	rs6854026[b]	PALLD	0.019172
1066	rs11670056[c]	ELL	0.3752

[a] 9 out of 49 top FGWAS regions (variant with posterior probability of association (PPA)≥0.5) the top prioritised variant agrees between the two approaches

[b] 2 out of 49 there is a different variant by accessible chromatin but is within the 95% credible set

[c] 1 out of 49 there is a different variant by accessible chromatin and is not within the 95% credible set

7.3.6 BP-associated variants in vascular accessible chromatin

As for the CAD associated SNPs, I considered the 86,811, 108,736, and 209,743 ATAC-seq peaks in unstimulated HCAEC, VEGFA-stimulated HCAEC, and HCASMC, respectively, for intersection with the 14,648 BP risk variants corresponding to 901 lead risk variants [304] and their proxies ($r^2 > 0.8$). I identified a total of 517 BP risk variants (Appendix, **Table 37**) located in vascular accessible chromatin peaks. In unstimulated HCAEC, I found 227 BP-associated SNPs (in 122 BP loci) to overlap with 86,811 ATAC-seq peaks, while in the VEGFA-stimulated HCAEC, I found 291 BP-associated SNPs (in 135 BP loci) to overlap with 108,736 ATAC-seq peaks. In total, 316 BP variants (in 147 BP loci) were found to map in ATAC-seq

peaks when combining unstimulated and VEGFA-stimulated HCAEC data. Of those, 89 variants were specific to the ATAC-seq peaks from VEGFA-stimulated cells (**Figure 69**). In HCASMC cells, I found 383 BP-associated SNPs (in 177 BP loci) that overlapped with 209,743 ATAC-seq peaks.
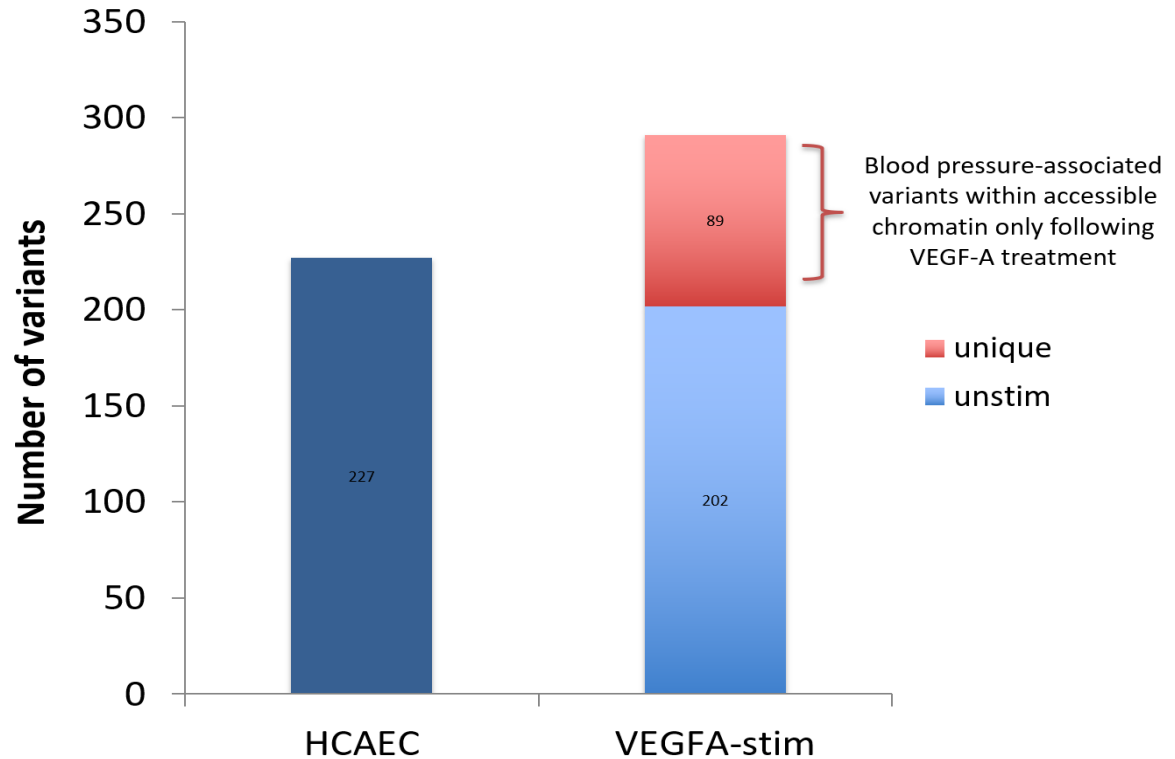
Figure 69. BP-associated variants in unstimulated and VEGFA-stimulated HCAEC. There were 89 BP additional SNPs following VEGFA-stimulation.

The 227 BP-associated variants in unstimulated HCAEC were present in the 1000 Genomes Phase 1 data. Of the 291 BP-associated variants in VEGFA-stimulated HCAEC ATAC-seq peaks, 1 was not found in the 1000 Genomes Phase 1 data. The 227 BP-associated variants in unstimulated HCAEC and the remaining 290 BP-associated variants in VEGFA-stimulated HCAEC overlapping ATAC-seq peaks in the study were compared to the overall set of 14, 648 BP-associated SNPs for enrichment in overlap with signatures of regulatory elements reported in Haploreg. As with CAD, there was clear enrichment in regulatory potential for the BP SNPs overlapping HCAEC ATAC-seq peaks with the exception of 'motif changed'. The 383 BP-associated variants in HCASMC were present in the 1000 Genomes Phase 1 data and their enrichment in Haploreg annotations compared to the overall set of 14,648 BP-associated SNPs, which follows very close pattern seen in HCAEC cells (**Figure 70**).
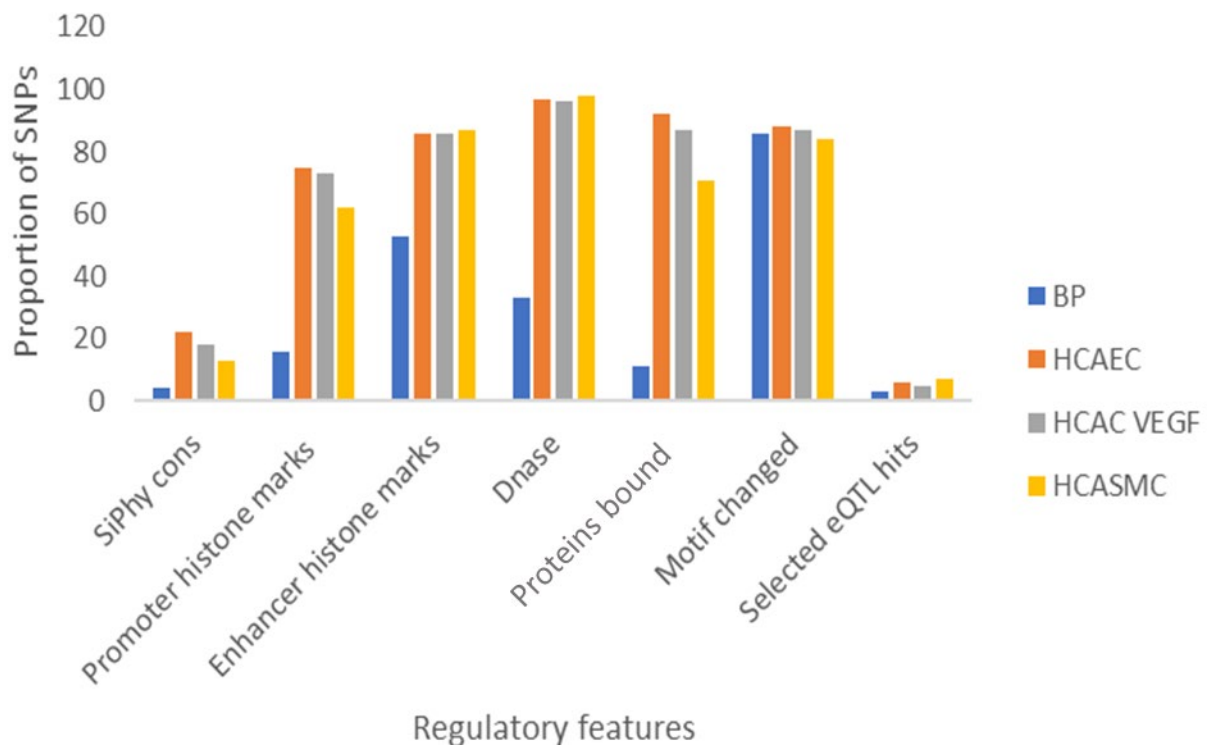
Figure 70. Distribution of the BP SNPs in open chromatin and HCAEC BP SNPs, VEGFA-stimulated BP SNPs and HCASMC BP SNPs within regulatory annotations using Haploreg. These regulatory annotations include conserved regions, promoter and enhancer histone marks, DNase, proteins bound, motifs changed and selected eQTL hits.

7.3.7 Prioritising of BP-associated variants in vascular accessible chromatin

I then examined the 517 BP-associated SNPs located in ATAC-seq peaks for possible regulatory functions using RegulomeDB. RegulomeDB classifies variants based on their likelihood to affect binding and expression of a target gene giving a score of 1 (most likely) to 7 (least likely). For variants with strong regulatory potential (score 1-2), RegulomeDB reports a more refined classification score based on supporting evidence (**Table 27**). BP variants that possessed strong regulatory potential with a score ≤2, include 80 HCAEC SNPs, 101 VEGFA-stimulated HCAECs and 133 HCASMCs. In total, 57 SNPs had a high priority score of 1 (a-f), of which 23 were shared across all three cell types and conditions tested (**Table 28**).

Table 27. Distribution of BP-associated variants within vascular accessible chromatin using RegulomeDB.

Category	Description	HCAEC	HCAEC VEGF	HCASMC
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak	2	2	2
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak	5	7	8
1c	eQTL + TF binding + matched TF motif + DNase peak	1	2	1
1d	eQTL + TF binding + any motif + DNase peak	2	2	2
1e	eQTL + TF binding + matched TF motif	0	0	0
1f	eQTL + TF binding/DNase peak	17	19	35
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak	22	29	22
2b	TF binding + any motif + DNase footprint + DNase peak	29	37	59
2c	TF binding + matched TF motif + DNase peak	2	3	4
3a	TF binding + any motif + DNase peak	9	12	30
3b	TF binding + matched TF motif	0	0	0
4	TF binding + DNase peak	128	162	189
5	TF binding or DNase peak	5	8	27
6	Motif hit	3	6	2
7	No evidence	2	1	2

eQTL = expression quantitative trait loci, TF = transcription factor, DNase = deoxyribonuclease

Table 28. Annotation of BP-associated variants within vascular accessible chromatin with RegulomeDB category of 1 (a-f).

dbSNP IDs	Nearest genes	HCAEC	HCAEC VEGF	HCASMC	Rank	eQTL	Proteins bound*	Motifs altered*
rs11826681	ARL14EP				1a	c11orf46	RELA,RELB	RELA,RELB
rs2037517	AC069368.3				1a	LOC348094	SPI1	SPI1
rs957634	RP1-212P9.2				1b	PHACTR4	44	CUX1,HBP1
rs4468203	AKR1A1				1b	Hs.396207	8	7
rs778143	AHSA2				1b	LOC339804, AHSA2	POLR2A	TRIM63
rs982146	THAP9-AS1				1b	SEC31A	232	PLAG1
rs10823977	PLA2G12B				1b	NUDT13	13	10
rs12255505	PLA2G12B				1b	NUDT13	247	TCF3,TP53
rs978458	IGF1				1b	CCDC53	14	AKR1B1,AR,NR3C1
rs12890225	RP11-638I2.4				1b		278	RARA,RARB
rs1916137	AC069368.3				1b	LOC348094	56	NF1
rs11642631	CLEC16A				1b	DEXI, LOC642755	202	VDR
rs16950065	PXN				1c	SIRT4	CTCF	CTCF
rs4808136	AC008397.1				1c	ELL	CTCF	CTCF
rs2581897	FUT10				1d	FUT10	913	MYF1
rs2394935	PLA2G12B				1d	NUDT13	6	TCF3,ZSCAN4
rs3014216	AKR1A1				1f	UROD, Hs.396207,AKR1A1	161	
rs72692299	FCF1P6				1f		35	
rs2065152	RP11-302M6.4				1f	LRRC8C	174	
rs72927138	THAP9-AS1				1f		100	
rs12643434	THAP9-AS1				1f	SEC31A	CEBPB,FOS	
rs4693555	THAP9-AS1				1f	SEC31A	FOS,JUN	
rs778594	NDUFA2				1f	WDR55	583	
rs2563335	NDUFA2				1f	SRA1, WDR55	616	
rs801168	NDUFA2				1f	SRA1, WDR55	143	
rs801183	NDUFA2				1f	SRA1, WDR55	170	
rs7807731	AC021218.2				1f	HSS00095245	13	
rs10097936	FUT10				1f	FUT10	518	
rs2732288	FUT10				1f	FUT10	276	
rs7836942	FUT10				1f	FUT10	5	
rs6996860	RP11-90P5.2				1f	BAG4, LSM1	371	
rs10819474	CRAT				1f	PPP2R4	149	
rs12253317	PLA2G12B				1f	NUDT13		AIRE
rs7094342	PLA2G12B				1f	NUDT13		
rs11000499	PLA2G12B				1f	NUDT13	6	
rs12242214	PLA2G12B				1f	NUDT13	NFIC, MEIS2	
rs12242222	PLA2G12B				1f	NUDT13	POLR2H, NFIC, MEIS2	
rs2280369	PLA2G12B				1f	NUDT13	195	
rs957486	PLA2G12B				1f	NUDT13		
rs3998306	PLA2G12B				1f	NUDT13	155	
rs3812623	PLA2G12B				1f	NUDT13	475	
rs7098453	PLA2G12B				1f	NUDT13	27	
rs10835653	ARL14EP				1f	c11orf46		9
rs10767838	ARL14EP				1f	c11orf46	RFX1, POLR2A	
rs11222385	RP11-890B15.2				1f	SNX19		5
rs11222386	RP11-890B15.2				1f	SNX19		8
rs1107479	BAZ2A				1f	NACA, PIP4K2C,RNF41,BAZ2A	703	
rs16950044	PXN				1f	SIRT4	799	
rs16950047	PXN				1f	SIRT4	823	
rs3847968	PXN				1f	SIRT4	796	
rs11633431	AC069368.3				1f	LOC348094	38	
rs1916138	AC069368.3				1f	LOC3448094	51	
rs11645809	TPSP2				1f		322	
rs9904964	RP11-227G15.3				1f	C17orf75,CDK5R1	JUND	
rs11083046	POLI				1f	C18orf54	130	
rs4808801	AC008397.1				1f	ELL	11	
rs8129943	WRB				1f	WRB	STAT3, FOS, ZNF316	

*Numbers show proteins bound where over 4 proteins bind at the locus

*Numbers show motifs altered where over 4 motifs alerted at the locus

ATAC-seq data indicated enrichment of AP-1 sites in vascular cells, and as with the CAD data this was explored in the BP data. To examine the potential role of AP-1 transcription factors in BP genetics, ENCODE HUVEC AP-1 binding and predicted effects on AP-1 motif were examined in the unstimulated and VEGFA-stimulated HCAEC ATAC-seq peaks. There were 90 BP SNPs that had either ENBCODE HUVEC AP-1 binding or predicted AP-1 motif change (**Table 29**).

Table 29. Enrichment of AP-1 sites in vascular accessible chromatin using Haploreg.

rsID	Nearest genes	HCAEC ATAC-seq peaks	HCAEC VEGFA ATAC-seq peaks	ENCODE HUVEC AP-1 binding	Predicted AP-1 motif change
rs369184	RP11-112H10.4				
rs1870735	AC021218.2				BATF
rs10109326	PREX2				ATF3
rs3998306	PLA2G12B				ATF3
rs957634	RP1-212P9.2				MAF
rs11605954	YAP1				
rs513425	LAYN				
rs4508210	RP11-890B15.2				
rs4403816	RP11-890B15.2				
rs1559671	RP11-890B15.2				
rs4403815	RP11-890B15.2				
rs16950044	PXN				
rs16950047	PXN				
rs77921082	PXN				
rs2208851	RXFP2				
rs7981648	SPRYD7				
rs7549876	TIE1				
rs3768046	TIE1				
rs4899527	LTBP2				
rs72692250	FCF1P6				
rs1981960	SETD6				
rs1981961	SETD6				
rs1343778	RP11-302M6.4				
rs57633263	FHL2				
rs56294457	FHL2				
rs134117	MN1				
rs4682994	ZNF502				
rs4680454	MLF1				
rs73156496	MLF1				
rs4693555	THAP9-AS1				

rs13109676	TET2			
rs3097175	AC005592.2			
rs114760566	HMGA1			
rs1800795	AC073072.5			
rs17558745	TGFB2			
rs12698354	AC021218.2			
rs6996860	RP11-90P5.2			
rs1375062	PTK2			
rs6578146	PTK2			
rs6578147	PTK2			
rs13259836	PTK2			
rs17369631	KANK1			
rs10819474	CRAT			
rs2271908	PLA2G12B			
rs7212271	RP11-112H10.4			
rs17396055	ARHGAP29			
rs10494474	POU2F1			
rs753641	RP11-574H6.1			
rs9419219	PPP2R2D			
rs11222383	RP11-890B15.2			
rs260508	RP11-181G12.4			
rs7219390	MXRA7			
rs6747242	FHL2			
rs6719765	FHL2			
rs61290608	PLA2G12B			
rs12149893	HAS3			
rs7936858	RP11-890B15.2			AP-1
rs1107479	BAZ2A			ATF3
rs12894937	LTBP2			AP-1
rs72692291	FCF1P6			MAF
rs3825877	ZSCAN2			AP-1
rs11645809	TPSP2			ATF3
rs1472932	EPN2			MAF
rs12945428	CCT6B			ATF3
rs9953520	NEDD4L			AP-1
rs3738484	RN7SL473P			ATF3/ATF4
rs76013375	ZMAT5			ATF6
rs2292180	ZNF502			ATF3
rs1727887	RP11-23D24.2			AP-1
rs1706003	TMEM44			MAF
rs2610989	LCORL			AP-1
rs114503346	DUSP1			ATF3
rs2732288	FUT10			AP-1

rs10819384	RP11-339B21.8				ATF3
rs3736460	FAM208B				AP-1
rs3740053	DNAJC12				AP-1/ATF3
rs115283512	CYBRD1				ATF3
rs13430254	EXOC6B				AP-1
rs143691840	PLA2G12B				AP-1
rs11000522	PLA2G12B				AP-1
rs3737801	FGR				ATF3
rs3820312	QSOX1				ATF3
rs2879813	NDUFAF6				ATF3
rs11990541	PTK2				AP-1
rs62541923	BNC2				AP-1
rs28431102	SENP7				ATF3
rs7671997	POLN				MAF
rs555754	SLC22A3				ATF3
rs11117322	BANP				AP-1
rs2237503	LIMK1				MAF

7.3.8 Differential motif enrichment analysis of BP-associated variants in vascular accessible chromatin

Differential motif enrichment analysis of BP-associated variants in vascular accessible chromatin was performed to identify sequence motifs that were differentially enriched in VEGFA-stimulated HCAEC prioritised BP SNPs ($n = 291$) relative to the unstimulated HCAEC prioritised BP SNPs ($n = 227$). This found interferon-regulatory factor (IRF) family members as the most differentially enriched motifs (**Figure 71**).
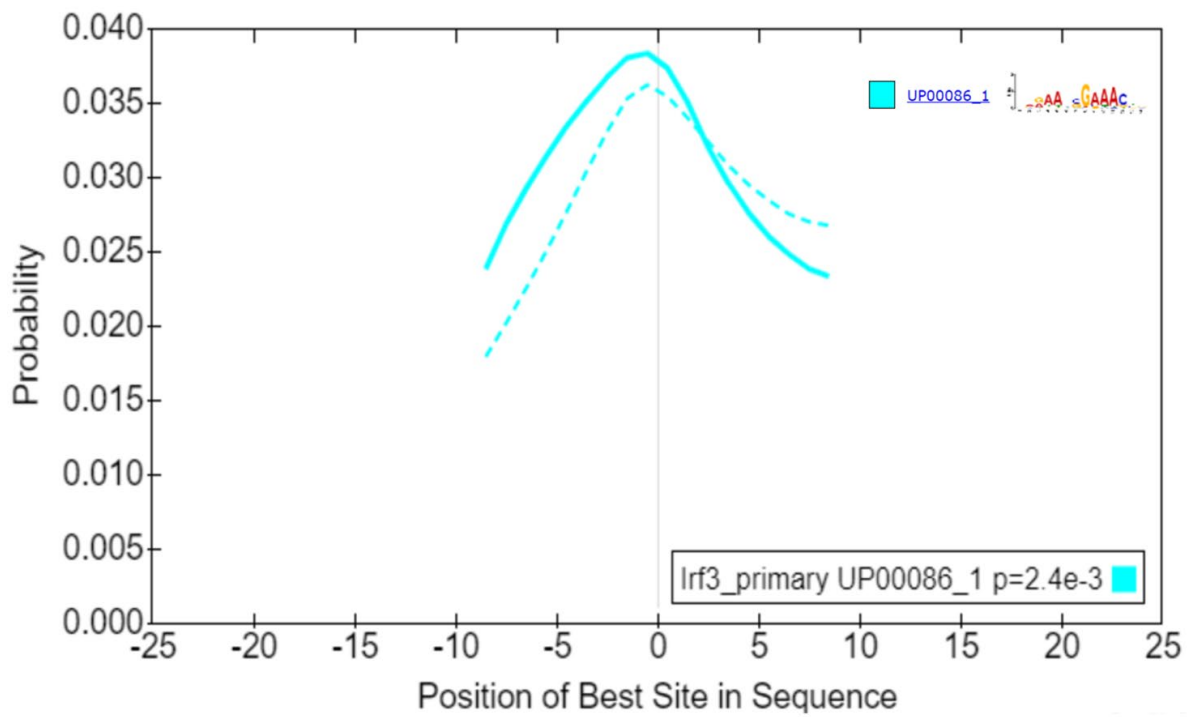
Figure 71. Differential motif enrichment analysis in unstimulated compared to VEGFA-stimulated HCAEC SNPs. This plot shows the central enrichment of IRF3 in VEGFA-stimulated HCAEC (solid line) and unstimulated HCAEC (dotted line).

Analysis to identify sequence motifs that are differentially enriched in HCASMC BP SNPs (383) relative to BP SNPs (14,648) found SP100 nuclear antigen (SP100), myeloid zinc finger 1 (MZF1) and oestrogen related receptor alpha (ESRRA) as the most significant motifs (**Figure 72**).
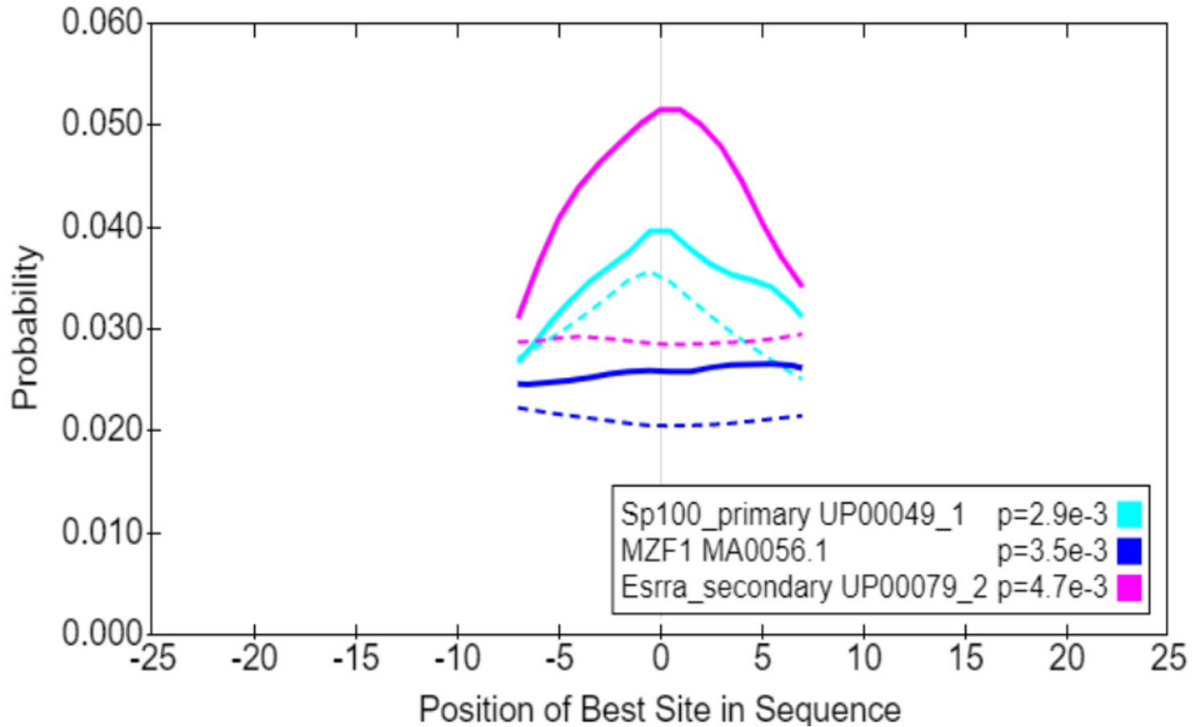
Figure 72. Differential motif enrichment analysis HCASMC BP SNPs relative to BP CAD SNPs. This plot shows the central enrichment of Sp100, MZF1, and Esrra in HCASMC (solid line) and BP SNPs (dotted line).

7.3.9 Candidate functional BP SNPs with regulatory potential upon VEGFA-stimulation

Given that the *VEGFA* locus is associated with both CAD and BP, I examined it in more detail. The SNP rs1317983 is located at the *VEGFA* locus and overlaps an ATAC-seq peak in VEGFA-stimulated HCAEC. In the unstimulated HCAEC cells there is no corresponding ATAC-seq peak. At the VEGFA locus there are eight additional BP associated proxies, but none are located within accessible chromatin or within a coding region and therefore less likely to be functional (**Figure 73**).
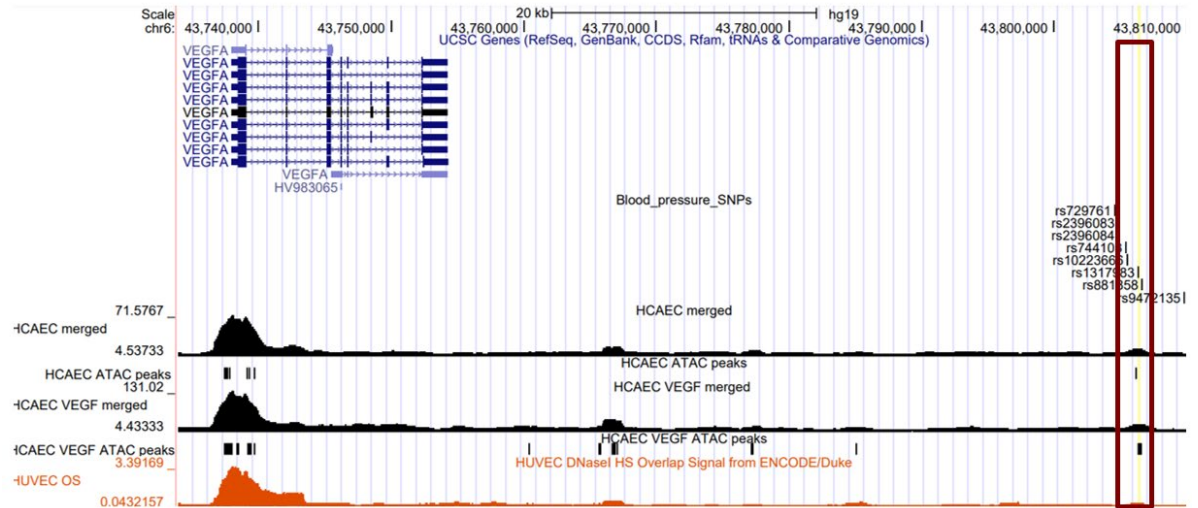
Figure 73. Open chromatin at the *VEGFA* locus in VEGFA-stimulated HCAEC as displayed by the UCSC Genome browser custom tracks. The figure above shows the candidate SNP rs1317983 is located within ATAC-seq peak upon VEGFA-stimulation.

From ENCODE project TF binding in HUVECs, CTCF is bound close to the candidate SNP rs1317983 (**Figure 74**).
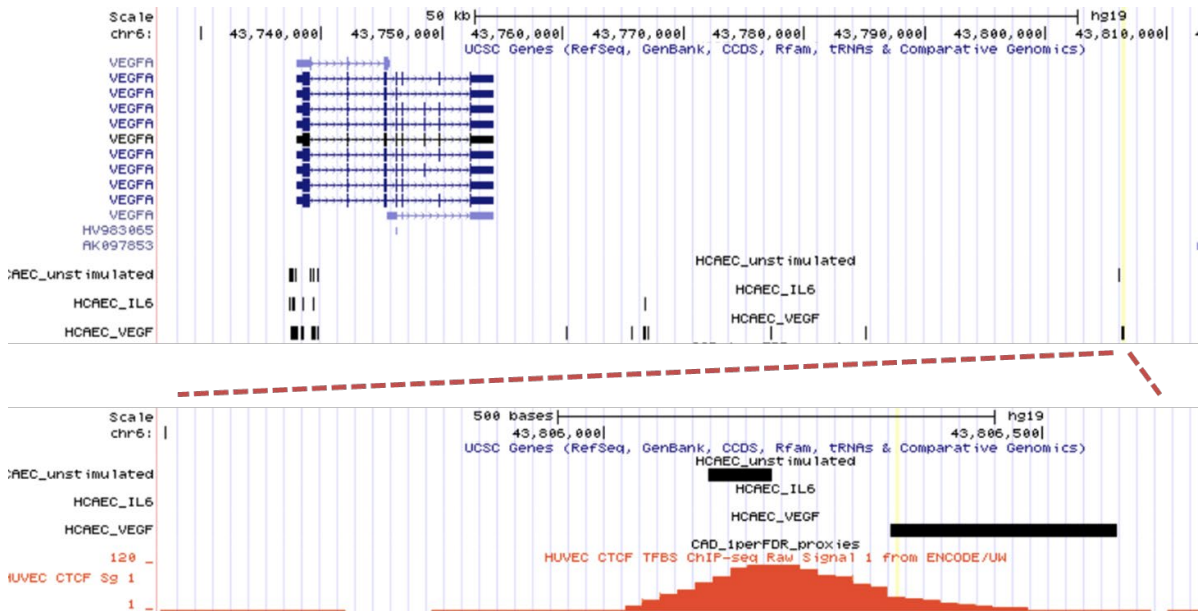
Figure 74. Accessible chromatin upon VEGFA-stimulation. This figure represents transcription factor binding sites of CTCF in HUVEC as identified by ChIP-seq.

Looking at the predicted DNA binding motifs altered by the candidate SNP rs1317983, however, the TRANSFAC database suggests binding of CTCF is destroyed by the BP risk allele (**Figure 75**).

Figure 75. Predicted CTCF binding at rs1317983 C>T locus. The variant changes the alternative allele of CTCF, changing a match from C to T. The sequence marked 10.5 indicates the reference allele, while the sequence marked 11.2 indicates the alternative allele.

7.3.10 Shared SNPs between CAD and BP

Given that BP is an established CAD risk factor, I examined which of the prioritised SNPs based on overlap with ATAC-seq peaks, were shared between CAD and BP. The 499 CAD-associated SNPs were intersected with the 517 BP-associated SNPs, which led to the identification of 9 shared SNPs. These SNPs might act through shared vascular mechanisms that play an important role in the pathogenesis of CAD and BP (**Table 30**).

Table 30. Shared SNPs between CAD and BP datasets.

Shared CAD/BP SNPs	Top eQTL (LD with CAD/BP SNPs)	Tissue	Gene	Effect Size	*P*-value
rs3738485* rs3738484*	rs41271951 (r2 = 0.0899)	Artery - Aorta	RP11-54A4.2	0.216	3x10-6
			GOLPH3L	0.093	1x10-4
			CTSS	-0.099	6x10-5
			CTSK	0.178	1x10-6
			CERS2	-0.115	8x10-6
			ARNT	-0.224	2x10-10
		Artery - Tibial	TARS2	0.104	1x10-4
			GOLPH3L	0.122	6x10-9
			ARNT	-0.173	2x10-14
			CTSS	-0.157	6x10-15
			ANXA9	0.137	2x10-4
			CERS2	-0.088	5x10-6
		Artery - Coronary	CTSS	-0.15	5x10-6
			HORMAD1	-0.363	4x10-6
			ARNT	-0.229	2x10-5
rs10819474	rs7037163 (r2 = 0.8484)	Artery - Aorta	PTPA	-0.334	8x10-25
			CRAT	-0.397	3x10-3
		Artery - Tibial	PTPA	-0.218	1x10-25
			CRAT	-0.247	2x10-24

			RP11-247A12.7	0.265	2x10-8
		Artery - Coronary	PTPA	-0.211	9x10-8
			CRAT	-0.275	4x10-10
			RP11-247A12.7	0.286	1x10-6
rs263533	rs2503701 (r2 = 0.2504)	Artery - Tibial	PRKCZ-AS1	0.335	2x10-12
		Artery - Aorta	PRKCZ-AS1	0.23	2x10-5
rs173396	rs5789616 (r2 = 0.6836)	Artery - Tibial	SWAP70	0.177	3x10-11
rs1727887* rs12493885* rs12497267✢	rs1617481 (r2 = 0.4394)[a] (r2 = 0.3684)[b]	Artery - Tibial	ARHGEF26	0.405	9x10-23
			ARHGEF26-AS1	0.361	1x10-7
			DHX36	-0.117	1x10-4
		Artery - Aorta	ARHGEF26	0.364	2x10-12
			DHX36	-0.199	6x10-8
			ARHGEF26-AS1	0.387	2x10-7
rs4808046	NA	NA	NA	NA	NA

* r2 = 1

✢ r2 = 0.9

7.4 Discussion

I used the ATAC-seq peaks in HCAEC (unstimulated and VEGFA stimulated) and HCASMC to prioritise 499 CAD-associated SNPs using RegulomeDB. I identified 114 in HCAEC, 135 in VEGFA-stimulated HCAEC, and 115 in HCASMC with a score ≤2. Moreover, of the total 499 CAD-associated loci, 29 SNPs possessed the strongest regulatory potential with a score of 1a-f. As examples:

i) SNP rs1894401 at the FES Proto-Oncogene, Tyrosine Kinase (*FES*) locus was highlighted as potentially functional following the effects of VEGFA-stimulation in HCAEC and HCASMC. The ATAC-seq peak in VEGFA-stimulated HCAEC and HCASMC and the eQTL SNP coincide in the promoter regions at the *FES* locus. The CAD risk allele increases *FES* expression, which suggests that increased gene expression is a risk factor for CAD. The SNP had a RegulomeDB score of 1b. GTEx shows this SNP to be an eQTL for *FES* in all three vascular tissues examined, aorta artery, coronary artery, and tibial artery (**Table 31**). A study by Liu *et al.,* [431], generated transcriptome and whole-genome datasets using HCASMC from 52 unrelated donors, as well as ATAC-seq in a subset of 8 donors. This study highlighted five genes, including *FES,* from jointly modelling eQTL and GWAS datasets and concluded that these genes play important role in vascular remodeling and modulate CAD risk through HCASMC [431].

ii) SNP rs10824083 at the Adenosine Kinase (*ADK*) locus was highlighted as potentially functional following the effects of VEGFA-stimulation in HCAEC. The CAD risk allele increases the *ADK* expression, which suggests that increased gene expression is a risk factor for CAD. The SNP has a RegulomeDB score of 1c and is predicted to affect the biding of CTCF. The SNP changes the alternative allele of CTCF, changing a match from C to G [432]. GTEx shows this SNP to be an eQTL for *ADK* in all three vascular tissues examined, aorta artery, coronary artery, and tibial artery (**Table 31**). A study by Xu *et al.,* [433], showed that *ADK* plays an important role in vascular inflammation. Pro-inflammatory stimuli lead to endothelial inflammation by increasing *ADK* expression reducing the level of intracellular adenosine in endothelial cells and activating the transmethylation pathway through increasing the association of *ADK* with S-adenosylhomocysteine hydrolase. Increasing intracellular adenosine by genetic *ADK* knockdown or exogenous adenosine reduces activation of the transmethylation pathway and attenuates the endothelial inflammatory response. Loss of *ADK* leads to reduced atherosclerosis and protects against ischemia/reperfusion injury of the cerebral cortex. These findings suggest that intracellular *ADK* influences vascular inflammation [433].

iii) SNP rs28558789 at the Cluster of Differentiation 151 (*CD151)* locus was highlighted as potentially functional in unstimulated and VEGFA-stimulated HCAEC. The CAD risk allele decreases the *CD151* expression, which suggests that decreased gene expression is protective for CAD. The SNP has a RegulomeDB score of 1f. GTEx shows that this SNP is an eQTL for *CD151* in all three vascular tissues examined, aorta artery, coronary artery, and tibial artery (**Table 31**). A study by Zhang et al., [434] showed *CD151* is required for endothelial capillary-like structures formed *in vitro* and *in vivo*. In addition, *CD151* plays an important role in vascular permeability. The loss of *CD151* increases actin cytoskeletol traction by up-regulating RhoA signalling and diminishes actin cortical meshwork by down-regulating Rac1 activity stabilizes *CD151*-silenced or null endothelial structure in vascular morphogenesis is stabilized by the inhibition of RhoA or activation of cAMP signalling. This showed that *CD151* plays an important role in vascular morphogenesis by promoting endothelial cell adhesions, especially cell-cell adhesion, and confining cytoskeletal tension.

Table 31. CAD-associated SNPs identified in vascular accessible chromatin.

Sentinel rsID	Ref. Allele	CAD Risk Effect size	CAD association (p-value)	Proxy rsID	eQTL gene	eQTL Coronary Artery (p-value)	eQTL effect size	eQTL Aorta Artery (p-value)	eQTL effect size	eQTL Tibial Artery (p-value)	eQTL effect size
rs7183988	0.468	0.060378	2.69E-33	rs1894401	FES	0.000011	0.22	8.10E-08	0.2	4.20E-10	0.15
rs10824133	0.1311	0.033591	0.00000646	rs10824083	ADK	0.9	0.011	4.20E-06	0.24	0.68	0.017
rs1130678	0.3175	-0.02316	0.0000223	rs28558789	CD151	1.40E-08	0.22	5.90E-14	0.29	2.70E-49	0.34
rs11466359	0.1621	-0.05522	5.47E-16	rs2241719	TGFB1	0.8	-0.014	0.37	-0.028	0.38	-0.021
rs11466359	0.1621	-0.05522	5.47E-16	rs2241718	TGFB1	0.8	-0.014	0.31	-0.032	0.31	-0.024
rs11466359	0.1621	-0.05522	5.47E-16	rs61459202	TGFB1	0.63	-0.028	0.39	-0.029	0.15	-0.037
rs11466359	0.1621	-0.05522	5.47E-16	rs7254679	TGFB1	0.88	-0.15	0.38	-0.87	0.24	-1.2
rs56062135	0.2251	-0.05779	2.27E-22	rs17293632	SMAD3	0.013	-0.16	0.0055	-0.098	0.011	-0.096

I used Haploreg to select 156 variants out of the 499 priorised variants which had either ENCODE HUVEC AP-1 binding or predicted AP-1 motif change. These included SNPs at both the *TGFB1* locus (rs2241719, rs2241718, rs61459202, rs7254679) and the *SMAD3* locus (rs17293632). The SNP rs61459202 changes the alternative allele of AP-1, changing the match from C to G [432]. GTEx shows this SNP to be an eQTL for *TGFB1* in all three vascular tissues examined, aorta artery, coronary artery, and tibial artery (**Table 31**). The SNP rs17293632 changes the alternative allele of AP-1, changing the match from C to T [432]. The CAD risk allele decreases *SMAD3* expression, which suggests that decreased gene expression is protective for CAD. GTEx shows this SNP to be an eQTL for *SMAD3* in all three vascular tissues examined, aorta artery, coronary artery, and tibial artery (**Table 31**). A study by Verrecchia *et al.,* [435], examined SMAD proteins that transduce signals from TGFB-beta receptors and regulate transcription of target genes either directly or in combination with other sequence-specific transcription factors. This study demonstrated that SMAD and AP-1 bind to their cognate cis-elements and do not interact with each other on-DNA, while off-DNA interactions occur between *SMAD3* and both c-Jun and JunB. Jun family members downregulate *SMAD3* gene transactivation, while AP-1 promoters are activated by *SMAD3* and Jun proteins. The presence of SMAD- and/or AP-1-specific cis-elements within TGF-beta-responsive genes allows dynamic modulation of gene expression, whereas interactions between SMAD and AP-1 proteins are merely an on/off mechanism to regulate TGF-beta/Smad targets. These findings demonstrated that there are several potential AP-1 sites in *TGFB* and *SMAD3* CAD loci [435].

Differential motif enrichment analysis in unstimulated HCAEC (n = 288 variants) relative to VEGFA-stimulated HCAEC (n = 403) found that the most differentially enriched motifs were members of EGR family of transcription factors. A study by Liu *et al.,* [436], demonstrated that *EGR3* was upregulated by VEGFA in an oligonucleotide microarray screen of endothelial cells. This study demonstrated that *EGR3* was upregulated in VEGFA in endothelial cells at 45 minutes. VEGFA induces a rapid increase in EGR-dependent transcriptional activation mediated via its major signalling receptor, VEGFAR2/KDR, and the PKC pathway. *EGR3* is upregulated by VEGFA to facilitate cell proliferation, migration, and tubulogenesis, suggesting that *EGR3* has an important role in VEGFA mediated endothelial functions leading to angiogenesis [436].

Differential motif enrichment analysis in the HCASMC CAD-associated SNPs (n = 292) relative to the 1% FDR CAD-associated SNPs (n = 21,461) identified members of the HEY2 family of transcription factors. A study by Baeten *et al.,* [437], highlighted the importance of the Notch-responsive gene *HEY2*, which promotes migration, proliferation and the formation of the neointimal layer in smooth muscle cells. Expression of members of the Notch signalling pathway are regulated in response to vascular injury, with an initial downregulation, followed by upregulation after several days. Several of these members have been specifically shown to regulate migration and proliferation of the neointima in injury models, including *HEY2*. This study suggested that the Notch signalling pathway drives neointimal formation after vascular injury [437].

Taken together, starting with a prioritised set of 499 CAD risk variants based on their location within accessible chromatin peaks in vascular EC and SMC cells I identified subsets of variants with strong regulatory potential that can be taken forward for functional studies. I had undertaken Electrophoretic Mobility Shift Assays (EMSAs) across a number of prioritised CAD variants from the different subgroups to experimentally assess the strength of evidence used to classify them having e.g., strong regulatory potential (RegulomeDB score 1 a-f) but this was hindered by lab closure and delays due to the COVID19 pandemic.

High blood pressure is a leading heritable risk factor for stroke and CAD. Similar to the CAD analyses, the 517 BP-associated SNPs in ATAC-seq peaks were examined using RegulomeDB. I identified 80 HCAEC SNPs, 101 VEGFA HCAEC SNPs, and 133 HCASMC

SNPs with a score ≤2. Moreover, of the total 517 BP-associated loci, 57 SNPs possessed strong regulatory potential with a score of 1 a-f. As examples:

i) SNP rs16950065 at the paxillin (PXN) locus was highlighted as potentially functional in unstimulated HCAEC, VEGFA-stimulated HCAEC and HCASMC. The BP risk allele decreases PXN expression, which suggests that decreased gene expression is protective for CAD. The SNP has a RegulomeDB score of 1c and is predicted to affect the binding of CTCF. The SNP changes the alternative allele of CTCF, changing a match from T to G [432]. GTEx shows this SNP to be an eQTL for PXN in all three vascular tissues examined, aorta artery, coronary artery, and tibial artery (Appendix, **Table 37**). A study by Yang et al., [438] examined the effect of PXN in VEGFA-induced angiogenesis in HUVECs. PXN silencing reduced the levels of lactic acid and adenosine triphosphate, downregulated HK1, HK2, and GLUT1, suppressed PI3K/AKT/mTOR signalling activation, and inhibited VEGFA-induced mitochondria injury in VEGF-A-induced HUVECs. This indicated that PXN silencing inhibited the VEGFA-induced invasion and angiogenesis of HUVECs via regulation of cell metabolism and mitochondrial damage, suggesting that PXN may be a potential target for antiangiogenic therapies [438].

I used Haploreg to prioritise SNPs which either have ENCODE HUVEC AP-1 binding or have predicted AP-1 motif change, identified 90 SNPs. These included SNPs at the Four And A Half LIM Domains 2 (*FHL2*) locus (rs57633263, rs56294457, rs6747242, rs6719765). The four SNPs at the *FHL2* locus have HUVEC AP-1 binding in the area, so these SNPs may not be altering the AP-1 consensus site. A study by Morlon and Sassone-Corsi [439], showed that *FHL2* stimulates Fos and Jun transcription, thus acting as an inducible coactivator of AP-1. This study showed that intracellular localisation of *FHL2* is controlled by signalling events [439]. A study by Hayashi *et al.,* [440], examined the mechanism of antiangiogenic signalling of *FHL2* in endothelial cells. Overexpressed *FHL2* inhibited VEGFA induced endothelial cell migration. Sphingosine kinase-1 (SK1) produces sphingosine-1-phosphate (S1P) as a potent angiogenic mediator in endothelial cells. Overexpression of *FHL2* in endothelial cells inhibited VEGFA-induced SK1 activity, phosphatidylinositol 3-kinase activity, and phosphorylation of Akt and eNOS. Overexpression of *FHL2* had no effect on S1P-induced Akt phosphorylation. VEGFA-stimulation decreased the binding of *FHL2* and *SK1*. Depletion of *FHL2* by siRNA increased migration accompanied with SK1 and Akt activation. Injection of *FHL2* mRNA into Xenopus embryos resulted in inhibition of vascular development. This study led to the

conclusion that *FHL2* may regulate phosphatidylinositol 3-kinase/Akt via direct suppression of the SK1-S1P pathway in endothelial cells [440].

Differential motif enrichment analysis was performed to identify the most significantly enriched motifs in the unstimulated HCAEC BP SNPs (n = 227) relative to the VEGFA-stimulated HCAEC BP SNPs (n = 291). The most differentially enriched motif is IRF3 is a member of the IRF family. IRF3 forms a complex with CREBBP, which is translocated to the nucleus activating interferons alpha and beta transcription. A study by Simons *et al.,* [441], studied mRNA expression of essential genes for arteriogenesis and angiogenesis. They found that mRNA expression levels of pro-inflammatory cytokines (tnfα, il6, ccl2) and growth factor receptor (VEGFR2), were decreased in gastrocnemius muscles of Irf3−/− mice compared to C57BL/6 mice. This study showed that IRF3 play an important role in angiogenesis and arteriogenesis [441].

Differential motif enrichment analysis was performed to identify the most significantly enriched motifs in the HCASMC BP-associated SNPs relative to total BP-associated SNPs. The most differentially enriched motifs were SP100, MZF1, and ESRRA. Binding of ETS1 to DNA is reduced by SP100, leading to inhibition of transcriptional activity on the MMP1 and uPA promoters. Interferons upregulate SP100 expression, which have antiangiogenic properties. SP100 plays an important role in HUVECs by directly antagonizing ETS1-mediated morphological changes. These data show that SP100 negatively modulates ETS1 processes [442, 443].

Given the established causal relationship between BP and CAD, I examined whether there were shared prioritised SNPs in my analyses to identify whether these SNPs may through the same vascular mechanisms leading to CAD. Intersection of the two datasets identified 9 SNPs in shared between CAD and BP that might act thought shared vascular mechanisms. As examples:

i) Cathepsin K (*CTSK*) and Cathepsin S (*CTSS*) loci, rs3738485, rs3738484, play an important for CAD. However, the SNP rs41271951 is top eQTL at the *CTSS* locus, which suggests that the two SNPs rs3738485 and rs3738484 may have independent effects compared to the top eQTL in the pathogenesis of CAD. A study by Cong-Lin Liu, *et al*., [444] have demonstrated that *CTSK* and *CTSS* play an important role in human atherosclerotic and abdominal aortic

aneurysm lesions [445, 446], studies of cultured cells, experimental animal models, and human biomarkers all suggest the participation of cathepsins in cardiovascular diseases. *CTSS* might have no direct role in cardiovascular disease, but proteases might contribute to cardiovascular disease by regulating the adaptive and innate immune response via antigen presentation in antigen-presenting cells (APCs), such as dendritic cells, B cells, and macrophages, and through Toll-like receptor (TLR) activation. *CTSS* have been implicated in antigen and autoantigen presentation in APCs and epithelial cells [447-449], which are cell types essential to atherogenesis [450]. *CTSK* and *CTSS* activate TLR7 and TLR9, thereby suppressing the immune regulatory activity of regulatory T cells [451, 452], which have a protective role against cardiovascular disease [453-455].

ii) Switching B Cell Complex Subunit SWAP70 (*SWAP70*) locus, rs173396, is genome-wide significant for CAD. However, the SNP rs5789616 is top eQTL at the *SWAP70* locus, which could indicate that SNP rs173396 might have either an independent effect or thought the top eQTL in the pathogenesis of CAD. *SWAP70* independently of RAS, transduces signals from tyrosine kinase receptors to RAC. Ras pathway functions downstream of a wide range of cytokines including VEGFA, is active in the growing vascular front of developing and pathological vascular networks [456]. A study by Zheng *et al.,* [457], identified 55 high-confidence causal genes for CAD, among which 15 genes ranked the highest priority. Among these 15 genes is *SWAP70*, which plays an important role in cellular proliferation and vascular remodeling.

iii) Rho Guanine Nucleotide Exchange Factor 26 (*ARHGEF26*) locus, rs1727887, rs12493883, rs12497266, is genome-wide significant for CAD. However, the SNP rs1617481 is top eQTL at the *ARHGEF26* locus could indicate that two SNPs rs1727887 and rs12493883 may have independent effects compared to the top eQTL in the pathogenesis of CAD. *ARHGEF26* is expressed in endothelial cells where it acts downstream of adhesion molecules to control the formation on membrane docking structures in leukocyte trans-endothelial migration [458]. Whilst the lead SNP causes a valine to leucine change at amino acid position 29, this is not conserved and not predicted to be damaging. A study by Klarin *et al.,* [234], identified 15 novel loci, among which is *ARHGEF26*. *ARHGEF26* impacts the transendothelial migration of leukocytes and appear to function through pathways beyond known CAD risk factors.

Overall, the prioritised variants in this study should be viewed in the context of possible limitations of the employed annotation tools. RegulomeDB, Haploreg and GTEx databases cover non-coding regions in the genome [306] but provide information only for assessing variant effects on chromatin state, conservation and regulatory motifs. Therefore, they do not cover all regulatory mechanisms or pathways. CAD depends on the interactions of multiple genetic and environmental factors which are likely to be linked with epigenetic effects on chromatin modification, histone modification and DNA methylation. Another caveat is that not all SNPs were present in these databases although this affected only a very small number of variants. Finally, I used cell lines in generating accessible chromatin maps. A study by Tcheandjieu *et al.,* [459] highlights the importance of cell cycle, replication, and growth gene-sets as well as endothelial, fibroblast, and smooth muscle cells in the pathogenesis of CAD. Expanding the study in primary cells as well as in tissue from CAD patients as the landscape of accessible chromatin may differ between patients and non-diseased individuals. To establish a comprehensive set of prioritised CAD risk variants additional data should be collected for all vascular cells and differentiation states, to overcome the above limitations.

8. Conclusion and future outlook

Interpretation of GWAS signals is challenging due to the fact that the majority of associated variants lie outside of protein-coding regions. Many diseases associated variants are likely to impose risk by altering functional DNA elements that regulate gene expression. The study established the accessible chromatin landscape of two cell types relevant to CAD and identified 499 putative disease-risk variants located in these regions. I employed a number of *in silico* prioritisation approaches to refine the above set, but ultimately experimental validation will be key to select variants/genes for future functional studies.

We are currently working on a Massively Parallel Reporter Assay (MPRA) using self-transcribing active regulatory region sequencing (STARR-seq) [460], which will assess whether the SNPs we have prioritised through ATAC-seq and CAD association affect gene expression in vascular cells. STARR-seq relies on regulatory DNA segments cloned downstream into the 3'UTR of a reporter. Active enhancers will activate the upstream promoter and undergo transcription, which results in reporter transcripts among cellular RNAs. Each reporter transcript contains the reporter gene combined with the enhancer sequence. These reporter transcripts can be isolated by targeted PCR and detected by high-throughput sequencing. One of the advantage of this technique over the conventional MPRA is that the tested sequence is used as a "barcode", quantifying enhancer activity [461]. We have employed an oligo design which incorporates a sliding shift frame format, where the variant is represented 5 times. Each variant allele is located in the centre for one of the five oligos and shifted 10 and 20 bp to the 5' and 3' for the other 4, providing the opportunity for verification of allele-specific signals. The oligo format comprises a 15b linker, 6b barcode, 194b sequence of interest, and 15b linker. The linkers are homologous to the h_STARR_ORI vector. The 6 bp barcode allows the downstream differentiation of wild-type, variant, and control sequences. The sequences represent ~499 CAD risk variants with 50 positive and negative control sequences. The positive and negative controls were designed using Human ACtive Enhancer to interpret Regulatory variants, which includes active, transcribed enhancers derived from global run-on sequencing, precision nuclear run-on sequencing and cap analysis of gene expression data [462].

A study by Selvarajan *et al.,* [463] included a STARR-seq MPRA to examine liver-specific CAD variants. This led to the identification of 177 variants that showed allele specific effects on expression. Examining the overlap between the 499 CAD SNPs in vascular accessible

chromatin and 177 CAD SNPs in liver-specific enhancer elements led to the identification of a shared variant: rs17293632 at the *SMAD3* locus. This variant is located in intron 1 and may be functional in multiple cell types; we aim to determine functionality in endothelial cells with and without VEGFA-stimulation.

There are several well-established tools examine the downstream effects of prioritised variants on vascular pathways, including CRISPR/Cas9 [464]. CRISPR/Cas9 relies on creating a targeted double strand break in the DNA by fusing a non-specific endonuclease to engineered protein with specific DNA recognition sites. In the case of CRISPR/Cas9 using a guide RNA that contains a base pairing DNA recognitions site with an additional motif that binds catalytically the active Cas9 protein. Once created the double strand breaks there are two possible outcomes, the most common will be for the cell repair mechanism to perform non-homologous end joining, which results in imperfect repair causing indel mutation at the cut sites. The other outcome will be for the cell to repair the break using a template DNA by process known as homology directed repair using existing genomic templates where we can add a DNA template to create a specific mutation [464].

A recent advance in genome-editing is a technique known as base editing [465]. This method does not rely on the creation of the double stranded DNA break. Rather than going through the non-homologous end joining process, this method is able to generate single base transition mutations, using modified Cas9. The targeting system is similar to CRISPR, but instead of the Cas9 creating a double strand break, inactive Cas9 is fused to nuclear base deaminase enzyme. Depending on the enzyme used, single nucleotides can be targeted and altered. One advantage of this technique is that mutations can be introduced into non-dividing cells without the introduction of unwanted indels caused by non-homologues end joining process [465].

CRISPR-based methods would ideally be used to create edited isogenic endothelial cell lines at several CAD-associated loci, including *NRP2*. One goal for generating an *NRP2* variant, for example, would be to examine the effects of gene expression by performing qPCR for *NRP2* expression or to examine the global effects of gene expression. Also, *NRP2* protein expression could be compared on the surface of endothelial cells by performing flow-cytometry. The edited cells would be examined how they affect vascular growth and remodelling, looking for four basic features, including matrix degradation, migration, proliferation, and morphogenesis.

There are several established techniques to look at these and most of these methods have been performed with HUVECs.

In matrix degradation assay, angiogenesis can be facilitated by activating angiogenetic proteins or releasing matrix or membrane bound growth factors. In a migration assay, the number of cells that pass thought semipermeable membrane over certain period would be measured [466]. In a proliferation assay, simply measuring the number of cells for the time-taken for cells to migrate over scratched area of endothelial cell monolayer [466]. In a morphogenesis assay, substrates and endothelial cells are plated on, in or between the substrates [466]. We would compare the vascular growth and remodeling of the two isogenic cell lines to determine whether this variant alone has the ability to affect angiogenesis *in vitro* [466].

Ultimately, the goal of this work is to demonstrate that certain CAD-associated variants are functioning through regulation of blood vessel growth or morphology, in order to develop novel drug targets for CAD and related diseases of the vasculature.

NRP2 is upregulated by vascular cells in response to injury/inflammation and is involved in many processes associated with vascular diseases. *NRP2* could be used as a novel therapeutic target or biomarker [467]. Several studies have explored methods to inhibit VEGFA-C/NRP2 signalling, which promotes endothelial cell migration, survival, and angiogenesis [468-470]. For example, a study found monoclonal antibodies targeting VEGFA-C/NRP2 are able to decrease metastasis in animal xenograft experiments [468], suggesting that VEGFA-C/NRP2 signalling could be developed therapeutically in the treatment of vascular diseases. *NRP2* inhibition may promote or prevent clinical complications associated with CAD depending on the context. Thus, a better understanding of *NRP2* function in CAD will be of interest as a potential therapeutic agent [467].

This study overall has demonstrated how the examination of chromatin structure in vascular cells can be used to prioritise functional variants in combination with bioinformatics data. In conjunction with additional experimental data this could lead to a new direction for future therapeutics directed towards vascular processes in the treatment/prevention of CAD and related traits.

9. Appendix

9.1 Supplementary tables

Table 32. A list of ATAC-seq oligos for PCR.

Custom Nextera PCR Primers	Index	Sequence
Custom Nextera PCR Primer 1	Ad1_noMX	AATGATACGGCGACCACCGA GATCTACACTCGTCGGCAGC GTCAGATGTG
Custom Nextera PCR Primer 2	Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATAC GAGATTCGCCTTAGTCTCGT GGGCTCGGAGATGT
	Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATAC GAGATCTAGTACGGTCTCGT GGGCTCGGAGATGT
	Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATAC GAGATTTCTGCCTGTCTCGT GGGCTCGGAGATGT
	Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATAC GAGATGCTCAGGAGTCTCGT GGGCTCGGAGATGT
	Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATAC GAGATAGGAGTCCGTCTCGT GGGCTCGGAGATGT
	Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATAC GAGATCATGCCTAGTCTCGT GGGCTCGGAGATGT
	Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATAC GAGATGTAGAGAGGTCTCGT GGGCTCGGAGATGT
	Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATAC GAGATCCTCTCTGGTCTCGT GGGCTCGGAGATGT
	Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATAC GAGATAGCGTAGCGTCTCGT GGGCTCGGAGATGT
	Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATAC GAGATCAGCCTCGGTCTCGT GGGCTCGGAGATGT

	Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATAC GAGATTGCCTCTTGTCTCGT GGGCTCGGAGATGT
	Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATAC GAGATTCCTCTACGTCTCGT GGGCTCGGAGATGT
	Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATAC GAGATATCACGACGTCTCGT GGGCTCGGAGATGT
	Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATAC GAGATACAGTGGTGTCTCGT GGGCTCGGAGATGT
	Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATAC GAGATCAGATCCAGTCTCGT GGGCTCGGAGATGT
	Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATAC GAGATACAAACGGGTCTCGT GGGCTCGGAGATGT
	Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATAC GAGATACCCAGCAGTCTCGT GGGCTCGGAGATGT
	Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATAC GAGATAACCCCTCGTCTCGT GGGCTCGGAGATGT
	Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATAC GAGATCCCAACCTGTCTCGT GGGCTCGGAGATGT
	Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATAC GAGATCACCACACGTCTCGT GGGCTCGGAGATGT
	Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATAC GAGATGAAACCCAGTCTCGT GGGCTCGGAGATGT
	Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATAC GAGATTGTGACCAGTCTCGT GGGCTCGGAGATGT

	Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATAC GAGATAGGGTCAAGTCTCGT GGGCTCGGAGATGT
	Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATAC GAGATAGGAGTGGGTCTCGT GGGCTCGGAGATGT

Table 33. Reactome pathway analysis of the 148 top up regulated genes and their underlying pathways upon VEGFA-stimulation.

Pathway	*p*-value	Genes
NGF-stimulated transcription	1.17E-13	EGR2;ARC;EGR3;EGR4;RRAD; NAB1;NAB2;F3
Nuclear Events (kinase and transcription factor activation)	8.94E-13	MEF2C;EGR2;ARC;EGR3;EGR4 ;RRAD;NAB1;NAB2;F3
Signalling by NTRK1 (TRKA)	2.93E-10	MEF2C;EGR2;ARC;EGR3;EGR4 ;RRAD;NAB1;NAB2;PIK3R1;F3
Signalling by NTRKs	2.41E-09	MEF2C;EGR2;ARC;EGR3;EGR4 ;RRAD;NAB1;NAB2;PIK3R1;F3
Circadian Clock	4.37E-09	PER1;ZC3H12C;MEF2C;NFIL3; CRY1;CREM;RORA;CCRN4L
Interleukin-4 and Interleukin-13 signalling	6.47E-08	VCAM1;CEBPD;BCL6;LIF;ROR A;PIK3R1;PTGS2;VEGFA
Unfolded Protein Response (UPR)	3.77E-07	XBP1;DDIT3;EIF2AK3;DNAJB9 ;ATF3;HERPUD1
Signalling by Receptor Tyrosine Kinases	6.25E-07	MEF2C;EGR2;EGR3;EGR4;RRA D;NAB1;NAB2;PRDM1;PIK3R1 ;F3;VEGFA;KITLG;ARC;MYCN ;SOCS6;HBEGF
PERK regulates gene expression	2.60E-06	DDIT3;EIF2AK3;ATF3;HERPU D1
ATF4 activates genes in response to endoplasmic reticulum stress	9.94E-06	DDIT3;ATF3;HERPUD1
Interleukin-10 signalling	3.35E-05	LIF;CXCL1;PTGS2;CXCL3;CX CL2
Cytokine Signalling in Immune system	5.63E-05	MEF2C;VCAM1;CEBPD;MX1;L IF;IFIT5;CXCL1;RORA;PIK3R1; CXCL3;PTGS2;CXCL2;VEGFA; RELB;ISG20;BCL6;REL;MAP3 K8;SOCS6;BIRC3
ATF6 (ATF6-alpha) activates chaperone genes	6.60E-05	XBP1;DDIT3
ATF6 (ATF6-alpha) activates chaperones	1.07E-04	XBP1;DDIT3
Signalling by Interleukins	1.10E-04	MEF2C;VCAM1;CEBPD;LIF;C XCL1;RORA;PIK3R1;PTGS2;C XCL3;CXCL2;VEGFA;BCL6;M AP3K8
Signalling by ALK	3.80E-04	MYCN;PRDM1;PIK3R1
TP53 regulates transcription of several additional cell death genes	6.96E-04	BCL6;NDRG1

whose specific roles in p53-dependent apoptosis remain uncertain		
Response of EIF2AK1 (HRI) to heme deficiency	7.92E-04	DDIT3;ATF3
Regulation of TP53 Expression	0.001495	PRDM1
EGR2 and SOX10-mediumted initiation of Schwann cell myelination	0.002326	EGR2;NAB1;NAB2
FOXO-mediated transcription of cell death genes	0.004295	BCL6;DDIT3
Transcriptional regulation of white adipocyte differentiation	0.004576	ZC3H12C;EGR2;KLF5;CEBPD
VEGFA ligand-receptor interactions	0.005766	VEGFA
VEGFA binds to VEGFAR leading to receptor dimerisation	0.005766	VEGFA
Response of EIF2AK4 (GCN2) to amino acid deficiency	0.005897	DDIT3;NAB1;ATF3
Nuclear Receptor transcription pathway	0.007544	NR4A2;NR4A1;NR4A3;RORA
rRNA modification in the mitochondrion	0.008847	MT-RNR2;MT-RNR1
MECP2 regulates transcription factors	0.008847	MEF2C
TNFR1-induced NFkappaB signalling pathway	0.008878	TNFAIP3;BIRC3
Calcineurin activates NFAT	0.012509	NFATC2
Cellular responses to stress	0.013712	KDM6B;ZC3H12C;MEF2C;XBP1;NAB1;DDIT3;EIF2AK3;DNAJB9;RORA;ATF3;HERPUD1;VEGFA
IRE1alpha activates chaperones	0.014287	XBP1;DNAJB9
Cellular responses to stimuli	0.016662	KDM6B;ZC3H12C;MEF2C;XBP1;NAB1;DDIT3;EIF2AK3;DNAJB9;RORA;ATF3;HERPUD1;VEGFA
PI3K events in ERBB4 signalling	0.019021	PIK3R1;HBEGF
Regulation of gene expression by Hypoxia-inducible Factor	0.019021	VEGFA
NOD1/2 Signalling Pathway	0.019027	TNFAIP3;BIRC3
Regulation of TNFR1 signalling	0.020285	TNFAIP3;BIRC3
BMAL1:CLOCK,NPAS2 activates circadian gene expression	0.022936	ZC3H12C;CCRN4L
TNF receptor superfamily (TNFSF) members mediating non-canonical NF-kB pathway	0.023993	BIRC3
Regulation of KIT signalling	0.026656	KITLG;SOCS6
CLEC7A (Dectin-1) induces NFAT activation	0.026656	NFATC2
STAT3 nuclear events downstream of ALK signalling	0.026656	PRDM1
TICAM1, RIP1-mediumted IKK complex recruitment	0.029434	BIRC3

TP53 Regulates Transcription of Cell Death Genes	0.029864	BCL6;NDRG1
TP53 Regulates Transcription of Genes Involved in G1 Cell Cycle Arrest	0.032321	E2F7
TFAP2 (AP-2) family regulates transcription of growth factors and their receptors	0.035315	VEGFA
Signalling by SCF-KIT	0.035329	KITLG;PIK3R1;SOCS6
TNF signalling	0.035329	TNFAIP3;BIRC3
Cellular response to starvation	0.038215	DDIT3;NAB1;ATF3
PI3K events in ERBB2 signalling	0.038411	PIK3R1;HBEGF
GAB1 signalosome	0.041607	PIK3R1;HBEGF
IKK complex recruitment mediated by RIP1	0.044899	BIRC3
Chemokine receptors bind chemokines	0.046467	CXCL1;CXCL3;CXCL2
Interferon alpha/beta signalling	0.047646	ISG20;MX1;IFIT5
Signal Transduction	0.048498	RRAD;NAB1;MIS12;NAB2;TNFAIP3;CXCL1;PRDM1;PIK3R1;CXCL3;HDAC8;CXCL2;RND1;ZC3H12C;HHAT;DUSP10;SOCS6;DUSP5;MEF2C;EGR2;EGR3;EGR4;FST;PTCH2;F3;DKK2;VEGFA;ARL4C;NR4A1;KITLG;ARC;BMP2;MYCN;MT-RNR2;RGL1;HBEGF;F2RL3;BIRC3

Table 34. 499 CAD-associated SNPs identified in vascular accessible chromatin.

Proxy rsID	Sentinel rsID	Nearest genes	*p*-value (artery-coronary)	*p*-value (artery-aorta)	*p*-value (artery-tibial)
rs34128716	rs6736093	MERTK	0.47	0.03	0.000000015
rs114192718	rs114192718	SAP130	0.096	0.19	0.000097
rs10171574	rs4954580	CXCR4	0.2	0.047	0.0071
rs10183863	rs4954580	CXCR4	0.19	0.041	0.054
rs1370526	rs16825477	DKFZp686O1327	NA	NA	NA
rs1370525	rs16825477	DKFZp686O1327	NA	NA	NA
rs13009734	rs35611688	ACVR2A	0.038	0.25	0.00012
rs4668380	rs66492942	MIR873	NA	NA	NA
rs13417165	rs62172372	CALCRL	0.00000001	1.40E-21	6.40E-69
rs6728904	rs199718686	RFTN2	0.03	0.051	0.02
rs144505847	rs148812085	NBEAL1	1.70E-14	3.00E-50	4.30E-60
rs149846585	rs148812085	NBEAL1	5.20E-14	5.50E-50	1.00E-65
2:203736202_CCG_C	rs148812085	NBEAL1	NA	NA	NA

rs149163995	rs148812085	NBEAL1	5.20E-14	5.50E-50	1.00E-65
rs3755237	rs13404404	NRP2	0.0098	3.10E-19	0.0000009
rs35949748	rs4140748	PID1	0.97	0.042	0.14
rs11677932	rs11677932	COL6A3	0.26	0.0000016	0.00097
rs3738484	rs11585169	MCL1	0.9	0.11	0.79
rs3738485	rs11585169	MCL1	0.79	0.1	0.78
rs200369604	rs11585169	MCL1	NA	NA	NA
20:8193159_C TT_C	rs6055645	PLCB1	NA	NA	NA
rs2378256	rs6088595	NCOA6	0.79	0.18	0.0001
rs6127200	rs59827666	TGM2	0.13	0.0016	0.078
rs73193808	rs73354869	NCRNA00189	NA	NA	NA
rs8141133	rs12484557	CABIN1	0.18	0.97	0.81
rs5756069	22:36249338_T A_T	NA	NA	NA	NA
rs11711766	rs7429594	CDC25A	5.90E-09	2.00E-26	5.00E-33
rs13063437	rs9790198	GXYLT2	0.0026	0.29	6.00E-11
rs13063441	rs9790198	GXYLT2	0.0026	0.33	9.10E-11
rs9832514	rs62268492	ZBTB20	0.73	0.66	0.5
rs2370576	rs9876658	ANAPC13	NA	NA	NA
rs1874883	rs9876658	ANAPC13	1.20E-42	2.60E-92	2.60E-153
rs1863912	rs9876658	ANAPC13	1.20E-42	2.60E-92	2.60E-153
rs1863911	rs9876658	ANAPC13	1.20E-42	2.60E-92	2.60E-153
rs10682759	rs9876658	ANAPC13	1.30E-41	1.80E-88	9.10E-140
rs62271373	rs62271373	TSC22D2	0.2	0.0058	0.46
rs6773653	rs11721038	MECOM	0.17	0.64	0.2
rs10912901	rs1057239	KIAA0040	0.99	0.0038	0.002
rs13137144	rs1377586	DGKQ	3.50E-13	9.50E-40	1.30E-43
rs4652304	rs10913568	C1orf220	0.00063	0.0018	0.017
rs66790703	rs781663	REST	0.097	0.0000029	0.00054
rs6853156	rs781663	REST	0.13	0.00000067	0.0011
rs6852182	rs781663	REST	0.085	0.0000014	0.00013
rs56155140	rs781663	REST	0.14	0.00000067	0.00048
rs10911205	rs12739316	LAMC1	0.0075	0.014	0.000000032
rs1518492	rs1401153	ADAMTS3	0.027	0.013	0.34
rs7677068	rs2127821	LOC645513	NA	NA	NA
rs71612597	rs2127821	LOC645513	NA	NA	NA
rs28571712	rs2127821	LOC645513	NA	NA	NA
rs28408407	rs2127821	LOC645513	NA	NA	NA
rs75396180	rs2127821	LOC645513	NA	NA	NA
rs80242894	rs2127821	LOC645513	NA	NA	NA
rs28634456	rs2127821	LOC645513	NA	NA	NA
rs78332141	rs2127821	LOC645513	NA	NA	NA
rs28714318	rs2127821	LOC645513	NA	NA	NA

rs28495013	rs2127821	LOC645513	NA	NA	NA
rs28441656	rs2127821	LOC645513	NA	NA	NA
4:120377089_T G T	rs2127821	LOC645513	NA	NA	NA
rs1809406	rs2127821	LOC645513	NA	NA	NA
rs2306456	rs2127821	LOC645513	NA	NA	NA
rs6823685	rs869396	PALLD	0.95	0.99	0.58
rs6854026	rs869396	PALLD	0.81	0.99	0.42
rs34826769	rs13163920	PAPD7	0.83	0.000014	0.000079
rs34338480	rs13163920	PAPD7	0.83	0.000014	0.000079
rs2477069	1:197604105_ CA_C	NA	NA	NA	NA
rs1499593	1:197604105_ CA_C	NA	NA	NA	NA
rs1032763	rs2652682	LOC285696	NA	NA	NA
1:201886769_T A_T	rs761491286	NA	NA	NA	NA
rs12516449	rs2452753	MIR4280	NA	NA	NA
rs113728457	rs112949822	FER	0.0016	1.50E-15	1.00E-10
rs79536589	rs112949822	FER	0.0016	1.50E-15	1.00E-10
rs75421844	rs112949822	FER	0.0014	2.40E-14	3.00E-10
rs7732639	rs6883598	FBN2	0.0035	7.10E-12	1.30E-18
rs60454446	rs6883598	FBN2	0.0017	8.10E-13	6.30E-21
rs56366217	rs62380939	KDM3B	0.19	0.3	0.012
rs55803185	rs62380939	KDM3B	0.33	0.41	0.029
rs11242456	rs773023786	NA	NA	NA	NA
rs1958604	rs10069494	EBF1	0.67	0.17	0.31
rs13153555	rs10069494	EBF1	0.6	0.34	0.34
rs4704938	rs10069494	EBF1	0.6	0.34	0.34
rs6597292	rs6597292	MUTED-TXNDC5	NA	NA	NA
rs6457791	rs9469899	UHRF1BP1	7.60E-35	2.30E-55	2.60E-120
rs13215181	rs9469899	UHRF1BP1	5.60E-34	5.30E-54	2.20E-112
rs6934662	rs9469899	UHRF1BP1	3.40E-30	3.60E-42	2.80E-92
rs7769961	rs9469899	UHRF1BP1	5.60E-34	1.30E-55	1.90E-119
rs16878812	rs4713895	FKBP5	0.57	0.12	0.67
rs9396266	rs9475818	PRIM2	0.36	0.074	0.31
6:57128340_C A_C	rs11752218	PRIM2	NA	NA	NA
rs9400476	rs9400480	TRAF3IP2-AS1	0.53	0.72	0.02
rs4945487	rs12528438	MARCKS	0.87	0.15	0.59
rs7770043	rs12528438	MARCKS	0.96	0.46	0.3
rs73534327	rs12528438	MARCKS	0.84	0.27	0.48
rs9388486	rs35510806	CENPW	0.000075	0.47	0.0007
rs6906478	rs142638448	EPB41L2	NA	NA	NA

rs13437326	rs142638448	EPB41L2	0.7	0.0003	0.00014
rs17080069	rs62434129	PLEKHG1	0.56	0.65	0.22
rs1619054	rs803463	MTHFD1L	0.00075	1.00E-13	0.00000011
rs1771817	rs803463	MTHFD1L	0.00016	2.90E-13	0.000000094
rs1620050	rs803463	MTHFD1L	0.00064	3.50E-14	0.000000075
rs1771816	rs803463	MTHFD1L	0.00064	3.50E-14	0.000000075
rs68144677	rs56263105	PARK2	0.31	0.85	0.0011
rs118099258	rs118099258	GNA12	0.68	0.0014	0.84
rs7807755	rs10951983	RAC1	0.018	0.86	0.013
rs2107595	rs2107595	HDAC9	0.43	0.73	0.19
rs2270223	rs2215614	TBX20	0.0014	0.0043	NA
rs17674947	rs2215614	TBX20	0.0014	0.0043	NA
rs11976145	rs2215614	TBX20	0.0037	0.0066	NA
7:35281692_A CT_A	rs2215614	TBX20	NA	NA	NA
rs73099190	rs2215614	TBX20	0.0005	0.0095	NA
rs1534310	rs35146811	CNPY4	0.81	0.97	0.054
rs764975725	rs35146811	CNPY4	NA	NA	NA
rs11765421	rs8137	LOC10012914 8	NA	NA	NA
rs10265	rs8137	LOC10012914 8	NA	NA	NA
rs6965212	rs8137	LOC10012914 8	NA	NA	NA
rs3924445	rs8137	LOC10012914 8	NA	NA	NA
rs38741	rs1089257	MRPS33	0.37	0.11	0.7
rs190692683	rs145928571	LOC10028701 5	NA	NA	NA
rs4233465	rs6668848	HEATR1	0.0011	0.00000014	1.60E-13
rs181657044	rs10094569	MTUS1	0.65	0.00043	0.32
rs73551707	rs56408342	BMP1	0.0023	3.30E-18	9.60E-16
rs73551705	rs56408342	BMP1	0.0038	4.60E-19	4.50E-14
rs112076731	rs896848	TP53INP1	0.71	0.000029	0.51
rs11782818	rs896848	TP53INP1	0.71	0.000029	0.51
rs11778553	rs896848	TP53INP1	0.71	0.000029	0.51
rs6982502	rs2001846	TRIB1	0.56	0.66	0.82
rs58914007	rs59413766	ST3GAL1	0.063	0.0097	0.00056
rs59413766	rs59413766	ST3GAL1	0.1	0.0065	0.0025
rs16904940	rs59413766	ST3GAL1	0.12	0.01	0.0021
rs72769357	rs61831074	KLF6	0.052	0.6	0.86
rs575142	rs10961206	C9orf146	NA	NA	NA
rs575856	rs10961206	C9orf146	NA	NA	NA
rs10964131	rs10811183	ACER2	0.99	0.64	0.24
rs1537373	rs2891168	CDKN2B-AS1	0.13	0.16	0.32
rs7861724	rs111344490	DCAF10	0.67	0.00081	0.00034

rs12555751	rs111344490	DCAF10	0.63	0.00011	0.00034
rs10814624	rs111344490	DCAF10	0.66	0.0005	0.00027
rs74987730	rs111344490	DCAF10	0.78	0.00013	0.000054
rs75986742	rs111344490	DCAF10	0.66	0.0005	0.00027
rs11394930	rs17566555	CDC123	0.79	0.0034	0.00089
rs11789185	rs11789185	ENG	0.23	0.079	0.17
rs3814492	rs10819473	IER5L	0.16	0.83	0.48
rs138415650	rs10819473	IER5L	0.44	0.76	0.53
rs10819474	rs10819473	IER5L	0.3	0.55	0.53
rs591673	rs653946	BAMBI	0.94	0.19	0.000012
rs3737180	rs220056	ZEB1	0.53	0.94	0.88
rs1576050	rs220056	ZEB1	0.61	0.54	0.68
rs66530629	rs6686889	SRRM1	0.69	0.0037	0.25
rs11000905	rs10824133	ADK	0.7	0.0000006	0.92
rs7070369	rs11202154	BMPR1A	0.0053	0.05	0.0046
rs61858562	rs11202154	BMPR1A	NA	NA	NA
rs1332327	rs1051338	LIPA	0.13	0.000000083	0.027
rs1332328	rs1051338	LIPA	0.12	0.000029	0.8
rs263533	rs1692580	SKI	0.91	0.29	0.64
rs2433027	rs11187752	PLCE1	NA	NA	NA
rs9665532	rs9665532	SLIT1	0.81	0.37	0.42
rs3824754	rs77787671	CNNM2	0.11	0.024	0.000021
rs79780963	rs77787671	CNNM2	0.045	0.046	0.000018
rs71485762	rs35007589	OBFC1	NA	NA	NA
rs34685262	rs35007589	OBFC1	NA	NA	NA
rs34102287	rs11146402	INPP5A	0.38	0.6	0.021
rs12762160	rs11146402	INPP5A	0.45	0.64	0.031
rs4627080	rs4537761	TMEM41B	0.12	0.53	0.34
rs173396	rs360153	SWAP70	0.13	0.0021	2.70E-11
rs2279286	11:13295751_AT_A	NA	NA	NA	NA
rs570129974	11:13295751_AT_A	NA	NA	NA	NA
rs12794768	rs4130469	SOX6	0.41	0.65	0.15
rs2593650	rs2034321	SLC17A6	NA	NA	NA
rs1337615	rs2467592	KCNA4	0.16	0.65	NA
rs2306363	rs2306363	SIPA1	0.1	0.4	0.0012
rs3016316	rs58221849	RIN1	0.05	0.64	0.72
rs11372961	rs59681006	PPFIA1	0.00036	4.50E-13	0.00061
rs7104718	rs59681006	PPFIA1	0.00019	2.50E-12	0.00031
rs2508619	rs2508619	KLHL35	0.000038	1.80E-26	6.30E-12
rs495889	rs618012	PAK1	0.00000036	7.30E-15	1.90E-27
rs4360494	rs61776719	FHL3	2.20E-11	2.00E-44	8.10E-65
rs604723	rs633185	ARHGAP42	6.10E-14	3.00E-28	5.50E-51

rs2019090	rs2839812	MIR4693	NA	NA	NA
rs10488763	rs10488763	FDX1	0.066	0.012	0.028
rs4252591	rs751537730	NA	NA	NA	NA
rs11216200	rs78044162	SIK3	0.18	0.23	0.078
rs11217133	rs1177562	VPS11	0.078	2.60E-13	3.60E-15
rs2509121	rs1177562	VPS11	0.043	3.20E-12	6.70E-15
rs1784459	rs1177562	VPS11	0.093	4.30E-13	8.10E-13
rs1786684	rs1177562	VPS11	0.093	4.30E-13	8.10E-13
rs1784460	rs1177562	VPS11	0.093	4.30E-13	8.10E-13
rs1786141	rs1177562	VPS11	0.093	4.30E-13	8.10E-13
rs1784461	rs1177562	VPS11	0.093	4.30E-13	8.10E-13
rs10790517	rs61377406	UBASH3B	0.086	0.058	0.65
rs10892870	rs61377406	UBASH3B	0.076	0.058	0.58
rs7127978	rs61377406	UBASH3B	0.24	0.065	0.68
rs7128198	rs61377406	UBASH3B	0.25	0.065	0.66
rs7129071	rs61377406	UBASH3B	0.25	0.065	0.66
rs10849546	12:7178440_T ATTTA_T	NA	NA	NA	NA
rs2275116	rs114786332	ZMYND12	0.98	0.88	0.38
rs12403025	rs11210939	SLC6A9	0.5	0.5	0.24
rs12403036	rs11210939	SLC6A9	0.5	0.55	0.23
rs147552389	rs75201095	CCNT1	0.79	0.5	0.51
rs2695789	12:56684496_T AGGAA_T	NA	NA	NA	NA
rs2371494	12:56684496_T AGGAA_T	NA	NA	NA	NA
12:56754371_T TA_T	12:56684496_T AGGAA_T	NA	N/A	N/A	N/A
rs555322	rs518634	FRS2	0.089	0.0061	0.00036
12:69921780_ GA_G	rs518634	FRS2	N/A	N/A	N/A
rs12817989	rs11107903	FGD6	0.014	8.20E-15	1.40E-11
rs4964274	rs968360	SSH1	0.053	0.044	0.84
rs2302700	rs968360	SSH1	0.076	0.042	0.8
rs2302701	rs968360	SSH1	0.076	0.042	0.8
rs11114051	rs968360	SSH1	0.19	0.041	0.73
rs35350651	rs10774625	ATXN2	0.51	0.56	0.3
rs7978610	rs7133378	DNAH10	0.2	0.0096	0.0000012
rs181299844	rs112958375	UBC	NA	NA	NA
rs11608885	rs111352470	ZNF268	1	0.7	0.64
rs17086701	rs663137	FLT1	0.18	0.94	0.68
rs9591145	rs7991314	N4BP2L2	0.5	0.000000071	0.00000069
rs10689124	rs734781	STARD13	0.33	0.21	0.15
rs9569666	rs734781	STARD13	0.26	0.23	0.13
rs9569667	rs734781	STARD13	0.26	0.2	0.12

rs12858634	rs734781	STARD13	0.22	0.36	0.24
rs80225305	rs8001982	NAA16	0.76	0.89	0.6
rs4942039	rs8001982	NAA16	0.75	0.8	0.7
rs58551993	rs8001982	NAA16	NA	NA	NA
rs11206803	rs11206803	PPAP2B	NA	NA	NA
rs117633128	rs11620127	DOCK9	0.6	0.0045	0.045
rs2391817	rs3783113	COL4A1	0.23	0.54	0.56
rs7333991	rs7333991	COL4A2	0.28	0.28	0.24
rs71445078	rs9521906	CARS2	NA	NA	NA
rs4981812	rs10686522	AP4S1	0.0082	0.31	0.018
rs179744	rs10686522	AP4S1	0.014	0.23	0.0039
rs175040	rs10131894	EIF2B2	0.000000063	6.30E-30	1.20E-25
rs175499	rs10131894	EIF2B2	0.000000032	1.60E-25	1.60E-20
rs3986303	rs3959554	EXD1	NA	NA	NA
rs11636934	rs12101466	LOXL1	0.78	0.73	0.0043
rs2507	rs12101466	LOXL1	0.78	0.71	0.0055
rs11072641	rs12148761	TSPAN3	0.00045	0.00000049	6.30E-20
rs12410336	rs11589090	RERE	0.66	0.99	0.94
rs76412132	rs7177201	ADAMTS7	0.043	0.051	0.00013
rs76681511	rs7177201	ADAMTS7	0.043	0.051	0.00013
rs75884254	rs7177201	ADAMTS7	0.043	0.051	0.00013
rs11852887	rs7177201	ADAMTS7	0.058	0.12	0.00035
rs4887091	rs7177201	ADAMTS7	0.057	0.017	0.00013
rs5029904	rs7173743	MORF4L1	0.56	1.30E-12	1.90E-16
rs1894401	rs7183988	FES	0.000011	0.000000081	4.20E-10
rs11121223	rs11589090	RERE	NA	NA	NA
rs9937801	rs9935770	DNAH3	NA	NA	NA
rs7186298	rs9935770	DNAH3	NA	NA	NA
rs564606546	rs564606546	NA	NA	NA	NA
rs4888392	rs8046696	CFDP1	0.035	0.1	0.27
rs4888409	rs8046696	CFDP1	0.042	0.13	0.23
rs7188231	rs8046696	CFDP1	0.042	0.16	0.094
rs3743609	rs8046696	CFDP1	0.042	0.15	0.078
rs3833048	rs8046696	CFDP1	0.037	0.13	0.15
rs34904236	rs8046696	CFDP1	0.57	0.093	0.67
rs7189106	rs7189462	PLCG2	0.047	0.022	0.52
rs9913156	rs8064504	ARRB2	0.1	0.73	0.33
rs4925125	rs12936927	SREBF1	0.12	0.000048	0.000000042
rs2470269	rs141077831	LRRC37B	0.82	0.54	0.2
rs17608766	rs17608766	GOSR2	0.79	0.29	0.95
rs2644383	rs62073965	NPEPPS	NA	NA	NA
rs62076549	rs62073965	NPEPPS	0.098	0.000011	0.000000017
rs4794283	rs62073965	NPEPPS	NA	NA	NA

rs55864353	rs62073965	NPEPPS	NA	NA	NA
rs1010322	rs5820757	ZNF652	0.001	0.00015	1.60E-09
19:11196356_AC_A	rs55997232	LDLR	N/A	N/A	N/A
rs57217136	rs55997232	LDLR	NA	NA	NA
rs4808046	rs3745318	KLF2	0.85	0.00043	0.0022
rs7246865	rs7246865	MYO9B	0.24	0.00097	8.80E-12
rs35976034	rs10410487	MAP1S	0.085	0.0067	0.0017
rs11670056	rs78030362	ELL	0.2	0.046	0.022
rs56255430	19:19432290_AG_A	NA	NA	NA	NA
rs2241719	rs11466359	TGFB1	0.8	0.37	0.38
rs2241718	rs11466359	TGFB1	0.8	0.31	0.31
rs12462166	rs1800469	B9D2	0.63	0.34	0.000011
rs2241712	rs1800469	B9D2	0.5	0.58	0.00015
rs35902819	rs8108474	RSPH6A	NA	NA	NA
rs7252126	rs8108474	RSPH6A	NA	NA	NA
rs60572996	rs62133081	HEATR5B	0.45	0.21	0.12
rs1549721	rs2023292	ZFP36L2	NA	NA	NA
rs2023292	rs2023292	ZFP36L2	NA	NA	NA
2:43552437_GT_G	rs111616286	THADA	N/A	N/A	N/A
rs582384	rs582384	PRKCE	0.85	0.8	0.43
rs12401351	rs56350134	NGF	0.59	0.33	0.45
rs12401352	rs56350134	NGF	0.59	0.33	0.45
rs6720415	rs10221683	AAK1	0.79	0.48	0.69
rs6546693	rs10207142	DYSF	0.59	0.35	0.02
rs753387787	rs12739316	LAMC1	NA	NA	NA
rs1316829	rs6686889	SRRM1	0.77	0.0031	0.26
rs5773100	rs6686889	SRRM1	0.75	0.0021	0.29
rs76341367	rs60808032	FOXC2	0.46	0.0093	0.000047
rs12952249	rs6505092	NEK8	0.77	0.27	0.74
rs6546510	rs10221683	AAK1	0.79	0.48	0.77
rs4315609	rs73354869	NCRNA00189	NA	NA	NA
rs4819290	rs35219138	RRP1B	0.022	0.82	0.11
rs2070533	rs35219138	RRP1B	0.045	0.66	0.085
rs12271690	rs528255324	NA	NA	NA	NA
rs12277907	rs528255324	NA	NA	NA	NA
rs12291203	rs528255324	NA	NA	NA	NA
rs11236430	rs2508619	KLHL35	0.000055	3.50E-25	2.60E-11
rs221072	rs518634	FRS2	0.11	0.0059	0.00048
rs7994229	rs7991314	N4BP2L2	0.71	8.10E-09	0.000000024
rs2275117	rs114786332	ZMYND12	0.98	0.88	0.38
rs7989823	rs7989823	COL4A2	0.61	0.33	0.15

rs57749886	rs1190344	RCOR1	0.6	0.67	0.069
rs533142015	rs533142015	NA	NA	NA	NA
rs5002487	rs12936927	SREBF1	0.17	0.000089	0.00000027
rs4794019	rs4643373	IGF2BP1	0.67	NA	0.24
rs4968721	rs11079536	PECAM1	0.81	0.56	0.0018
rs61459202	rs11466359	TGFB1	0.63	0.39	0.15
rs7254679	rs11466359	TGFB1	0.88	0.38	0.24
rs12983775	rs1800469	B9D2	0.63	0.38	0.0000095
rs10692845	rs515135	APOB	0.026	0.63	0.77
rs35507976	rs6736093	MERTK	0.65	0.017	0.00000004
rs1250259	rs1250247	FN1	0.15	0.4	0.2
rs11676780	rs74431246	LRRFIP1	0.23	0.03	0.73
rs762737767	rs6088595	NCOA6	NA	NA	NA
rs16987150	rs59827666	TGM2	0.041	0.0037	0.11
rs17196927	rs59827666	TGM2	0.11	0.007	0.31
rs6123388	rs59827666	TGM2	0.025	0.00094	0.13
rs17196913	rs59827666	TGM2	0.098	0.0064	0.34
rs80346118	rs2008614	PREX1	0.63	0.23	0.068
rs75587772	rs2008614	PREX1	0.57	0.14	0.028
rs73260214	rs2008614	PREX1	0.63	0.23	0.068
rs2051407	rs35219138	RRP1B	0.016	0.93	0.079
rs11204859	rs12035531	TUFT1	0.17	0.55	0.25
rs55644935	rs17002947	LOC388906	NA	NA	NA
rs6442109	rs7429594	CDC25A	0.000000018	5.80E-27	3.40E-31
rs6442108	rs7429594	CDC25A	0.000000018	5.80E-27	3.40E-31
rs6762186	rs1483891	FRMD4B	0.00086	0.000015	0.0052
rs13089164	rs9790198	GXYLT2	0.0054	0.47	2.50E-10
rs1727887	rs357494	ARHGEF26	0.000069	2.10E-12	9.30E-23
rs12493885	rs357494	ARHGEF26	0.000069	2.10E-12	9.30E-23
rs11248061	rs1377586	DGKQ	8.00E-14	1.70E-30	9.60E-42
rs10902762	rs1377586	DGKQ	1.70E-13	1.10E-29	1.70E-41
rs74768960	rs74768960	C5orf13	NA	NA	NA
rs335428	rs753144400	NA	NA	NA	NA
rs234376	rs753144400	NA	NA	NA	NA
rs13193957	rs12528438	MARCKS	0.92	0.46	0.28
rs3035619	rs12528438	MARCKS	0.92	0.46	0.28
rs2473937	rs12528438	MARCKS	0.96	0.46	0.3
rs2502400	rs12528438	MARCKS	0.82	0.26	0.68
rs17080102	rs62434129	PLEKHG1	0.43	0.71	0.25
rs7773781	rs6557122	AKAP12	0.84	0.17	0.58
rs6965143	rs8137	LOC100129148	NA	NA	NA
rs6947215	rs8137	LOC100129148	NA	NA	NA

rs6962366	rs8137	LOC100129148	NA	NA	NA
rs7781964	rs756142636	NA	NA	NA	NA
rs111862082	rs111862082	MKRN1	0.57	0.19	0.72
rs188227226	rs10094569	MTUS1	0.65	0.00043	0.53
rs77055525	rs896848	TP53INP1	0.71	0.000029	0.51
rs10087329	rs59413766	ST3GAL1	0.063	0.0097	0.00056
rs62520290	rs59413766	ST3GAL1	0.087	0.0064	0.0018
rs73447281	rs111344490	DCAF10	0.66	0.0005	0.00027
rs9337951	rs9337951	KIAA1462	0.49	1.80E-25	0.0000036
rs3737178	rs220056	ZEB1	0.53	0.94	0.96
rs3737179	rs220056	ZEB1	0.53	0.94	0.88
rs11000897	rs10824133	ADK	0.71	0.00000093	0.94
rs10824083	rs10824133	ADK	0.9	0.0000042	0.68
rs10824084	rs10824133	ADK	0.9	0.0000042	0.68
rs4934258	rs11202154	BMPR1A	0.04	0.15	0.062
rs34755052	rs11202154	BMPR1A	0.0042	0.084	0.0059
rs61871680	rs61871680	BTBD16	0.00066	0.000000001	2.30E-10
rs28558789	rs1130678	CD151	0.000000014	5.90E-14	2.70E-49
rs1002707	rs6686889	SRRM1	0.64	0.0038	0.27
rs56348932	rs56170783	PPAP2B	NA	NA	NA
rs4888403	rs8046696	CFDP1	0.042	0.13	0.23
rs113985803	rs2365468	BPTF	NA	NA	NA
rs71378928	rs2909217	WIPI1	0.67	0.39	0.32
rs470389	rs470802	ZNF236	0.39	0.15	0.014
rs470469	rs470802	ZNF236	0.35	0.12	0.0095
rs7769820	rs9469899	UHRF1BP1	5.60E-34	1.30E-55	1.90E-119
rs396207	6:164482304_ATCT_A	NA	NA	NA	NA
rs7793613	rs12112877	COG5	0.75	0.015	0.81
rs11192991	10:108405474_ATGTTTATTT_A	NA	NA	NA	NA
rs11192992	10:108405474_ATGTTTATTT_A	NA	NA	NA	NA
rs3901289	rs11200519	BTBD16	3.00E-17	1.00E-47	1.30E-43
rs28372895	rs1130678	CD151	0.00000011	1.10E-13	7.80E-49
rs221078	rs518634	FRS2	0.089	0.0061	0.00031
rs112389360	rs8001982	NAA16	0.75	0.8	0.7
rs216152	rs12691049	MYH11	0.5	0.11	0.29
rs8077485	rs62073965	NPEPPS	0.14	0.0000016	0.00000018
rs2215417	rs175382	UTP18	0.87	0.39	0.003
rs74847504	rs10221683	AAK1	0.74	0.41	0.74
rs6750832	rs10176176	LOC100630918	NA	NA	NA

rs771405767	rs2008614	PREX1	NA	NA	NA
rs9607871	rs17002947	LOC388906	NA	NA	NA
rs2037119	rs59971314	ATXN7	0.35	0.44	0.66
rs113075206	rs1377586	DGKQ	3.70E-17	1.90E-44	5.10E-58
rs4001207	rs2127821	LOC645513	NA	NA	NA
rs6837828	rs2127821	LOC645513	NA	NA	NA
rs10476052	rs11746837	ADAM19	0.055	0.0012	0.00000001
6:160427471_CG_C	6:160427471_CG_C	NA	N/A	N/A	N/A
rs75672964	rs746014	PODXL	0.66	0.1	0.97
rs2564741	rs6668848	HEATR1	0.00085	0.00000023	5.10E-13
rs10818578	rs885150	DAB2IP	0.33	0.7	0.69
rs7911904	rs10764741	PTPRE	0.02	0.05	0.094
rs11537930	rs58221849	RIN1	0.037	0.79	0.42
rs113722375	rs59681006	PPFIA1	0.00019	2.50E-12	0.00031
rs7932431	rs59681006	PPFIA1	0.00011	4.40E-13	0.00003
rs7950280	rs59681006	PPFIA1	0.00038	2.70E-15	0.00000038
rs534812	rs2508619	KLHL35	0.000022	2.60E-25	1.60E-11
rs4396302	rs4245078	ETS1	0.17	0.96	0.3
rs12823740	rs7133378	DNAH10	0.0015	0.000013	2.60E-18
rs7328306	rs663137	FLT1	0.17	0.91	0.73
rs9569668	rs734781	STARD13	0.26	0.23	0.13
rs35776301	rs10686522	AP4S1	NA	NA	NA
rs17293632	rs56062135	SMAD3	0.013	0.0055	0.011
rs10625725	rs7177201	ADAMTS7	0.077	0.00054	0.000024
rs9936936	rs9936936	CMIP	0.35	0.54	0.9
rs2290771	rs2410859	UNC13D	0.32	0.86	0.03
rs11650046	rs57453217	CYTH1	0.11	0.47	0.81
rs12150986	rs12150986	37316	NA	NA	NA
rs2070737	rs8108474	RSPH6A	NA	NA	NA
rs35675382	rs2052745	GEN1	0.00000033	5.60E-12	8.40E-22
rs13031876	rs2052745	GEN1	0.0000002	7.10E-12	1.30E-21
rs6546521	rs10221683	AAK1	0.79	0.33	0.9
rs7421891	rs11123825	LONRF2	0.81	0.084	0.47
rs4851285	rs11123825	LONRF2	0.8	0.16	0.091
rs7604403	rs6736093	MERTK	0.65	0.017	0.00000004
rs12105229	rs10804106	AOX2P	0.77	NA	0.1
rs79996022	rs74431246	LRRFIP1	0.69	0.035	0.76
rs113119118	rs112875281	CLDN5	0.32	0.26	0.51
rs112875281	rs112875281	CLDN5	0.45	0.16	0.59
rs1781416	rs12035531	TUFT1	0.18	0.58	0.17
rs9606203	rs71313931	ARVCF	0.000000012	3.40E-44	5.70E-21
rs176157	rs12484557	CABIN1	0.051	0.36	0.86

3:134205445_ACCTT_A	rs9876658	ANAPC13	N/A	N/A	N/A
rs12497267	rs357494	ARHGEF26	0.000069	2.10E-12	9.30E-23
rs4652305	rs10913568	C1orf220	0.00063	0.0018	0.017
rs4652306	rs10913568	C1orf220	0.00063	0.0018	0.015
rs4652307	rs10913568	C1orf220	0.00063	0.0018	0.017
rs113983715	rs10512704	IRX1	0.0047	0.28	1.60E-12
rs62355881	5:56121794_CT_C	NA	NA	NA	NA
rs72758040	5:56121794_CT_C	NA	NA	NA	NA
rs62380877	rs62380939	KDM3B	0.19	0.32	0.018
rs56316522	rs62380939	KDM3B	0.19	0.3	0.012
rs251023	rs6881581	DIAPH1	0.12	0.013	0.16
rs335429	rs753144400	NA	NA	NA	NA
rs79754943	rs79754943	PLG	NA	NA	NA
rs4724806	rs10951983	RAC1	0.018	0.9	0.011
rs3828944	rs10951983	RAC1	0.018	0.9	0.0073
rs1060544	rs35146811	CNPY4	0.91	0.83	0.042
rs2116095	rs10100333	SOX7	0.33	0.6	0.033
rs2281721	rs4846913	GALNT2	0.93	0.26	0.47
rs7848132	rs111344490	DCAF10	0.54	0.0012	0.00011
rs56045355	rs10981435	KIAA1958	0.94	0.44	0.24
rs10818579	rs885150	DAB2IP	0.33	0.7	0.69
rs35007589	rs35007589	OBFC1	NA	NA	NA
rs7091055	rs7091055	CHST15	0.67	0.14	0.000000042
rs55905028	rs59681006	PPFIA1	0.00011	4.40E-13	0.00003
rs192240803	rs75201095	CCNT1	0.4	0.19	0.8
rs185029099	rs75201095	CCNT1	0.82	0.46	0.52
rs184368932	rs75201095	CCNT1	0.79	0.5	0.51
rs56082942	rs75201095	CCNT1	0.67	0.077	0.97
rs150374862	rs75201095	CCNT1	0.79	0.5	0.51
rs2483099	rs8001982	NAA16	0.95	0.8	0.67
rs2869868	rs7177201	ADAMTS7	0.062	0.0016	0.000017
rs2002854	rs7177201	ADAMTS7	0.023	0.0053	0.000034
rs11591147	rs11591147	PCSK9	0.85	0.34	0.69
rs17114036	rs56170783	PPAP2B	NA	NA	NA
rs72664324	rs56170783	PPAP2B	NA	NA	NA
rs17881561	rs62049432	LOC100506083	NA	NA	NA
rs35251820	rs8046696	CFDP1	0.055	0.09	0.14
rs116843064	rs116843064	ANGPTL4	0.94	0.45	0.59
rs7256001	rs7255739	ZNF568	0.019	0.000000038	0.000069
rs117955557	rs183657985	EXOC3L2	0.94	0.69	0.12
rs6741336	rs10207142	DYSF	0.58	0.25	0.044

rs6750847	rs10176176	LOC100630918	NA	NA	NA
rs7512500	rs61797068	BAI3	NA	NA	NA
rs7512502	rs61797068	BAI3	NA	NA	NA
rs10804106	rs10804106	AOX2P	0.38	NA	0.11
rs2345383	rs59827666	TGM2	0.18	0.0064	0.055
rs6000285	rs6000285	MED15	0.93	0.42	0.33
rs5755961	22:36249338_TA_T	NA	NA	NA	NA
rs56386062	rs2127821	LOC645513	NA	NA	NA
rs112379863	rs62380939	KDM3B	0.19	0.31	0.021
rs150198102	rs62380939	KDM3B	0.19	0.36	0.015
rs769204262	rs6881581	DIAPH1	NA	NA	NA
rs631556	rs511763	NAV1	0.025	0.2	0.01
rs667340	rs577449462	NA	NA	NA	NA
rs4895389	rs2327426	TCF21	0.0000022	0.000017	0.0000048
rs6943607	rs1089257	MRPS33	0.14	0.16	0.84
rs6997760	rs1510758	DOCK5	0.44	0.15	0.00009
rs77493007	rs28583665	ZFPM2	NA	NA	NA
rs76473630	rs28583665	ZFPM2	NA	NA	NA
rs11249174	rs6686889	SRRM1	0.31	0.014	0.53
rs12121345	rs6686889	SRRM1	0.16	0.003	0.49
rs7927208	rs61377406	UBASH3B	0.12	0.049	0.6
rs41276688	rs112958375	UBC	0.19	0.12	0.02
rs9569651	rs734781	STARD13	0.33	0.21	0.2
rs1057881	rs1190344	RCOR1	NA	NA	NA
rs139704890	rs114786332	ZMYND12	0.98	0.88	0.43
rs59409959	rs8062169	BCAR1	0.46	0.00068	0.034
rs13306736	rs12936927	SREBF1	0.051	0.00042	0.00001
rs62086046	rs2365468	BPTF	0.00051	0.0002	0.018
rs183657985	rs183657985	EXOC3L2	NA	NA	NA
rs36091860	rs8108474	RSPH6A	NA	NA	NA
rs1123648	rs62133081	HEATR5B	0.45	0.21	0.12
rs6442107	rs7429594	CDC25A	0.000000017	2.20E-26	2.10E-33
rs10911228	rs12739316	LAMC1	NA	NA	NA
rs9689516	rs9469899	UHRF1BP1	5.60E-34	1.30E-55	1.90E-119
rs78785901	rs77612673	SASH1	0.49	0.44	0.56

Table 35. Further prioritisation of functional CAD SNPs

FGWAS region	FGWAS rsID	PPA*	CAD rsID	Nearest Genes
34	rs61776719	0.49		FHL3
	rs34655914	0.1007		
	rs35267671	0.0986		

	rs67631072	0.0903		
	rs4360494	0.0719	rs4360494	FHL3
	rs28470722	0.0130		
	rs17465420	0.0116		
	rs28493395	0.0088		
	rs28391281	0.0085		
	rs11485595	0.0081		
	rs28435150	0.0081		
	rs4072980	0.0077		
	rs28683126	0.0076		
	rs28570969	0.0074		
	rs28605759	0.0070		
	rs201741844	0.0070		
	rs28645205	0.0069		
47	rs11591147	1.00	rs11591147	PCSK9
48	rs56170783	0.1786		
	rs17114046	0.1632		
	rs9970807	0.1569		
	rs72664332	0.1563		
	rs72664355	0.0880		
	rs2404715	0.0639		
	rs72664341	0.0555		
	rs72664354	0.0406		
	rs72664335	0.0245		
	rs56348932	0.0187	rs56348932	PPAP2B
	rs72664353	0.0161		
231	rs9337951	1.00	rs9337951	KIAA1462
284	rs1412444	0.3757		
	rs1412445	0.2363		
	rs1332327	0.1197	rs1332327	LIPA
	rs1051338	0.1083		
	rs2246941	0.0510		
	rs1332328	0.0471	rs1332328	LIPA
	rs2246942	0.0207		
314	rs2672592	0.70		
	rs4237541	0.0534		
	rs4328161	0.0420		
	rs4547034	0.0362		
	rs77494534	0.0177		
	rs7907130	0.0135		
	rs911775	0.0116		
	rs7072288	0.0113		
	rs4752700	0.0100		
	rs7086403	0.0092		
	rs10788274	0.0092		

	rs7090214	0.0088		
	rs4752678	0.0084		
	rs7076349	0.0074		
	rs7092500	0.0073		
	rs7076102	0.0059		
	rs61871680	0.0041	rs61871680	BTBD16
340	rs11316597	0.1544		
	rs1351525	0.0931		
	rs4603287	0.0803		
	rs1481892	0.0741		
	rs60521023	0.0269		
	rs4757140	0.0262		
	rs10832013	0.0226		
	rs12361893	0.0204		
	rs7926712	0.0203		
	rs3993105	0.0201		
	rs4757139	0.0192		
	rs2279285	0.0187		
	rs4146386	0.0186		
	rs7928655	0.0185		
	rs2279286	0.0182	rs2279286	NA
	rs2279287	0.0181		
	rs4757138	0.0180		
	rs1384030	0.0178		
	rs7109016	0.0178		
	rs11022735	0.0177		
	rs900147	0.0166		
	rs900146	0.0163		
	rs1384024	0.0155		
	rs4414197	0.0151		
	rs2403662	0.0151		
	rs1384027	0.0141		
	rs10832014	0.0137		
	rs1384028	0.0136		
	rs1384025	0.0135		
	rs1384026	0.0134		
	rs900145	0.0125		
	rs34148132	0.0122		
	rs61882122	0.0108		
	rs11022752	0.0096		
	rs11022738	0.0088		
	rs11022751	0.0077		
	rs1384029	0.0063		
	rs36170258	0.0055		
	rs4757137	0.0053		

	rs12295734	0.0052		
387	rs2306363	0.91	rs2306363	SIPA1
	rs12801636	0.0383		
	rs66864335	0.0229		
429	rs10488763	0.51	rs10488763	FDX1
	rs1443120	0.4770		
459	rs72447384	0.4771		
	rs11838267	0.0483		
	rs11064498	0.0452		
	rs12146727	0.0450		
	rs10849546	0.0445	rs10849546	NA
	rs111453032	0.0442		
	rs16933084	0.0432		
	rs77620124	0.0422		
	rs12368181	0.0346		
	rs12366520	0.0305		
	rs7962629	0.0292		
	rs112796495	0.0272		
	rs12368783	0.0217		
	rs12367198	0.0154		
	rs7183	0.0146		
540	rs12817989	0.2899	rs12817989	FGD6
	rs11107909	0.1877		
	rs12372152	0.1423		
	rs11107908	0.1418		
	rs11107903	0.0840		
	rs11107900	0.0537		
	rs12830164	0.0306		
	rs11107904	0.0272		
556	rs10774625	0.78		
	rs4766578	0.0898		
	rs35350651	0.0475	rs35350651	ATXN2
	rs3184504	0.0448		
721	rs10131894	0.54		
	rs175040	0.1135	rs175040	EIF2B2
	rs91144	0.0867		
	rs175047	0.0765		
	rs12882664	0.0460		
	rs175036	0.0164		
	rs175016	0.0144		
	rs175037	0.0087		
	rs12894419	0.0067		
	rs175007	0.0060		
	rs175057	0.0047		
	rs175499	0.0044	rs175499	EIF2B2

	rs ID	Value	rs ID	Gene
	rs8013780	0.0042		
	rs12588981	0.0041		
	rs175080	0.0039		
	rs1548807	0.0032		
	rs12882111	0.0032		
	rs8017642	0.0024		
	rs175042	0.0023		
	rs4556	0.0021		
	rs175012	0.0021		
	rs4903284	0.0018		
792	rs72743461	0.50		
	rs17293632	0.3743	rs17293632	SMAD3
	rs56062135	0.0665		
	rs17228058	0.0337		
812	rs7183988	0.3751		
	rs7177338	0.3404		
	rs11539637	0.1627		
	rs6224	0.0583		
	rs1894401	0.0343	rs1894401	FES
890	rs7189462	0.1127		
	rs7200626	0.0852		
	rs9924101	0.0841		
	rs7189106	0.0839	rs7189106	PLCG2
	rs10493891	0.0501		
	rs7200324	0.0429		
	rs12716921	0.0422		
	rs7198813	0.0414		
	rs7199578	0.0402		
	rs10871423	0.0400		
	rs7199941	0.0396		
	rs12102309	0.0367		
	rs3934985	0.0299		
	rs7197689	0.0268		
	rs7197685	0.0219		
	rs7197832	0.0203		
	rs12928482	0.0201		
	rs7342694	0.0157		
	rs12716922	0.0139		
	rs4997772	0.0137		
	rs12931627	0.0120		
	rs56215311	0.0092		
	rs4997770	0.0091		
	rs4997769	0.0088		
	rs55873405	0.0068		
	rs12918486	0.0062		

	rs7192509	0.0060		
	rs4997771	0.0060		
	rs7192535	0.0059		
	rs7187163	0.0052		
	rs7187837	0.0051		
	rs55822905	0.0051		
	rs4997773	0.0050		
895	rs299946	0.3703		
	rs423984	0.1104		
	rs60753168	0.0574		
	rs62042066	0.0551		
	rs299945	0.0335		
	rs61403071	0.0296		
	rs12926098	0.0272		
	rs12444314	0.0272		
	rs34076489	0.0255		
	rs526882	0.0207		
	rs12446277	0.0180		
	rs527031	0.0162		
	rs8051014	0.0139		
	rs8064031	0.0122		
	rs4843171	0.0115		
	rs12444333	0.0108		
	rs78285663	0.0084		
	rs79430710	0.0078		
	rs380539	0.0071		
	rs77563985	0.0071		
	rs60808032	0.0067		
	rs414347	0.0064		
	rs79088793	0.0059		
	rs78689656	0.0053		
	rs4843167	0.0052		
	rs117700120	0.0050		
	rs12922111	0.0042		
	rs74870060	0.0042		
	rs735344	0.0039		
	rs118004807	0.0038		
	rs12918331	0.0035		
	rs7206713	0.0031		
	rs79667308	0.0030		
	rs4843173	0.0025		
	rs299937	0.0018		
	rs11863161	0.0018		
	rs62040912	0.0016		
	rs899245	0.0013		

	rs11642652	0.0011		
	rs4843160	0.0009		
	rs8049102	0.0009		
	rs74387953	0.0008		
	rs434772	0.0007		
	rs4843453	0.0006		
	rs80284611	0.0006		
	rs76341367	0.0005	rs76341367	FOXC2
	rs9938127	0.0005		
	rs35670068	0.0005		
	rs450260	0.0004		
	rs12919475	0.0004		
	rs28460639	0.0004		
	16:86603991_T_T TA	0.0004		
	rs76062880	0.0003		
	rs7206530	0.0003		
	rs13338127	0.0003		
	rs1088953	0.0003		
	rs4843438	0.0003		
	rs4843437	0.0003		
	rs9930902	0.0003		
	rs79805292	0.0003		
953	rs11079536	0.4017		
	rs1122800	0.1752		
	rs1867624	0.0778		
	rs4968720	0.0675		
	rs4968721	0.0387	rs4968721	PECAM1
	rs1962147	0.0301		
	rs2070783	0.0229		
	rs9303469	0.0201		
	rs9303470	0.0195		
	rs8070460	0.0163		
	rs1108592	0.0155		
	rs12936766	0.0144		
	rs2070784	0.0134		
	rs6504218	0.0134		
	17:62409725_C_C TT	0.0119		
	rs6504213	0.0112		
	rs9902260	0.0069		
957	rs2909218	0.3198		
	rs2952289	0.3161		
	rs2952286	0.0726		
	rs2361953	0.0485		

	rs2909220	0.0452		
	rs2909217	0.0427		
	rs71378928	0.0403	rs71378928	WIPI1
	rs2909221	0.0338		
	rs2111508	0.0298		
	rs2909211	0.0113		
967	rs8066695	0.51		
	rs8075861	0.3459		
	rs57453217	0.0262		
	rs11650046	0.0081	rs11650046	CYTH1
	rs7502213	0.0059		
	rs12942140	0.0052		
	rs12941476	0.0046		
	rs8080283	0.0038		
	rs7405712	0.0037		
	rs12943621	0.0037		
	rs8065768	0.0035		
	rs1531798	0.0035		
	17:76811791_G_G_A	0.0027		
	rs1531797	0.0020		
	rs7212662	0.0020		
	rs11649885	0.0017		
	rs2306527	0.0016		
	rs62075567	0.0016		
	rs12604076	0.0015		
	rs6501270	0.0014		
	rs7224711	0.0014		
	rs17657522	0.0012		
	rs2306526	0.0012		
	rs8076588	0.0011		
	rs62075585	0.0011		
	rs6501250	0.0010		
	rs9916809	0.0010		
	rs17657767	0.0010		
	rs11658134	0.0009		
	rs11656741	0.0009		
	rs1110274	0.0009		
	rs6501269	0.0009		
	rs11077397	0.0009		
	rs11656742	0.0009		
	rs17736494	0.0009		
1055	rs116843064	1.00	rs116843064	ANGPTL4
1064	rs7246865	0.71	rs7246865	MYO9B
	rs10421727	0.1089		

	rs56937714	0.0220		
	rs1060367	0.0212		
	rs3745348	0.0212		
	rs55957788	0.0173		
	rs10427132	0.0162		
	rs10401600	0.0159		
	rs8101673	0.0151		
1065	rs10410487	0.4125		
	rs35976034	0.1817	rs35976034	MAP1S
	rs10423961	0.1472		
	rs28689805	0.0722		
	rs55752165	0.0290		
	rs12979186	0.0229		
	rs13382133	0.0184		
	rs28398825	0.0146		
	rs2382631	0.0130		
	rs28719946	0.0122		
	rs28403253	0.0117		
	rs10422213	0.0076		
	rs28736584	0.0076		
1066	rs78030362	0.62		
	rs11670056	0.3752	rs11670056	ELL
1091	rs8108474	0.2342		
	rs8108864	0.1248		
	rs8108590	0.1243		
	rs34633566	0.0883		
	rs17571725	0.0639		
	rs7252126	0.0603	rs7252126	RSPH6A
	rs9676288	0.0533		
	rs2070737	0.0520	rs2070737	RSPH6A
	rs8109951	0.0500		
	rs2341097	0.0413		
	rs725660	0.0210		
	rs2879934	0.0188		
	rs1972423	0.0110		
	rs36091860	0.0103	rs36091860	RSPH6A
1126	rs16986953	1.00		
1149	rs4245791	0.3889		
	rs6544713	0.3260		
	rs4299376	0.2850		
1151	rs582384	0.57	rs582384	PRKCE
	rs4369911	0.1971		
	rs616381	0.0438		
	rs2344662	0.0403		
	rs6716707	0.0352		

	rs632357	0.0262		
	rs610115	0.0224		
	rs546639	0.0096		
	rs474453	0.0057		
1219	rs114192718	0.56	rs114192718	SAP130
	rs77836623	0.4126		
1293	rs1250247	0.3802		
	rs1250240	0.1971		
	rs1250259	0.0966	rs1250259	FN1
	rs1250239	0.0871		
	rs1250242	0.0733		
	rs1250258	0.0659		
	rs1250244	0.0457		
	rs1250241	0.0335		
1359	rs2008614	0.2555		
	rs4809731	0.0804		
	rs2004772	0.0734		
	rs6095302	0.0521		
	rs113216630	0.0479		
	rs17366885	0.0459		
	rs6095299	0.0403		
	rs56313611	0.0373		
	rs2093348	0.0358		
	rs4809729	0.0336		
	rs2869677	0.0310		
	rs77120386	0.0276		
	rs75587772	0.0228	rs75587772	PREX1
	rs754064	0.0204		
	rs2008603	0.0198		
	rs73260214	0.0142	rs73260214	PREX1
	rs73609701	0.0130		
	rs79044887	0.0105		
	rs73609690	0.0100		
	rs80346118	0.0094	rs80346118	PREX1
	rs6090908	0.0093		
	rs17366644	0.0092		
	rs78473917	0.0091		
	rs73609692	0.0088		
	rs17366684	0.0087		
	rs73609697	0.0085		
	rs6095326	0.0079		
	rs73609696	0.0077		
1415	rs71313931	0.3776		
	rs71313932	0.3243		
	rs756653	0.0488		

	rs2238787	0.0400		
	rs1034564	0.0364		
	rs1034566	0.0243		
	rs2073743	0.0210		
	rs9606204	0.0185		
	rs2238786	0.0167		
	rs4819852	0.0133		
	rs7285377	0.0114		
	rs9606203	0.0114	rs9606203	ARVCF
	rs2238792	0.0091		
1685	rs781663	0.3126		
	rs7687767	0.1065		
	rs56155140	0.0848	rs56155140	REST
	rs781657	0.0773		
	rs6853156	0.0753	rs6853156	REST
	rs6554401	0.0558		
	rs781656	0.0501		
	rs66790703	0.0501	rs66790703	REST
	rs72606404	0.0262		
	rs17081935	0.0183		
	rs781658	0.0156		
	rs17081933	0.0125		
	rs6852997	0.0095		
	rs72627509	0.0085		
	rs12645070	0.0079		
	rs7658601	0.0073		
	rs57265257	0.0071		
	rs58408429	0.0071		
	rs17087335	0.0067		
	rs72627508	0.0065		
	rs3796529	0.0064		
1790	rs869396	0.69		
	rs7696431	0.1720		
	rs6823685	0.0316	rs6823685	PALLD
	rs869397	0.0262		
	rs869398	0.0231		
	rs6854026	0.0192	rs6854026	PALLD
1910	rs112949822	0.1240		
	rs113792119	0.1007		
	rs76325227	0.0962		
	rs79001324	0.0875		
	rs79243757	0.0874		
	rs77630811	0.0825		
	rs112472930	0.0804		
	rs113728457	0.0625	rs113728457	FER

	rs79536589	0.0547	rs79536589	FER
	rs111465548	0.0526		
	rs78633399	0.0507		
	rs77844865	0.0125		
	rs71592765	0.0112		
	rs79980741	0.0096		
	rs111636466	0.0095		
	rs75271047	0.0090		
	rs112828293	0.0083		
	rs79621178	0.0077		
	rs77606373	0.0076		
2150	rs4724806	0.3246	rs4724806	RAC1
	rs3828944	0.2961	rs3828944	RAC1
	rs10951983	0.1137		
	rs7807755	0.0993	rs7807755	RAC1
	rs7797644	0.0529		
	rs33986000	0.0090		
	rs12538142	0.0081		
	rs4724802	0.0064		
	rs836545	0.0055		
	rs836546	0.0046		
	rs12539027	0.0045		
	rs34420518	0.0044		
	rs113080138	0.0044		
	rs35444732	0.0030		
	rs1880118	0.0026		
	rs12537894	0.0026		
	rs11760556	0.0025		
	rs12532493	0.0025		
	rs11767078	0.0025		
	rs12537483	0.0025		
2164	rs2107595	0.98	rs2107595	HDAC9
2179	rs4723406	0.0491		
	rs17675131	0.0421		
	rs73099190	0.0294	rs73099190	TBX20
	rs28561590	0.0277		
	rs10235849	0.0246		
	rs73099185	0.0201		
	rs17606844	0.0198		
	rs73099187	0.0196		
	rs28628585	0.0193		
	rs4720170	0.0187		
	rs17675142	0.0184		
	rs12538515	0.0179		

rs714956	0.0178		
rs17675148	0.0177		
rs714958	0.0173		
rs7797172	0.0172		
rs10249005	0.0164		
rs4723407	0.0164		
rs17606698	0.0159		
rs715112	0.0158		
rs2270223	0.0157	rs2270223	TBX20
rs6972725	0.0153		
rs10280486	0.0152		
rs4723405	0.0152		
rs1346021	0.0152		
rs6976630	0.0152		
rs10263165	0.0151		
rs10255047	0.0151		
rs73099148	0.0150		
rs6977095	0.0149		
rs73099202	0.0148		
rs17606602	0.0147		
rs6963934	0.0147		
rs6959445	0.0144		
rs17674947	0.0144	rs17674947	TBX20
rs10224889	0.0137		
rs6976783	0.0125		
rs6959846	0.0124		
rs7790964	0.0122		
rs2215614	0.0109		
rs10264933	0.0106		
rs12538541	0.0105		
rs2024343	0.0098		
rs10264435	0.0098		
rs2024345	0.0098		
rs6959920	0.0096		
rs2024344	0.0094		
rs6946632	0.0092		
rs10280159	0.0091		
rs988270	0.0090		
rs7795344	0.0083		
rs2109090	0.0083		
rs2075125	0.0079		
rs2024346	0.0075		
rs10486660	0.0073		
rs1362207	0.0071		
rs7790713	0.0069		

	rs7790990	0.0069		
	rs6462599	0.0068		
	rs7790968	0.0065		
	rs2191884	0.0062		
	rs2191885	0.0059		
	rs56834372	0.0057		
	rs57812663	0.0055		
	rs17674881	0.0051		
	rs6959887	0.0050		
	rs10254475	0.0050		
	rs17606656	0.0050		
	rs17606554	0.0050		
	rs1833126	0.0049		
2270	rs756142636	0.67		
	rs2001621	0.0446		
	rs10435198	0.0364		
	rs10435199	0.0253		
	rs7781964	0.0175	rs7781964	NA
	rs111324112	0.0131		
	7:139709728_C_CCA	0.0101		
	rs2286196	0.0092		
	rs2286195	0.0079		
	rs111862082	0.0076	rs111862082	MKRN1
	7:139714881_A_AAAAC	0.0042		
	rs6954547	0.0038		
	rs7795996	0.0036		
	rs3823715	0.0036		
	rs35586793	0.0032		
	rs2284213	0.0028		
	rs6945612	0.0027		
	rs9640375	0.0027		
	rs6969159	0.0024		
	rs2286198	0.0023		
	rs10229358	0.0021		
	rs2269997	0.0019		
	rs3923186	0.0018		
	rs7788777	0.0017		
	rs3922848	0.0017		
	rs757835	0.0016		
	rs757834	0.0016		
	rs6949963	0.0016		
	rs6950143	0.0016		
	rs117102789	0.0015		

rs10229249	0.0015		
rs10274480	0.0015		
rs6954531	0.0014		
rs10265	0.0013	rs10265	LOC100129148
rs6962053	0.0013		
rs4398843	0.0013		
rs3924445	0.0013	rs3924445	LOC100129148
rs11765421	0.0013	rs11765421	LOC100129148
rs6961812	0.0012		
rs4379397	0.0012		
rs4398844	0.0012		
rs11768023	0.0011		
rs6467838	0.0011		
rs11771715	0.0011		
rs6962366	0.0011	rs6962366	LOC100129148
7:139759161_T_T AC	0.0011		
rs6965143	0.0010	rs6965143	LOC100129148
rs2269996	0.0009		
rs7791455	0.0009		
rs7789470	0.0009		
rs6947215	0.0009	rs6947215	LOC100129148
rs73163251	0.0007		
rs2286200	0.0007		
rs2005751	0.0007		
rs11772026	0.0007		
rs10237377	0.0007		
rs781269942	0.0006		
rs6952262	0.0006		
7:139046517_T_T GA	0.0006		
rs10952541	0.0006		
rs1544460	0.0005		
rs753309490	0.0005		
rs6945359	0.0005		
rs8137	0.0005		
rs2355787	0.0005		
rs2362462	0.0005		
rs2362457	0.0005		
rs118154483	0.0005		
rs10156003	0.0005		
rs17621391	0.0004		
rs10253126	0.0004		
rs6467841	0.0004		
rs2284212	0.0004		

rs112596441	0.0003		
rs10155912	0.0003		
rs11410381	0.0003		
rs73468542	0.0003		
rs10085555	0.0003		
rs34762928	0.0003		
rs6965212	0.0003	rs6965212	LOC100129148
rs6947309	0.0003		
rs6467854	0.0003		
rs10230407	0.0003		
rs7781459	0.0003		
rs10954650	0.0003		
rs10244513	0.0003		
rs4732373	0.0003		
rs4728463	0.0003		
rs2267704	0.0003		
rs7779390	0.0003		
rs112945674	0.0003		
rs10259374	0.0002		
rs7805363	0.0002		
rs9642120	0.0002		
rs6467849	0.0002		
rs740204	0.0002		
rs7779900	0.0002		
rs7786469	0.0002		
rs6946113	0.0002		
rs10228643	0.0002		
rs7795520	0.0002		
rs140274527	0.0002		
rs6960876	0.0002		
rs7781157	0.0002		
rs10269108	0.0002		
rs12537434	0.0002		
rs2355788	0.0002		
rs17613690	0.0002		
rs7779314	0.0002		
rs10085902	0.0002		
rs9683	0.0002		
rs7781579	0.0002		
rs3778931	0.0002		
rs7805824	0.0002		
rs6978680	0.0002		
rs7799165	0.0002		
rs11760420	0.0002		
rs3779130	0.0002		

rs6467852	0.0002		
rs35209997	0.0002		
rs111461944	0.0002		
rs6946393	0.0002		
rs13068	0.0002		
rs10278377	0.0002		
rs9642121	0.0002		
rs6943788	0.0002		
rs10085708	0.0002		
rs4732370	0.0002		
rs10242419	0.0002		
rs1137297	0.0002		
rs7799201	0.0002		
rs10271297	0.0002		
rs6467845	0.0002		
rs1986447	0.0002		
rs6467839	0.0002		
rs7789620	0.0002		
rs10265718	0.0002		
rs4732366	0.0002		
rs2355784	0.0002		
rs6973964	0.0002		
rs4283989	0.0002		
rs10085796	0.0002		
rs6467840	0.0002		
rs6953963	0.0002		
rs10085814	0.0002		
rs11973507	0.0002		
rs12154842	0.0002		
rs10261557	0.0002		
rs10085527	0.0002		
rs10281085	0.0002		
rs6967836	0.0002		
rs6974099	0.0002		
rs6974076	0.0002		
rs7792523	0.0002		
rs4732376	0.0002		
rs4732369	0.0002		
rs6962637	0.0002		
rs10085842	0.0002		
rs2355786	0.0002		
rs17613864	0.0002		
rs7792752	0.0002		
rs2355785	0.0002		
rs4728466	0.0002		

	rs6467846	0.0002		
	rs6944104	0.0002		
2452	rs10964131	0.257159	rs10964131	ACER2
	rs4327934	0.161998		
	rs10811183	0.097079		
	rs7875622	0.0753503		
	rs13297629	0.0661864		
	rs7021844	0.0527169		
	rs12003940	0.0505834		
	9:19431017_G_GT	0.0424043		
	rs10811182	0.0190786		
	rs10120929	0.0184035		
	rs4610844	0.0119707		
	rs7037772	0.0111509		
	rs7023514	0.00822637		
	rs7027346	0.00740958		
	rs6475330	0.0068996		
	rs10964135	0.00667917		
	rs7869763	0.00661206		
	rs7027076	0.00652961		
	rs7023819	0.00633114		
	rs13287117	0.00478995		
	rs10964137	0.00464273		
	rs2001604	0.00358888		
	rs10118089	0.00317964		
	rs10125261	0.00275931		
	rs6475314	0.00212731		
	rs10125954	0.00202583		
	rs4418423	0.00201858		
	rs6475315	0.00196841		
	rs10114121	0.00194281		
	rs4977489	0.00189657		
	rs6475329	0.00164079		
	rs12340681	0.00133952		
	rs10120186	0.00124641		
	9:19434762_C_CT	0.00121552		
	rs10120306	0.00121146		
2522	rs885150	0.1626		
	rs10818578	0.1486	rs10818578	DAB2IP
	rs10818580	0.1411		
	rs10985344	0.1326		
	rs10985347	0.1258		
	rs10985348	0.1121		
	rs10818579	0.1058	rs10818579	DAB2IP
	rs62572789	0.0162		

| | | rs10985343 | 0.0147 | | |

*PPA is defined as variant with posterior probability of association (PPA)≥0.5) the top prioritised variant agrees between the two approaches.

Table 36. Summary of prioritised functional CAD SNPs.

CAD lead / proxy SNP (1%FDR set) located in vascular accessible chromatin (out of 499)	Nearest gene	156 CAD-associated SNPs with either ENCODE HUVEC AP-1 binding or predicted AP-1 motif changed using Haploreg	29 CAD-associated SNPs with a high priority score of 1 (a-f) using RegulomeDB	62 CAD-associated variants overlapping between CAD data and FGWAS approach
rs10171574	CXCR4			
rs10183863	CXCR4			
rs1370526	DKFZp686O1327			
rs1370525	DKFZp686O1327			
rs4668380	MIR873			
rs149846585	NBEAL1			
rs3755237	NRP2			
rs3738484	MCL1			
rs6127200	TGM2	▓	▓	
rs8141133	CABIN1			
rs2370576	ANAPC13			
rs1863911	ANAPC13			
rs62271373	TSC22D2			
rs10912901	KIAA0040			
rs6853156	REST	▓		▓
rs80242894	LOC645513			
rs28634456	LOC645513			
rs78332141	LOC645513			
rs28714318	LOC645513			
rs28495013	LOC645513			
rs6823685	PALLD	▓		▓
rs6854026	PALLD	▓		▓
rs1032763	LOC285696			
rs113728457	FER	▓		▓
rs79536589	FER	▓		▓
rs75421844	FER			
rs7732639	FBN2			
rs1958604	EBF1			
rs16878812	FKBP5			
rs9400476	TRAF3IP2-AS1			
rs4945487	MARCKS			
rs7770043	MARCKS			

rs73534327	MARCKS			
rs9388486	CENPW			
rs6906478	EPB41L2			
rs118099258	GNA12			
rs7807755	RAC1			
rs6965212	LOC100129148			
rs181657044	MTUS1			
rs73551707	BMP1			
rs6982502	TRIB1			
rs58914007	ST3GAL1			
rs59413766	ST3GAL1			
rs16904940	ST3GAL1			
rs1537373	CDKN2B-AS1			
rs7861724	DCAF10			
rs12555751	DCAF10			
rs10814624	DCAF10			
rs75986742	DCAF10			
rs11789185	ENG			
rs138415650	IER5L			
rs10819474	IER5L			
rs3737180	ZEB1			
rs2433027	PLCE1			
rs71485762	OBFC1			
rs4627080	TMEM41B			
rs2593650	SLC17A6			
rs3016316	RIN1			
rs7104718	PPFIA1			
rs2508619	KLHL35			
rs4252591	NA			
rs1784459	VPS11			
rs1786141	VPS11			
rs7127978	UBASH3B			
rs7128198	UBASH3B			
rs7129071	UBASH3B			
rs12403025	SLC6A9			
rs2371494	NA			
rs12817989	FGD6			
rs2302700	SSH1			
rs2302701	SSH1			
rs17086701	FLT1			
rs10689124	STARD13			
rs9569666	STARD13			
rs9569667	STARD13			
rs12858634	STARD13			
rs4942039	NAA16			

rs11206803	PPAP2B			
rs71445078	CARS2			
rs3986303	EXD1			
rs11636934	LOXL1	▓	▓	
rs2507	LOXL1			
rs4888409	CFDP1			
rs9913156	ARRB2			
rs2644383	NPEPPS			
rs1010322	ZNF652			
rs7246865	MYO9B	▓		▓
rs35976034	MAP1S	▓		▓
rs2241719	TGFB1			
rs2241718	TGFB1	▓	▓	
rs12462166	B9D2			
rs2241712	B9D2			
rs60572996	HEATR5B			
rs1549721	ZFP36L2			
rs2023292	ZFP36L2			
rs582384	PRKCE	▓		▓
rs6720415	AAK1			
rs4315609	NCRNA00189			
rs12277907	NA			
rs2275117	ZMYND12			
rs57749886	RCOR1			
rs4794019	IGF2BP1	▓	▓	
rs4968721	PECAM1	▓		▓
rs61459202	TGFB1			
rs7254679	TGFB1			
rs10692845	APOB			
rs11676780	LRRFIP1			
rs16987150	TGM2			
rs17196927	TGM2			
rs17196913	TGM2			
rs11204859	TUFT1			
rs6762186	FRMD4B			
rs1727887	ARHGEF26			
rs11248061	DGKQ			
rs10902762	DGKQ			
rs13193957	MARCKS			
rs2502400	MARCKS			
rs17080102	PLEKHG1			
rs188227226	MTUS1			
rs3737178	ZEB1			
rs3737179	ZEB1			
rs11000897	ADK			

rs4934258	BMPR1A			
rs61871680	BTBD16	▓		▓
rs1002707	SRRM1			
rs56348932	PPAP2B	▓		▓
rs4888403	CFDP1			
rs113985803	BPTF			
rs71378928	WIPI1	▓		▓
rs3901289	BTBD16			
rs28372895	CD151			
rs8077485	NPEPPS	▓		▓
rs7328306	FLT1			
rs9569668	STARD13			
rs17293632	SMAD3	▓		▓
rs9936936	CMIP			
rs13031876	GEN1			
rs12105229	AOX2P			
rs1781416	TUFT1			
rs9606203	ARVCF	▓		▓
rs62355881	NA			
rs62380877	KDM3B			
rs251023	DIAPH1			
rs192240803	CCNT1			
rs185029099	CCNT1			
rs184368932	CCNT1			
rs17114036	PPAP2B			
rs7256001	ZNF568			
rs117955557	EXOC3L2			
rs2345383	TGM2			
rs112379863	KDM3B			
rs631556	NAV1			
rs6997760	DOCK5			
rs9569651	STARD13			
rs10911228	LAMC1			
rs78785901	SASH1			
rs1874883	ANAPC13			
rs34826769	PAPD7			
rs6728904	RFTN2			
rs56316522	KDM3B			
rs1534310	CNPY4			
rs1332327	LIPA		▓	▓
rs7950280	PPFIA1			
rs175040	EIF2B2		▓	▓
rs1894401	FES		▓	▓
rs2070737	RSPH6A		▓	▓
rs7252126	RSPH6A			

rs10824083	ADK			
rs6442108	CDC25A			
rs6442109	CDC25A			
rs7793613	COG5			
rs10824084	ADK			
rs28558789	CD151			
rs1784460	VPS11			
rs179744	AP4S1			
rs5002487	SREBF1			
rs4360494	FHL3			
rs11591147	PCSK9			
rs9337951	KIAA1462			
rs1332328	LIPA			
rs2279286	NA			
rs2306363	SIPA1			
rs10488763	FDX1			
rs10849546	NA			
rs35350651	ATXN2			
rs175499	EIF2B2			
rs7189106	PLCG2			
rs76341367	FOXC2			
rs11650046	CYTH1			
rs116843064	ANGPTL4			
rs11670056	ELL			
rs36091860	RSPH6A			
rs114192718	SAP130			
rs1250259	FN1			
rs75587772	PREX1			
rs73260214	PREX1			
rs80346118	PREX1			
rs56155140	REST			
rs66790703	REST			
rs4724806	RAC1			
rs3828944	RAC1			
rs2107595	HDAC9			
rs73099190	TBX20			
rs2270223	TBX20			
rs17674947	TBX20			
rs7781964	NA			
rs111862082	MKRN1			
rs10265	LOC100129148			
rs3924445	LOC100129148			
rs11765421	LOC100129148			
rs6962366	LOC100129148			
rs6965143	LOC100129148			

rs6947215	LOC100129148				
rs10964131	ACER2				
rs10818578	DAB2IP				
rs10818579	DAB2IP				

Table 37. 517 BP-associated SNPs identified in vascular accessible chromatin.

Proxy rsid	Sentinel rsid	Genes	*p*-value (artery-coronary)	*p*-value (artery-aorta)	*p*-value (artery-tibial)
rs2612211	rs4385883	OR4A5	NA	NA	NA
rs11820037	rs2298807	P2RY6	0.63	0.17	3.4E-06
rs203283	rs1565716	RP1-212P9.2	NA	NA	NA
rs571080	rs1565716	RP1-212P9.2	NA	NA	NA
rs11605954	rs12807220	YAP1	0.14	0.93	0.22
rs263533	rs260508	RP11-181G12.4	0.57	0.19	NA
rs513425	rs12362593	LAYN	0.51	0.017	0.83
rs2850247	rs12362593	LAYN	0.7	0.0097	0.69
rs11607161	rs12574332	UBASH3B	0.0015	0.75	0.065
rs7129204	rs12574332	UBASH3B	0.00065	0.76	0.2
rs7936858	rs11222386	RP11-890B15.2	0.064	0.86	0.014
rs4508210	rs11222386	RP11-890B15.2	0.068	0.39	0.0053
rs4403816	rs11222386	RP11-890B15.2	0.068	0.39	0.0053
rs1559671	rs11222386	RP11-890B15.2	0.068	0.39	0.0053
rs4403815	rs11222386	RP11-890B15.2	0.068	0.39	0.0053
rs1107479	rs7137749	BAZ2A	0.84	0.61	0.19
rs3214051	rs7137749	BAZ2A	0.79	0.94	0.18
rs11112548	rs11112548	C12orf75	0.19	0.36	0.57
rs16950044	rs3898618	PXN	0.086	0.42	0.66
rs16950047	rs3898618	PXN	0.086	0.42	0.66
rs77921082	rs3898618	PXN	0.32	0.82	0.83
rs3847969	rs3898618	PXN	0.086	0.48	0.7
rs3847968	rs3898618	PXN	0.053	0.47	0.69
rs61665022	rs3898618	PXN	0.1	0.42	0.77
rs16950065	rs3898618	PXN	0.1	0.42	0.77
rs76603099	rs3898618	PXN	0.094	0.6	0.72
rs2208851	rs9532243	RXFP2	NA	0.003	0.45
rs3812891	rs9549297	FOXO1	0.64	0.67	0.19
rs4942566	rs912434	LRCH1	0.011	0.0000044	0.00007
rs1063181	rs12583637	SPRYD7	0.019	0.032	1.9E-08

rs74078717	rs12583637	SPRYD7	0.018	0.029	1.4E-08
rs9670083	rs12583637	SPRYD7	0.01	0.038	5.1E-08
rs28411370	rs12583637	SPRYD7	0.019	0.032	1.9E-08
rs7981648	rs12583637	SPRYD7	0.059	0.011	3.9E-08
rs3861113	rs3861113	DACH1	0.36	0.082	0.12
rs78307732	rs72659998	ZMYND12	0.000087	2.7E-09	1.6E-16
rs7549876	rs839755	TIE1	0.013	0.0000002	1.6E-07
rs3768046	rs839755	TIE1	0.023	9.9E-09	7E-08
rs839752	rs839755	TIE1	0.02	0.00000013	0.000048
rs4899527	rs11159091	LTBP2	0.053	2.1E-12	0.00025
rs12894937	rs11159091	LTBP2	0.092	5.9E-09	0.00077
rs11852134	rs11159091	LTBP2	0.092	7.4E-10	0.00014
rs1077662	rs11159091	LTBP2	0.079	2.8E-09	0.00049
rs6429576	rs512083	AKR1A1	0.85	0.54	0.65
rs12890225	rs28470843	RP11-638I2.4	NA	NA	0.8
rs72692250	rs6681713	FCF1P6	NA	NA	NA
rs72692291	rs6681713	FCF1P6	NA	NA	NA
rs2037517	rs832890	AC069368.3	NA	NA	NA
rs3825877	rs7180952	ZSCAN2	0.73	0.67	0.00023
rs698620	rs7180952	ZSCAN2	0.91	0.64	0.00021
rs4842978	rs7180952	ZSCAN2	0.96	0.93	0.00035
rs12934086	rs9932866	AL022341.3	0.092	0.000073	0.000015
rs11865749	rs9932866	AL022341.3	0.0067	0.00096	0.000035
rs4984805	rs11248862	TPSP2	NA	NA	NA
rs11645809	rs11248862	TPSP2	NA	NA	NA
rs1646010	rs37060	SETD6	0.28	0.32	0.083
rs1981960	rs37060	SETD6	0.28	0.34	0.083
rs1981961	rs37060	SETD6	0.28	0.34	0.083
rs27097	rs37060	SETD6	0.39	0.24	0.091
rs45474499	rs45474499	CMTM4	0.85	0.22	1
rs9894790	rs8069739	RP11-599B13.6	NA	NA	NA
rs9901637	rs8069739	RP11-599B13.6	NA	NA	NA
rs11650713	rs8069739	RP11-599B13.6	NA	NA	NA
rs1472932	rs12938803	EPN2	9.9E-12	1.5E-20	1.5E-77
rs2233072	rs12938803	EPN2	5.9E-10	5.4E-16	1.9E-62
rs12945428	rs3135967	CCT6B	0.036	0.0027	4.5E-08
rs369184	rs34430710	RP11-112H10.4	NA	NA	NA
rs430973	rs34430710	RP11-112H10.4	NA	NA	NA

rs444393	rs34430710	RP11-112H10.4	NA	NA	NA
rs654743	rs34430710	RP11-112H10.4	NA	NA	NA
rs57067187	rs112280096	RP11-1055B8.3	0.017	0.000000012	5.1E-06
rs8067402	rs112280096	RP11-1055B8.3	0.0086	1.9E-09	0.000016
rs35699948	rs963920	SPIRE1	0.93	0.000011	3.8E-14
rs1343778	rs2065152	RP11-302M6.4	NA	NA	NA
rs3730668	rs34163044	POLI	9.7E-24	6.2E-38	7.2E-67
rs2065152	rs2065152	RP11-302M6.4	NA	NA	NA
rs553039	rs2065152	RP11-302M6.4	NA	NA	NA
rs2065153	rs2065152	RP11-302M6.4	NA	NA	NA
rs9953520	rs7235890	NEDD4L	0.13	0.0000062	2.6E-12
rs6511291	rs6511291	RP11-420K14.3	NA	NA	NA
rs874513	rs2707238	LINC00211	0.0066	0.038	0.74
rs146829035	rs11681462	E ML4	0.64	0.16	0.85
rs6545867	rs7608483	AHSA2	0.01	0.00053	0.000024
rs3198055	rs4851462	ZAP70	0.89	0.81	0.00063
rs57633263	rs150194832	FHL2	0.0053	0.000065	1.2E-08
rs56294457	rs150194832	FHL2	0.0053	0.000065	1.2E-08
rs12053113	rs12990959	ACVR2A	0.000028	1.3E-12	1.5E-20
rs897172	rs12990959	ACVR2A	0.000028	1.3E-12	1.3E-20
rs6746656	rs6758859	MLTK	NA	NA	NA
rs2176550	rs10184839	UBE2E3	0.26	0.00032	0.0001
rs6711592	rs1996992	STK36	0.67	0.41	0.0066
rs2280039	rs12474050	SPEG	0.9	0.14	0.83
rs11698137	rs11087740	BMP2	0.19	0.1	0.74
rs2876049	rs11087740	BMP2	0.11	0.07	0.58
rs2206799	rs1232482	LINC00687	NA	NA	NA
rs1998107	rs2618647	SNX5	0.95	0.86	0.23
rs67611724	rs13042148	CDK5RAP1	1	0.38	0.11
rs6021247	rs6021247	NFATC2	0.71	0.54	0.53
rs62229372	rs62229372	AP000692.10	0.0075	0.00023	0.0028
rs2277788	rs2277788	WRB	0.000018	0.0000029	1.2E-10
rs3738484	rs11585169	RN7SL473P	0.098	0.13	0.073
rs3738485	rs11585169	RN7SL473P	0.086	0.12	0.081
rs2330592	rs11701512	PCNT	0.026	0.026	0.0016
rs134117	rs134041	MN1	0.044	0.0063	0.000035
rs76013375	rs737721	ZMAT5	0.088	0.77	0.57
rs2240366	rs737721	ZMAT5	0.63	0.3	0.64

rs5994293	rs5753103	RP1-130H16.18	NA	NA	NA
rs6801957	rs6801957	SCN10A	NA	0.012	NA
rs4682994	rs141979279	ZNF502	0.56	0.39	0.57
rs2292180	rs141979279	ZNF502	0.72	0.34	0.54
rs2292181	rs141979279	ZNF502	0.72	0.34	0.54
rs1727887	rs73158427	RP11-23D24.2	NA	NA	NA
rs12493885	rs73158427	RP11-23D24.2	NA	NA	NA
rs4680454	rs78151625	MLF1	0.16	0.58	0.74
rs73156496	rs78151625	MLF1	0.18	0.91	0.72
rs7611674	rs7611674	GNB4	0.0022	0.00000043	0.00024
rs1706003	rs1706003	TMEM44	0.00055	0.0000003	0.000013
rs61792599	rs55829085	POLN	0.38	0.17	0.0023
rs115944162	rs55829085	POLN	0.23	0.64	0.0076
rs3822239	rs3796822	SLC2A9	0.0000014	0.0000001	1.5E-10
rs2610989	rs2610990	LCORL	0.65	0.076	0.51
rs982146	rs6823199	THAP9-AS1	2.8E-12	8.2E-35	2.8E-67
rs4693555	rs6823199	THAP9-AS1	0.00000008	8.8E-24	4.3E-40
rs11929841	rs223361	MANBA	0.59	0.00012	0.00091
rs13109676	rs4699165	TET2	0.41	0.26	0.3
rs778593	rs702395	NDUFA2	0.0056	0.001	0.00014
rs778594	rs702395	NDUFA2	0.0056	0.001	0.00014
rs2563335	rs702395	NDUFA2	0.0066	0.0049	0.0031
rs801183	rs702395	NDUFA2	0.0066	0.0049	0.0027
rs3097175	rs1650911	AC005592.2	0.79	0.41	0.45
rs114503346	rs114503346	DUSP1	0.82	0.56	0.065
rs117012610	rs114503346	DUSP1	0.9	0.76	0.073
rs6879874	rs28362590	RAB24	0.098	0.43	1.9E-07
rs114760566	rs115245297	HMGA1	0.95	0.1	0.86
rs2298307	rs7753695	BCKDHB	0.066	0.023	0.0098
rs35447745	rs7753695	BCKDHB	0.0083	0.0055	0.0025
rs12134698	rs7555285	TRAF3IP3	0.7	0.57	0.16
rs1800797	rs1800795	AC073072.5	3.4E-21	1.4E-22	5.9E-33
rs1800795	rs1800795	AC073072.5	1E-20	1.7E-23	5E-31
rs7357209	rs17454517	GRB10	0.03	0.00000019	0.21
rs194505	rs11770630	STEAP2-AS1	NA	NA	NA
rs17558745	rs35981664	TGFB2	0.89	0.26	0.69
rs145411968	rs115172170	HBP1	0.56	0.99	0.31
rs1870735	rs1870735	AC021218.2	NA	NA	NA
rs12698354	rs1870735	AC021218.2	NA	NA	NA
rs10097936	rs7845722	FUT10	7.6E-15	2.3E-31	1E-36

rs2732317	rs7845722	FUT10	1.2E-16	5.5E-28	1E-36
rs2732318	rs7845722	FUT10	1.2E-16	5.5E-28	1E-36
rs10097189	rs7845722	FUT10	9.9E-15	1.6E-30	2.6E-35
rs2732288	rs7845722	FUT10	5.8E-15	4.4E-29	1.6E-34
rs2581897	rs7845722	FUT10	5.8E-15	4.4E-29	1.6E-34
rs55962869	rs1906672	RP11-90P5.2	NA	NA	NA
rs2270376	rs1906672	RP11-90P5.2	NA	NA	NA
rs1149	rs1906672	RP11-90P5.2	NA	NA	NA
rs12678205	rs1906672	RP11-90P5.2	NA	NA	NA
rs6996860	rs1906672	RP11-90P5.2	NA	NA	NA
rs10109326	rs13253358	PREX2	0.39	0.19	0.8
rs7822260	rs7009170	LRRC69	0.097	0.56	0.91
rs10097617	rs4582532	NDUFAF6	0.0021	0.029	4.6E-07
rs10108430	rs4582532	NDUFAF6	0.011	0.014	3E-07
rs2247355	rs2513877	AZIN1	0.00081	0.0000081	1.1E-21
rs1375062	rs10087782	PTK2	0.83	0.0019	0.06
rs6578146	rs10087782	PTK2	0.61	0.003	0.34
rs6578147	rs10087782	PTK2	0.61	0.003	0.34
rs13259836	rs10087782	PTK2	0.83	0.0019	0.06
rs481905	rs520015	CBWD1	2E-12	6.7E-11	7.6E-26
rs17369631	rs60191654	KANK1	0.27	0.6	0.97
rs4741961	rs28558845	GLIS3	0.28	0.57	0.51
rs10819384	rs7869756	RP11-339B21.8	0.06	0.00000096	0.00022
rs10819474	rs184457	CRAT	3.8E-10	3E-31	2.5E-24
rs12379849	rs11145807	EGFL7	0.026	0.67	0.4
rs3736460	rs56352451	FAM208B	0.025	0.093	0.5
rs1890876	rs10906391	BEND7	0.043	0.14	1.5E-08
rs11597888	rs76164690	RP11-166N17.1	0.78	0.63	0.98
rs3740053	rs10823136	DNAJC12	0.94	0.96	0.77
rs3740051	rs10823136	DNAJC12	0.94	0.79	0.86
rs1490008	rs10823136	DNAJC12	0.44	0.53	0.39
rs10823977	rs12572586	PLA2G12B	NA	NA	NA
rs28384495	rs12572586	PLA2G12B	NA	NA	NA
rs2271908	rs12572586	PLA2G12B	NA	NA	NA
rs3998306	rs12572586	PLA2G12B	NA	NA	NA
rs3812623	rs12572586	PLA2G12B	NA	NA	NA
rs78581136	rs12572586	PLA2G12B	NA	NA	NA
rs12255505	rs12572586	PLA2G12B	NA	NA	NA
rs10887881	rs10887914	RP11-137H2.6	NA	NA	NA
rs11202929	rs10887914	RP11-137H2.6	NA	NA	NA

rs116857892	rs10887914	RP11-137H2.6	NA	NA	NA
rs113816392	rs111777102	MXI1	0.65	0.000005	0.0063
rs11826681	rs11031051	ARL14EP	0.19	0.012	0.35
rs3747560	rs7965392	GXYLT1	0.036	0.0000041	4.6E-08
rs56733641	rs3898618	PXN	0.17	0.57	0.79
rs57279831	rs3898618	PXN	0.1	0.42	0.77
rs10773037	rs28498002	BCL7A	0.013	0.45	0.73
rs748676	rs3861113	DACH1	0.36	0.082	0.12
rs17190203	rs55684003	MBNL2	0.76	0.19	0.027
rs1051533	rs57786342	ZFP36L1	0.92	0.19	0.84
rs7212271	rs34430710	RP11-112H10.4	NA	NA	NA
rs17396055	rs17396055	ARHGAP29	0.0022	0.036	0.011
rs62189009	rs55732192	TBR1	NA	NA	NA
rs115283512	rs151054210	CYBRD1	5.7E-10	2E-25	2.2E-58
rs10494474	rs7524019	POU2F1	0.75	0.28	0.058
rs72927138	rs6823199	THAP9-AS1	1.4E-10	3.3E-33	2.7E-62
rs60558877	rs1906672	RP11-90P5.2	NA	NA	NA
rs11992645	rs10087782	PTK2	0.83	0.0019	0.06
rs3124498	rs184457	CRAT	1E-12	5E-37	2.7E-29
rs1612013	rs12362593	LAYN	0.76	0.0041	0.75
rs1611989	rs12362593	LAYN	0.76	0.0041	0.75
rs957634	rs1565716	RP1-212P9.2	NA	NA	NA
rs9526704	rs9526707	RNASEH2B-AS1	8.4E-11	1E-22	6.1E-34
rs9429085	rs512083	AKR1A1	0.23	0.68	0.65
rs59333122	rs59333122	CTD-2349B8.1	NA	NA	NA
rs11117317	rs6540125	BANP	0.32	0.14	0.96
rs11117318	rs6540125	BANP	0.26	0.14	0.58
rs11117319	rs6540125	BANP	0.26	0.14	0.58
rs9957225	rs7235890	NEDD4L	0.13	0.00001	4.8E-13
rs7525880	rs7514579	BCAR3	0.01	8.9E-15	5.9E-19
rs13430254	rs10193543	EXOC6B	0.066	0.000014	0.00078
rs6737270	rs150194832	FHL2	0.0053	0.000065	1.2E-08
rs753641	rs1432457	RP11-574H6.1	NA	NA	NA
rs2531347	rs702395	NDUFA2	0.0049	0.024	0.037
rs675162	rs555754	SLC22A3	0.07	0.95	0.0056
rs3735665	rs6593297	PSPH	0.015	0.0051	0.00029
rs688791	rs2513877	AZIN1	0.0016	0.00013	2.4E-19
rs143691840	rs12572586	PLA2G12B	NA	NA	NA
rs11000522	rs12572586	PLA2G12B	NA	NA	NA
rs11000521	rs12572586	PLA2G12B	NA	NA	NA

rs2394935	rs12572586	PLA2G12B	NA	NA	NA
rs9419219	rs7912283	PPP2R2D	7.9E-11	6.1E-19	7.5E-40
rs173396	rs360153	0	NA	NA	NA
rs3737803	rs3737801	FGR	0.76	0.0057	6.5E-06
rs3737801	rs3737801	FGR	0.87	0.023	0.000025
rs729354	rs2450128	USP35	0.69	0.56	0.019
rs12804962	rs12362593	LAYN	0.77	0.0068	0.98
rs11222383	rs11222386	RP11-890B15.2	0.068	0.39	0.0053
rs260508	rs260508	RP11-181G12.4	0.87	0.3	NA
rs10783796	rs7137749	BAZ2A	0.79	0.99	0.11
rs16950061	rs3898618	PXN	0.1	0.42	0.77
rs140025566	rs28498002	BCL7A	0.028	0.53	0.79
rs839995	rs839755	TIE1	0.017	0.00000049	8.8E-07
rs3014216	rs512083	AKR1A1	0.23	0.6	0.75
rs4468203	rs512083	AKR1A1	0.69	0.64	0.66
rs77150581	rs4926923	FLJ00388	NA	NA	NA
rs80020621	rs4926923	FLJ00388	NA	NA	NA
rs1916137	rs832890	AC069368.3	NA	NA	NA
rs1916138	rs832890	AC069368.3	NA	NA	NA
rs12439517	rs62004794	PIAS1	0.013	0.05	0.051
rs12439519	rs62004794	PIAS1	0.013	0.05	0.051
rs11639761	rs6540125	BANP	0.29	0.076	0.83
rs2304909	rs8069739	RP11-599B13.6	NA	NA	NA
rs57349131	rs9904409	FZD2	0.53	0.94	0.08
rs7219390	rs35504735	MXRA7	8E-12	8.3E-27	3.4E-49
rs4496245	rs7235890	NEDD4L	0.16	0.0000083	2.8E-12
rs34746918	rs8111708	AC008397.1	NA	NA	NA
rs10413237	rs8111708	AC008397.1	NA	NA	NA
rs8103622	rs8111708	AC008397.1	NA	NA	NA
rs6747242	rs150194832	FHL2	0.0053	0.000065	1.3E-08
rs6719765	rs150194832	FHL2	0.0053	0.000065	1.2E-08
rs884724	rs150194832	FHL2	0.0035	4.8E-09	3.3E-13
rs57198343	rs2618647	SNX5	0.82	0.42	0.26
rs2016807	rs11701512	PCNT	0.026	0.026	0.0016
rs12497267	rs73158427	RP11-23D24.2	NA	NA	NA
rs35481032	rs10913934	QSOX1	0.0021	1.3E-12	1.4E-32
rs3820312	rs10913934	QSOX1	0.0021	1.3E-12	1.4E-32
rs3211270	rs1848510	PLK2	0.82	0.91	0.021
rs12521065	rs2400509	RP11-373N22.3	0.33	0.06	0.0055

rs34013988	rs114503346	DUSP1	0.69	0.85	0.17
rs116836493	rs115245297	HMGA1	0.82	0.19	0.8
rs13227860	rs4507656	RAPGEF5	0.26	0.0041	0.019
rs4947936	rs17454517	GRB10	0.033	0.0000015	0.15
rs7807731	rs1870735	AC021218.2	NA	NA	NA
rs4131010	rs13253358	PREX2	0.37	0.23	0.84
rs2879813	rs4582532	NDUFAF6	0.0012	0.031	1.1E-06
rs7003387	rs4582532	NDUFAF6	0.004	0.024	5.1E-08
rs6982393	rs4582532	NDUFAF6	0.00091	0.01	7.3E-07
rs11990541	rs10087782	PTK2	0.75	0.0029	0.04
rs28558845	rs28558845	GLIS3	0.13	0.44	0.72
rs62541923	rs11789875	BNC2	0.35	0.027	0.042
rs10123342	rs7045409	BEND3P2	NA	NA	NA
rs1413299	rs10988442	COL15A1	0.11	0.0044	0.0086
rs12254769	rs12572586	PLA2G12B	NA	NA	NA
rs61290608	rs12572586	PLA2G12B	NA	NA	NA
rs11000614	rs12572586	PLA2G12B	NA	NA	NA
rs7098453	rs12572586	PLA2G12B	NA	NA	NA
rs4934170	rs10887914	RP11-137H2.6	NA	NA	NA
rs208078	rs919045	DCDC1	NA	NA	NA
rs57336947	rs11222386	RP11-890B15.2	0.068	0.46	0.0062
rs3794133	rs11222386	RP11-890B15.2	0.055	0.47	0.0037
rs3214009	rs11615689	MED13L	0.1	0.41	0.043
rs10468043	rs62004794	PIAS1	0.0094	0.029	0.028
rs112940674	rs62004794	PIAS1	0.0094	0.033	0.031
rs12149893	rs7185555	HAS3	0.97	0.43	0.12
rs4808801	rs8111708	AC008397.1	NA	NA	NA
rs4808136	rs8111708	AC008397.1	NA	NA	NA
rs11898059	rs1996992	STK36	0.67	0.41	0.0066
rs76354454	rs77692990	BRD1	0.99	0.31	6.2E-07
rs9682313	rs11923667	SENP7	0.000024	8.2E-12	1.6E-30
rs28496953	rs11923667	SENP7	0.000035	4.1E-12	2.7E-29
rs28431102	rs11923667	SENP7	0.000041	1E-11	5E-29
rs7671997	rs55829085	POLN	0.41	0.22	0.0054
rs13116100	rs3097937	LINC01091	0.78	0.55	0.31
rs801168	rs702395	NDUFA2	0.0066	0.0036	0.0027
rs1317983	rs9472135	VEGFA	0.64	0.96	0.48
rs555754	rs555754	SLC22A3	0.041	0.84	0.035
rs668871	rs555754	SLC22A3	0.09	0.96	0.028
rs11760961	rs9638084	LINC01006	0.35	0.93	0.82

rs55967138	rs10087782	PTK2	0.61	0.0023	0.35
rs12246277	rs12572586	PLA2G12B	NA	NA	NA
rs7074274	rs12572586	PLA2G12B	NA	NA	NA
rs579187	rs1565716	RP1-212P9.2	NA	NA	NA
rs842406	rs912434	LRCH1	0.011	0.0000044	0.00007
rs56234729	rs62004794	PIAS1	0.0012	0.1	0.091
rs11117321	rs6540125	BANP	0.26	0.14	0.58
rs11117322	rs6540125	BANP	0.31	0.12	0.63
rs2237503	rs1091811	LIMK1	0.29	0.2	0.031
rs4735340	rs4582532	NDUFAF6	0.00091	0.01	7.3E-07
rs6983948	rs4582532	NDUFAF6	0.0011	0.009	3.4E-07
rs11689538	rs6723509	TFCP2L1	0.27	0.027	0.9
rs67155347	rs13001283	GYPC	0.9	0.35	0.81
rs359708	rs13001283	GYPC	0.9	0.35	0.81
rs2034261	rs4954192	CCNT2-AS1	0.34	0.025	0.14
rs723681	rs12990959	ACVR2A	0.00013	4E-11	4E-17
rs61049574	rs3175	PRPF40A	0.43	0.23	0.00034
rs3175	rs3175	PRPF40A	0.64	0.17	0.000051
rs16846026	rs55732192	TBR1	NA	NA	NA
rs17221848	rs151054210	CYBRD1	0.000000039	1.9E-20	7.3E-45
rs6433880	rs10184839	UBE2E3	0.28	0.00041	0.00015
rs6433881	rs10184839	UBE2E3	0.28	0.00041	0.00015
rs10208742	rs10184839	UBE2E3	0.26	0.00032	0.0001
rs1518012	rs11901929	AC079613.1	NA	NA	NA
rs2138960	rs11901929	AC079613.1	NA	NA	NA
rs2138959	rs11901929	AC079613.1	NA	NA	NA
rs7569460	rs11901929	AC079613.1	NA	NA	NA
rs7569461	rs11901929	AC079613.1	NA	NA	NA
rs11124186	rs4507125	TWIST2	0.37	0.12	0.91
rs2144351	rs1232482	LINC00687	NA	NA	NA
rs8129943	rs2277788	WRB	0.000038	0.000027	4.3E-13
rs78998444	rs737721	ZMAT5	0.58	0.46	0.3
rs11914096	rs77692990	BRD1	0.99	0.31	6.2E-07
rs10918683	rs7524019	POU2F1	0.75	0.28	0.038
rs793499	rs9860290	COL8A1	0.91	0.23	0.88
rs9864437	rs9860290	COL8A1	0.79	0.31	0.91
rs7634759	rs11923667	SENP7	0.000029	5.6E-12	4.2E-30
rs9826148	rs1882289	ZBTB20	0.17	0.45	0.69
rs668437	rs863930	MSL2	0.61	0.02	0.011
rs634559	rs863930	MSL2	0.61	0.02	0.011
rs7641039	rs73158427	RP11-23D24.2	NA	NA	NA

rs61791545	rs73158427	RP11-23D24.2	NA	NA	NA
rs73242122	rs2610990	LCORL	0.61	0.12	0.97
rs3753812	rs10913934	QSOX1	0.0019	1.2E-12	1E-30
rs3789369	rs1043069	XPR1	0.77	0.055	0.087
rs35895091	rs6823199	THAP9-AS1	0.00000017	3.6E-29	7.8E-53
rs12643434	rs6823199	THAP9-AS1	2.7E-12	1.5E-29	2E-63
rs4699831	rs1347345	BMPR1B	0.94	0.65	0.42
rs4698875	rs223361	MANBA	0.097	0.000062	0.00032
rs10035032	rs7439567	PCDH18	0.59	0.21	0.88
rs7728821	rs6875372	CWC27	0.2	0.5	0.17
rs12518430	rs4286632	MAST4	0.14	0.0096	1E-12
rs12566249	rs33996239	ADORA1	0.6	0.29	0.0012
rs35399227	rs1432457	RP11-574H6.1	NA	NA	NA
rs10036322	rs1432457	RP11-574H6.1	NA	NA	NA
rs891905	rs62385385	EBF1	0.42	0.31	0.094
rs891904	rs62385385	EBF1	0.91	0.29	0.1
rs10864035	rs2629665	C1orf116	0.25	0.018	0.065
rs2306714	rs2629665	C1orf116	0.39	0.01	0.074
rs881858	rs9472135	VEGFA	0.64	0.96	0.56
rs9353012	rs9449350	AL359693.1	NA	NA	NA
rs6698945	rs9431431	HLX	0.12	0.37	0.68
rs607284	rs7763294	CITED2	0.033	0.0086	0.000079
rs9403140	rs7763294	CITED2	NA	NA	NA
rs6944243	rs1468520	C1GALT1	NA	NA	NA
rs702814	rs10274928	JAZF1	0.027	2.4E-10	9.2E-19
rs7805470	rs2222544	AUTS2	0.093	0.62	0.5
rs2106274	rs11770630	STEAP2-AS1	NA	NA	NA
rs4727241	rs10245696	CDK14	0.86	0.61	0.12
rs6977966	rs115172170	HBP1	0.19	0.76	0.56
rs73187349	rs115172170	HBP1	0.98	1	0.23
rs73187350	rs115172170	HBP1	0.46	0.96	0.73
rs2442633	rs2922895	MCPH1	0.084	0.36	0.21
rs2442632	rs2922895	MCPH1	0.084	0.36	0.21
rs7836942	rs7845722	FUT10	5E-15	3.6E-31	6.4E-34
rs1043782	rs1906672	RP11-90P5.2	NA	NA	NA
rs55703821	rs1906672	RP11-90P5.2	NA	NA	NA
rs2016875	rs1906672	RP11-90P5.2	NA	NA	NA
rs9643511	rs6996733	TOX	0.2	0.89	0.22
rs1569339	rs2205260	TRPS1	0.13	0.034	0.17
rs6470626	rs4598218	TMEM75	0.48	NA	NA
rs13273943	rs10087782	PTK2	0.94	0.0018	0.1

rs4378011	rs10087782	PTK2	0.63	0.0019	0.099
rs7815191	rs10087782	PTK2	0.88	0.0025	0.28
rs11786859	rs10087782	PTK2	0.69	0.0017	0.037
rs11780030	rs10087782	PTK2	0.82	0.0021	0.033
rs520015	rs520015	CBWD1	1.6E-15	2.9E-15	4.7E-34
rs12002168	rs60191654	KANK1	0.88	0.45	0.41
rs62541920	rs11789875	BNC2	0.26	0.037	0.042
rs77507622	rs11789875	BNC2	0.29	0.021	0.038
rs955975	rs11139596	RP11-15B24.5	NA	NA	NA
rs11139596	rs11139596	RP11-15B24.5	NA	NA	NA
rs11141732	rs11141731	SNORA26	NA	NA	NA
rs1053446	rs7045409	BEND3P2	NA	NA	NA
rs10982910	rs10982910	PAPPA	0.23	0.45	0.4
rs10982925	rs10982910	PAPPA	0.19	0.99	0.41
rs2275260	rs7869756	RP11-339B21.8	0.035	0.000003	0.0013
rs332119	rs1265842	WAC	0.087	0.000069	0.00034
rs332138	rs1265842	WAC	0.081	0.000023	0.00014
rs2273771	rs10823136	DNAJC12	0.94	0.96	0.77
rs10823116	rs10823136	DNAJC12	0.94	0.67	0.93
rs12253317	rs12572586	PLA2G12B	NA	NA	NA
rs7094342	rs12572586	PLA2G12B	NA	NA	NA
rs11000499	rs12572586	PLA2G12B	NA	NA	NA
rs79232872	rs12572586	PLA2G12B	NA	NA	NA
rs12242214	rs12572586	PLA2G12B	NA	NA	NA
rs12242222	rs12572586	PLA2G12B	NA	NA	NA
rs2280369	rs12572586	PLA2G12B	NA	NA	NA
rs36152134	rs12572586	PLA2G12B	NA	NA	NA
rs7922374	rs12572586	PLA2G12B	NA	NA	NA
rs957486	rs12572586	PLA2G12B	NA	NA	NA
rs263535	rs260508	RP11-181G12.4	0.87	0.3	NA
rs61874877	rs11197813	HSPA12A	0.8	0.00000086	0.00013
rs4962723	rs4411245	CTBP2	0.91	0.77	0.8
rs11042064	rs10743086	ST5	0.032	0.079	0.015
rs3763919	rs10743086	ST5	0.041	0.18	0.03
rs11821932	rs11026586	CTD-2140G10.1	NA	NA	NA
rs11026601	rs11026586	CTD-2140G10.1	NA	NA	NA
rs11026603	rs11026586	CTD-2140G10.1	NA	NA	NA
rs10835653	rs11031051	ARL14EP	0.24	0.0087	0.31
rs10767838	rs11031051	ARL14EP	0.3	0.0086	0.34

rs2616816	rs919045	DCDC1	NA	NA	NA
rs7116854	rs504217	PHOX2A	NA	NA	NA
rs7130652	rs504217	PHOX2A	NA	NA	NA
rs11824614	rs2298807	P2RY6	0.85	0.31	0.000022
rs2510042	rs2450128	USP35	0.98	0.41	0.06
rs2511168	rs2450128	USP35	0.96	0.67	0.0068
rs2510032	rs2450128	USP35	0.81	0.55	0.033
rs2512532	rs2450128	USP35	0.9	0.52	0.0054
rs2512533	rs2450128	USP35	0.95	0.5	0.028
rs7104222	rs2450128	USP35	1	0.68	0.0089
rs10899474	rs2450128	USP35	0.85	0.46	0.015
rs11237471	rs2450128	USP35	0.92	0.62	0.0021
rs1784784	rs12362593	LAYN	0.96	0.07	0.66
rs7938360	rs12362593	LAYN	0.66	0.022	0.76
rs17337899	rs12574332	UBASH3B	0.0011	0.77	0.085
rs61911494	rs12574332	UBASH3B	0.0011	0.77	0.085
rs11222384	rs11222386	RP11-890B15.2	0.068	0.39	0.0053
rs11222385	rs11222386	RP11-890B15.2	0.14	0.31	0.0074
rs11222386	rs11222386	RP11-890B15.2	0.14	0.26	0.014
rs2406674	rs7965392	GXYLT1	0.036	0.0000041	4.6E-08
rs7980317	rs7137749	BAZ2A	0.64	0.6	0.11
rs775648	rs4143175	CAND1	0.00011	0.00001	1.4E-10
rs978458	rs5742643	IGF1	0.28	0.18	0.73
rs6539036	rs5742643	IGF1	0.14	0.22	0.63
rs72692299	rs6681713	FCF1P6	NA	NA	NA
rs17498131	rs11615689	MED13L	0.1	0.4	0.043
rs79589896	rs3898618	PXN	0.086	0.42	0.66
rs34532891	rs28498002	BCL7A	0.017	0.49	0.66
rs34974633	rs28498002	BCL7A	0.017	0.52	0.44
rs7485380	rs28498002	BCL7A	0.0045	0.41	0.45
rs61950149	rs2480171	snoU13	NA	NA	NA
rs7326823	rs2480171	snoU13	NA	NA	NA
rs73172203	rs73187288	DGKH	0.92	0.0097	0.12
rs41308548	rs12583637	SPRYD7	0.034	0.019	3.5E-08
rs7325993	rs12583637	SPRYD7	0.023	0.036	7.3E-08
rs17073834	rs12583637	SPRYD7	0.049	0.0062	7.2E-09
rs72675474	rs34983854	MIA2	NA	NA	NA
rs7144602	rs7144602	SAMD4A	0.55	0.15	0.51
rs8007201	rs7144602	SAMD4A	0.67	0.12	0.63

rs11628933	rs11628933	PPM1A	0.01	0.000064	0.022
rs11626509	rs11627326	RP11-497E19.2	NA	NA	NA
rs61980102	rs11627326	RP11-497E19.2	NA	NA	NA
rs41405244	rs20354	SGIP1	0.00044	0.00027	0.000045
rs74087510	rs20354	SGIP1	0.002	0.00025	4.4E-06
rs7151531	rs11160085	RIN3	0.000042	0.019	0.000058
rs11633431	rs832890	AC069368.3	NA	NA	NA
rs2176333	rs62004794	PIAS1	0.00014	0.057	0.027
rs55954678	rs62004794	PIAS1	0.0019	0.081	0.1
rs7194305	rs11642631	CLEC16A	0.61	0.17	0.29
rs11642631	rs11642631	CLEC16A	0.59	0.54	0.062
rs6498160	rs11642631	CLEC16A	0.42	0.28	0.12
rs4843758	rs6540125	BANP	0.29	0.11	0.69
rs4843768	rs6540125	BANP	0.65	0.12	0.65
rs9904964	rs9899540	RP11-227G15.3	NA	NA	NA
rs11165681	rs17516329	TGFBR3	0.095	0.015	0.61
rs9652856	rs34430710	RP11-112H10.4	NA	NA	NA
rs12967888	rs4800420	CTAGE1	NA	0.86	NA
rs4355035	rs4800420	CTAGE1	NA	0.81	NA
rs11876341	rs11876341	MEX3C	0.46	0.44	0.84
rs11083046	rs34163044	POLI	1.4E-23	4.8E-37	2.2E-65
rs4808046	rs3745318	CTD-2562J15.6	NA	NA	NA
rs28375303	rs8111708	AC008397.1	NA	NA	NA
rs3795945	rs10779936	KLHL29	0.14	0.61	0.033
rs7573831	rs2707238	LINC00211	0.0099	0.13	0.92
rs10176023	rs11681462	E ML4	0.39	0.14	0.17
rs243066	rs925484	AC007381.2	NA	NA	NA
rs7561672	rs7608483	AHSA2	0.011	0.0005	0.000036
rs778143	rs7608483	AHSA2	0.0099	0.00093	0.0001
rs10186325	rs7608483	AHSA2	0.0067	0.001	0.000023
rs2518934	rs7608483	AHSA2	0.19	0.31	0.12
rs6742187	rs13014371	WDPCP	0.00085	0.012	0.000011
rs2553052	rs2631669	AC074391.1	0.045	0.16	0.29
rs2553053	rs2631669	AC074391.1	0.021	0.16	0.66
rs2631669	rs2631669	AC074391.1	0.028	0.19	0.88
rs2631668	rs2631669	AC074391.1	0.028	0.19	0.88
rs6711257	rs10193543	EXOC6B	0.017	0.0000017	0.0068
rs7763294	rs7763294	CITED2	0.12	0.042	0.00019
rs11563340	rs11770630	STEAP2-AS1	NA	NA	NA

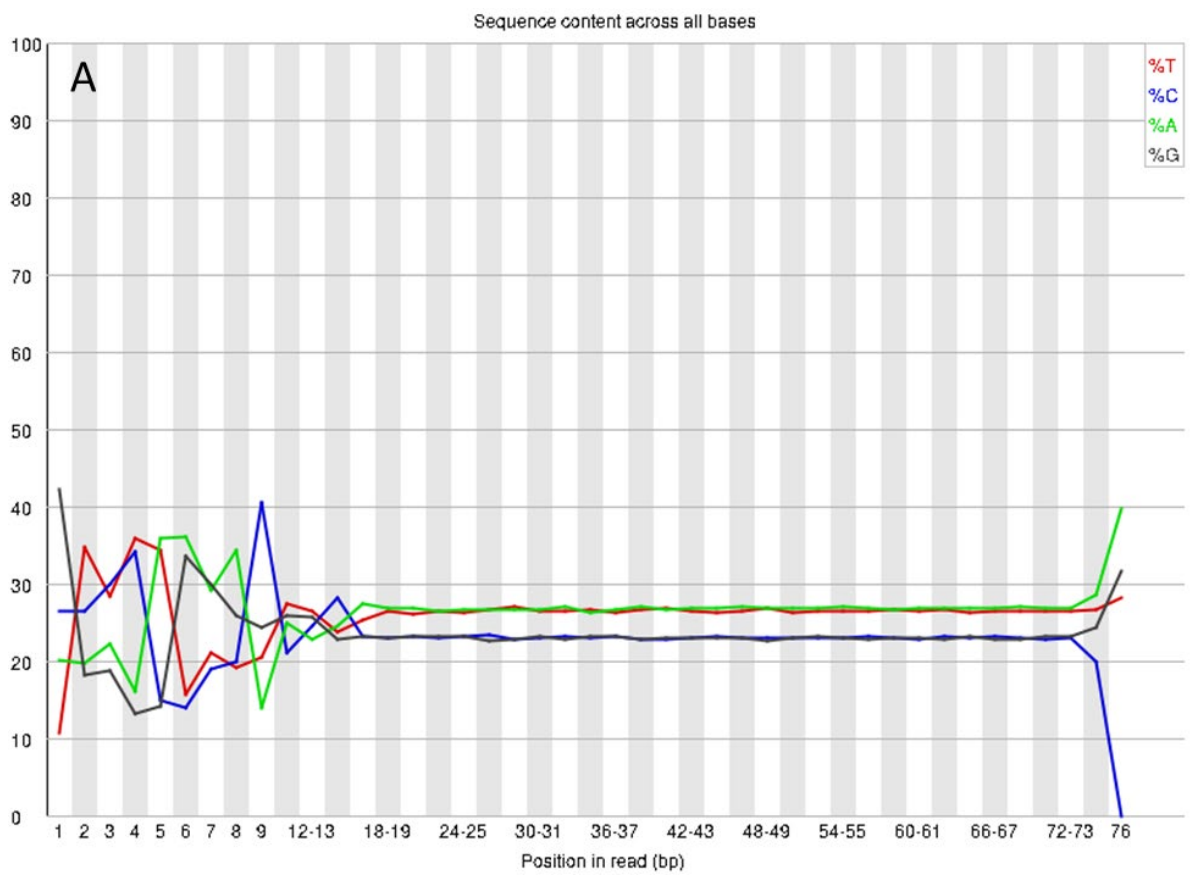| rs10102100	rs61040371	CLDN23	0.73	0.032	0.91
rs7007564	rs61040371	CLDN23	0.8	0.043	0.93
rs7812637	rs2205260	TRPS1	0.33	0.064	0.48
rs4877233	rs11139596	RP11-15B24.5	NA	NA	NA
rs506807	rs1565716	RP1-212P9.2	NA	NA	NA
rs56381931	rs11615689	MED13L	0.12	0.4	0.023
rs11608792	rs28498002	BCL7A	0.0067	0.43	0.56
rs9526707	rs9526707	RNASEH2B-AS1	8.6E-09	6.6E-19	2.2E-28
rs9535528	rs9526707	RNASEH2B-AS1	8.4E-11	1E-22	3.7E-33
rs12147199	rs57786342	ZFP36L1	0.41	0.54	0.89
rs8066162	rs9904409	FZD2	0.45	0.86	0.23
rs9303774	rs963920	SPIRE1	0.83	0.0000062	1.9E-12
rs4851462	rs4851462	ZAP70	0.73	0.89	0.004

9.2 Supplementary figures

252
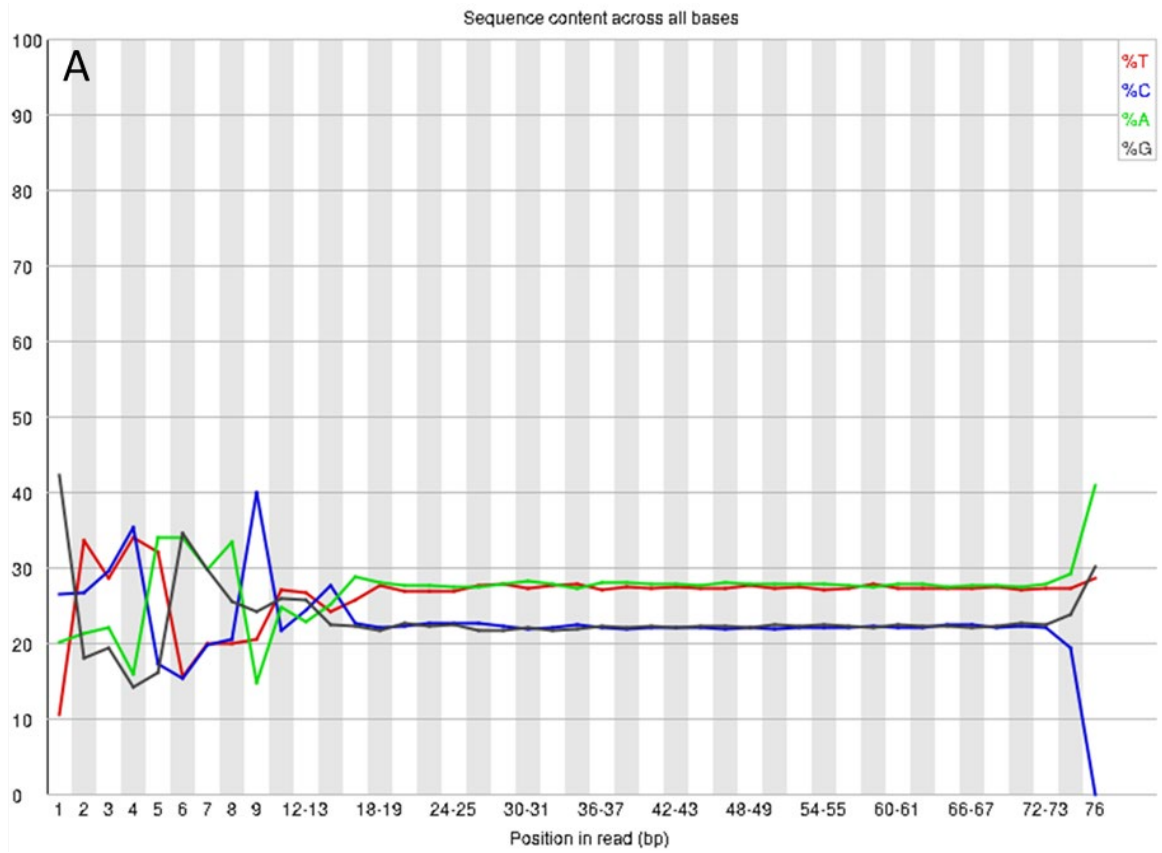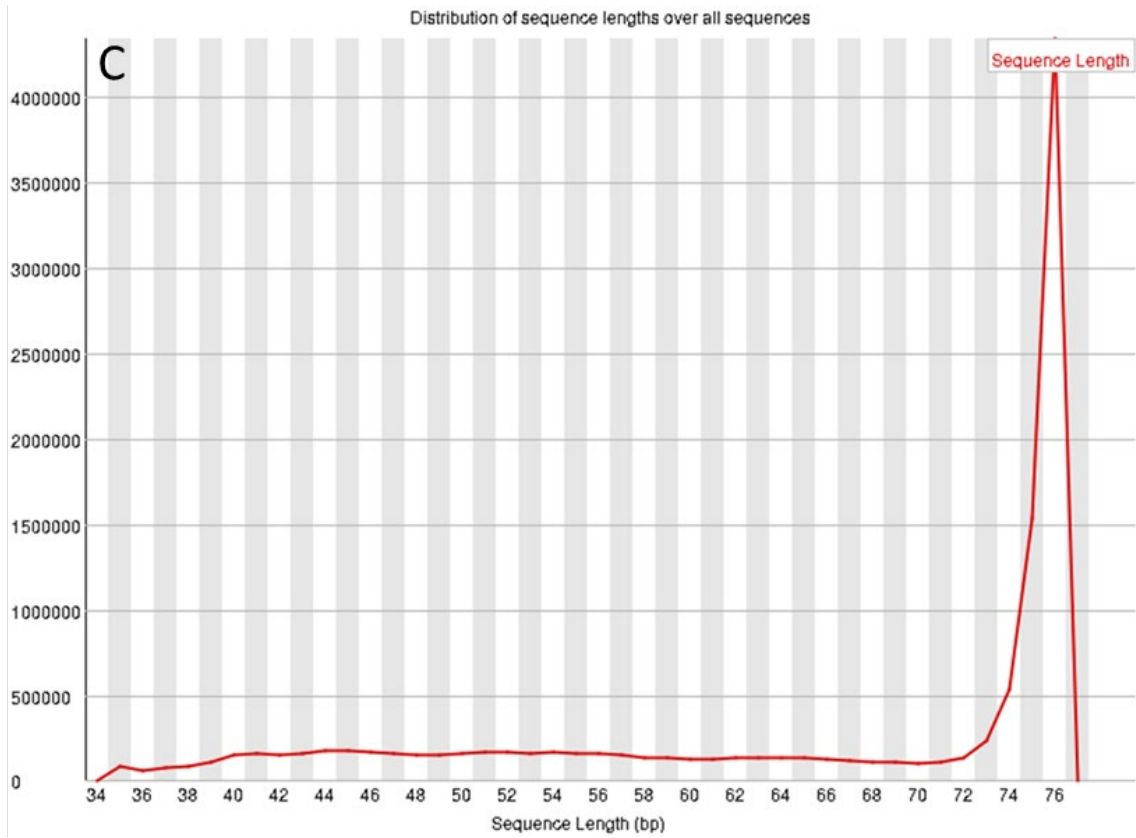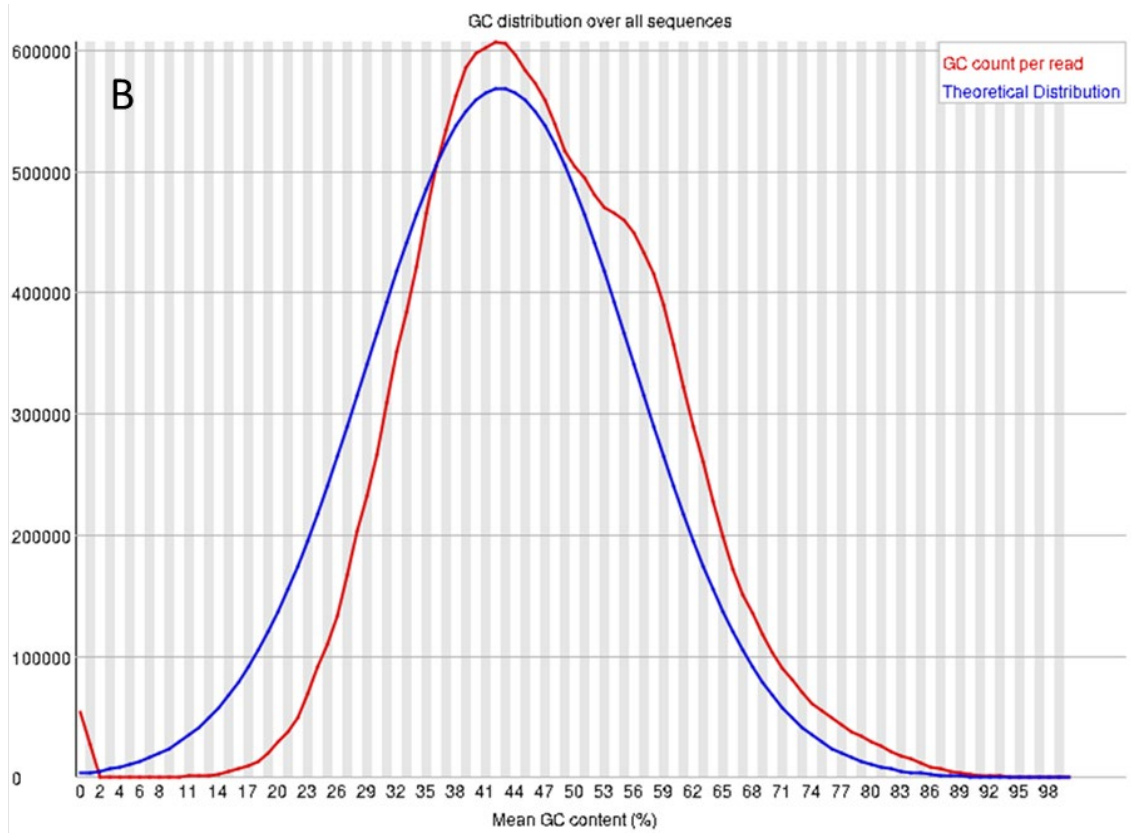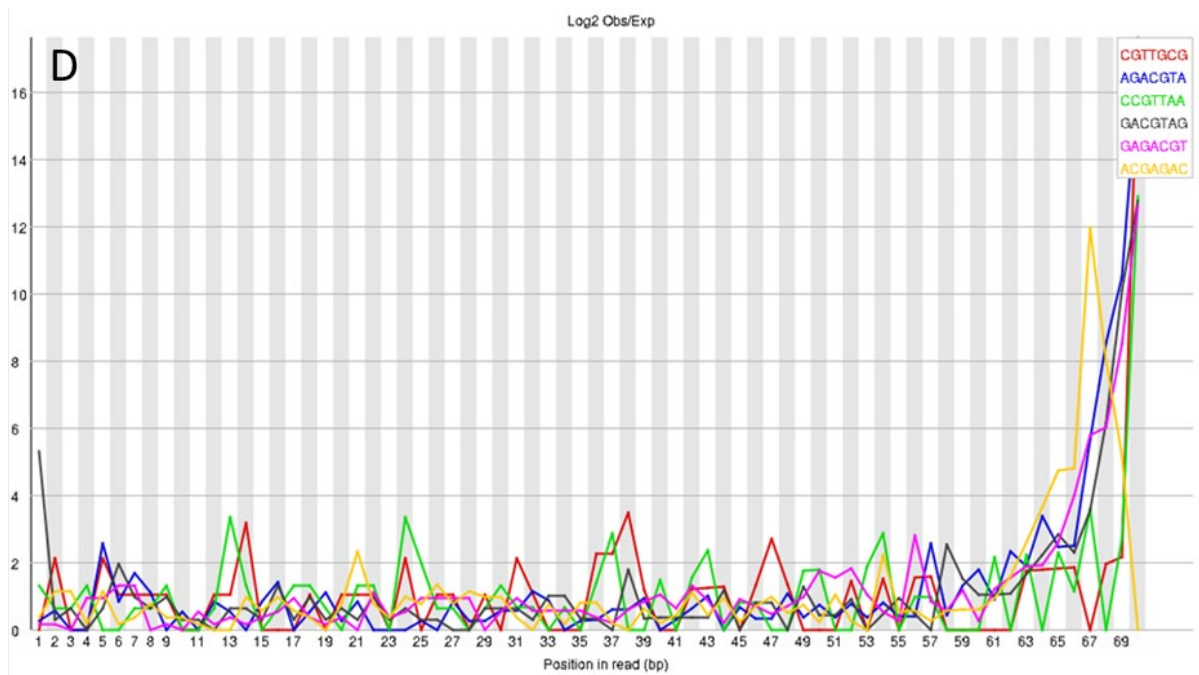
Figure 76. A characteristic example of a FastQC report of unstimulated HCAEC based on ATAC-seq data. A) per base sequence content plot out the proportion of each base position in a file for which each of the four normal DNA bases has been called. B) per sequence GC content plot shows the GC content across the whole length of each sequence in a file and compares it to a modelled normal distribution of GC content. C) per base N content plot shows the percentage of base calls at each position for which an N was called. D) kmer content plot shows the number of each 7-mer at each position in the library and then uses a binomial test to look for significant deviations from an even coverage at all positions.
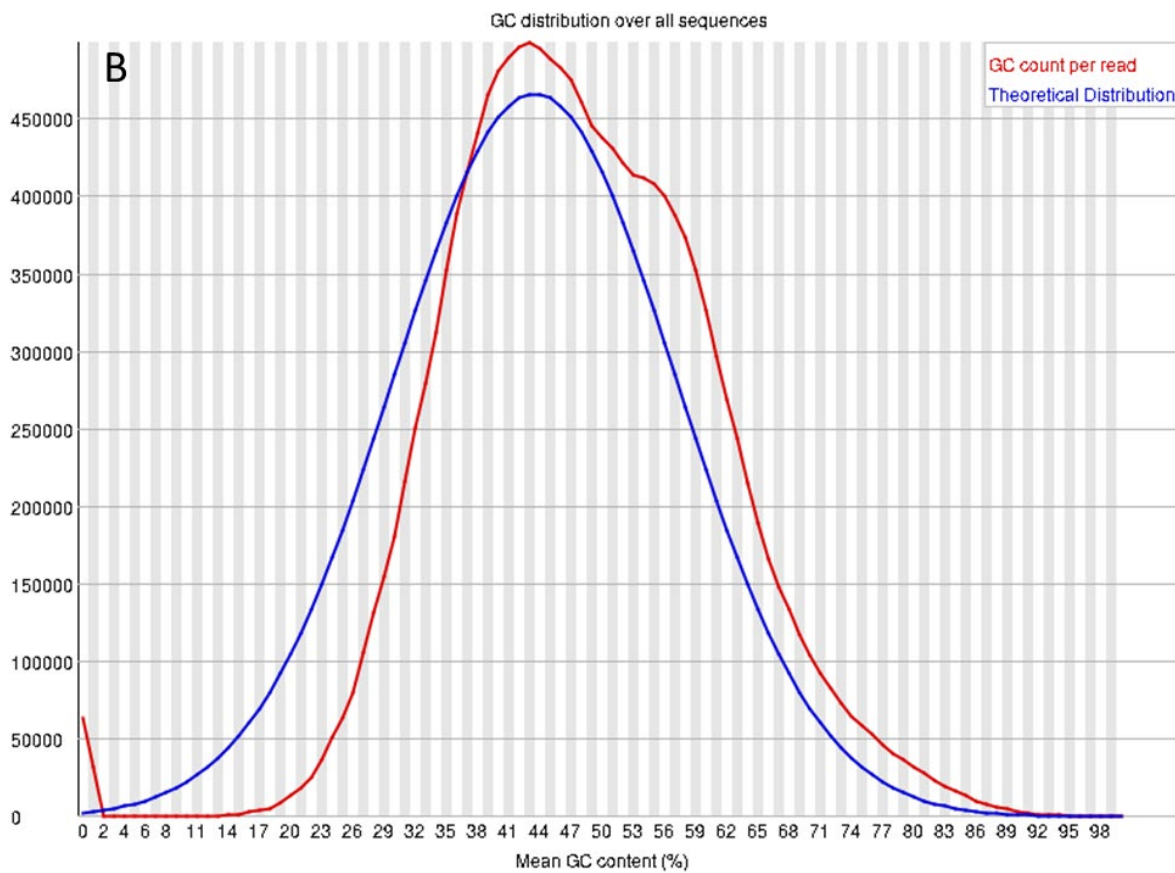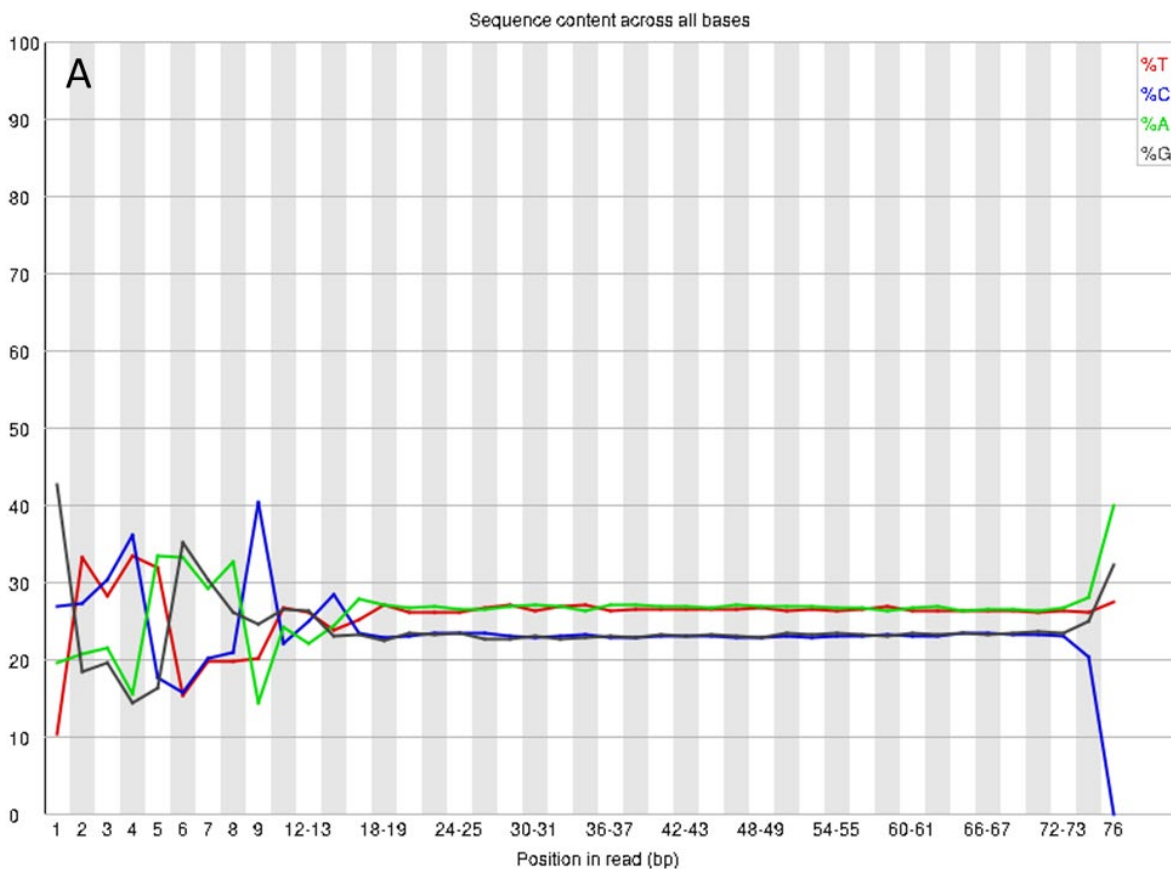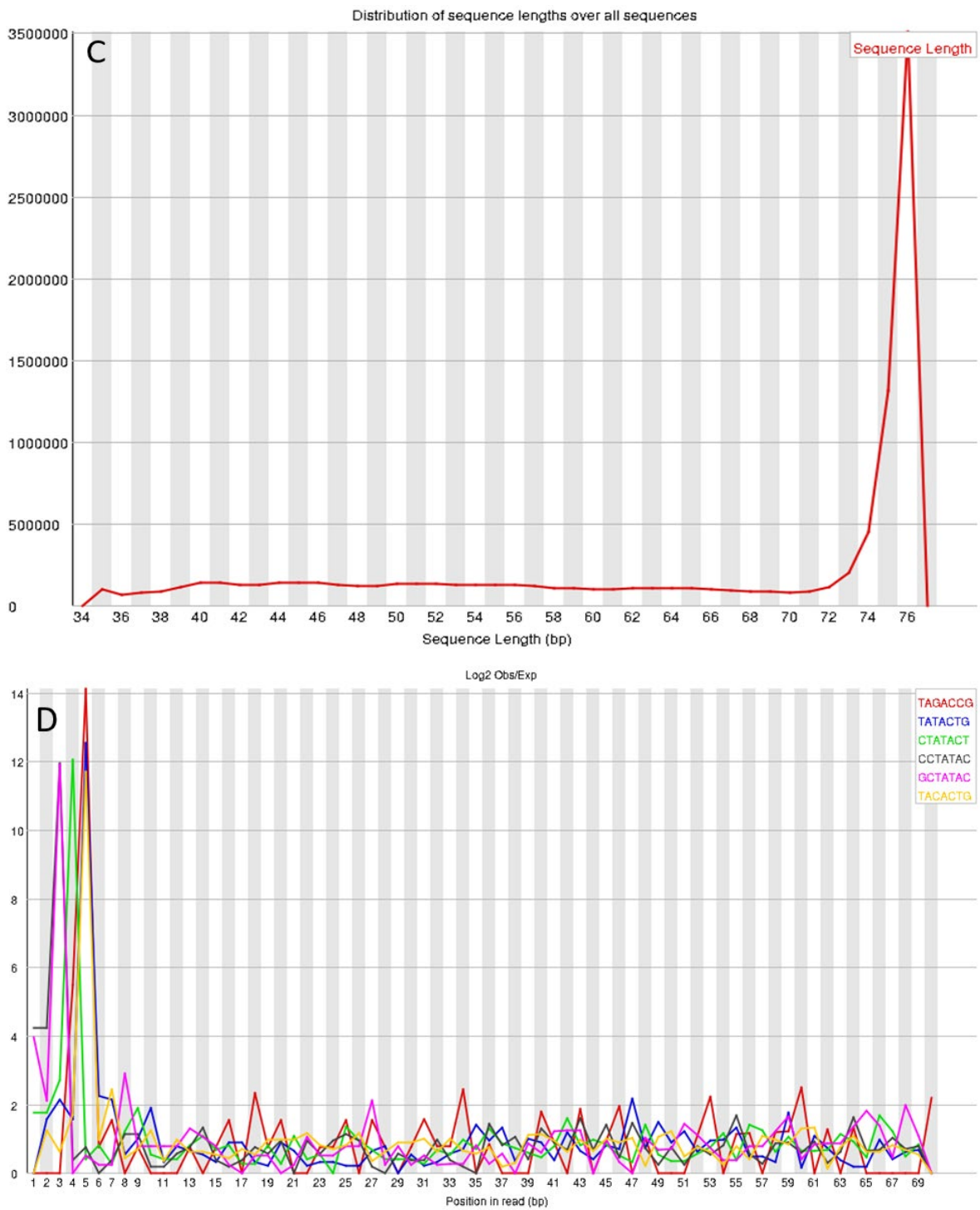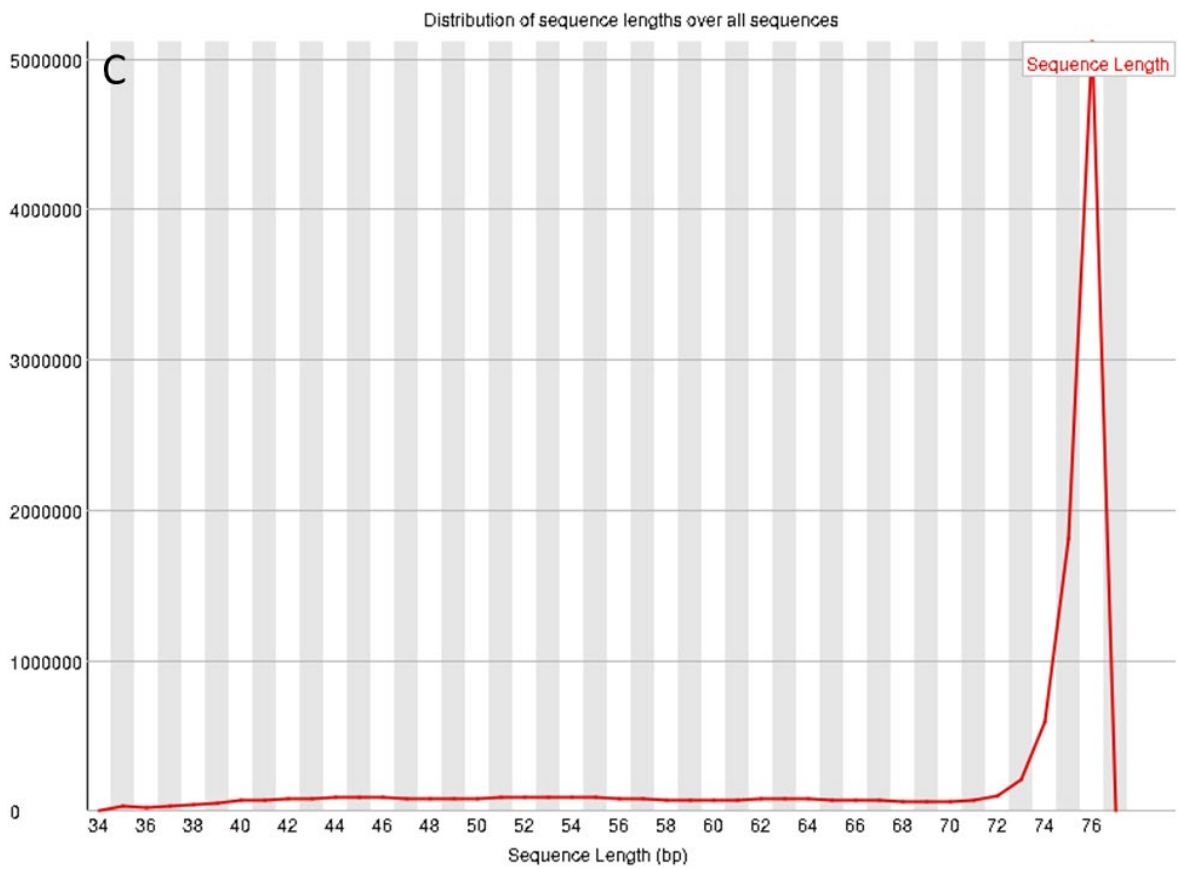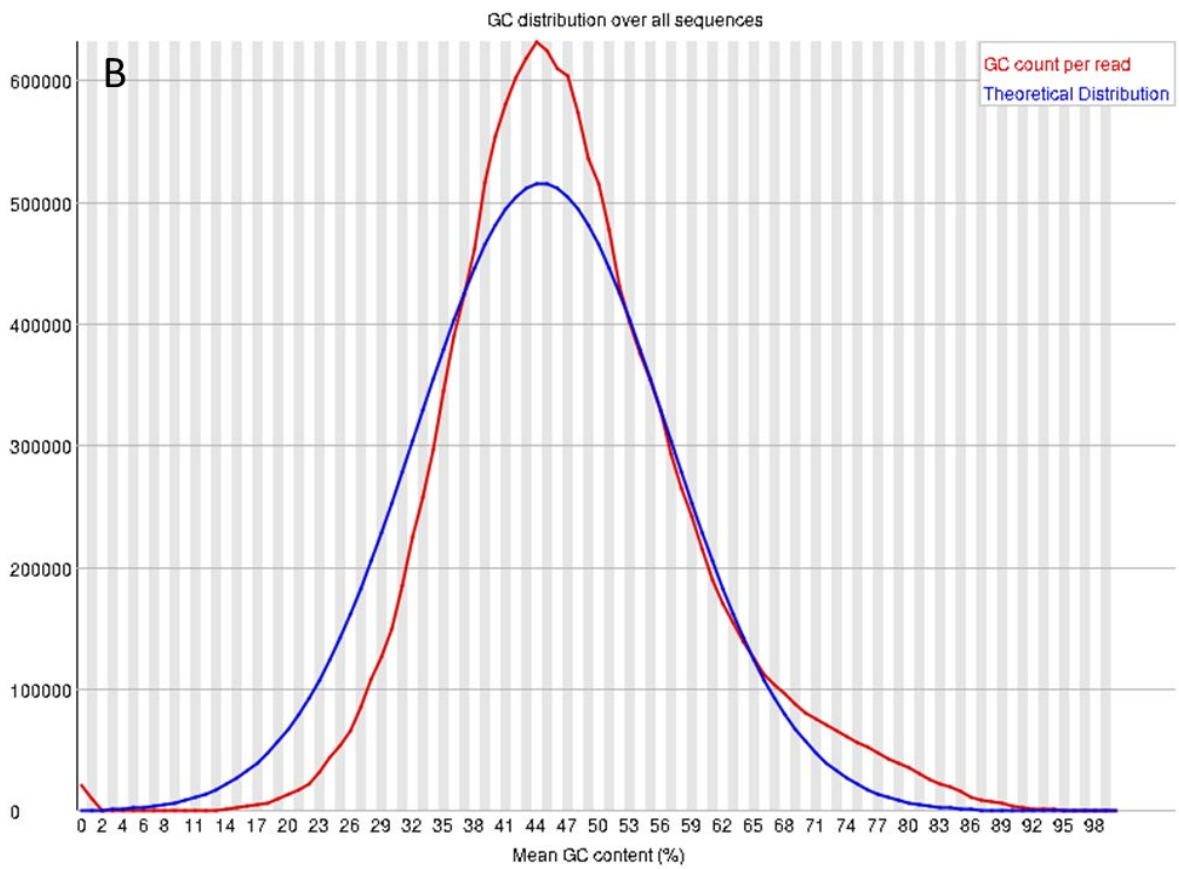
Figure 77. A characteristic example of a FastQC report of VEGFA-stimulated HCAEC based on ATAC-seq data. A) per base sequence content plot out the proportion of each base position in a file for which each of the four normal DNA bases has been called. B) per sequence GC content plot shows the GC content across the whole length of each sequence in a file and compares it to a modelled normal distribution of GC content. C) per base N content plot shows the percentage of base calls at each position for which an N was called. D) kmer content plot shows the number of each 7-mer at each position in the library and then uses a binomial test to look for significant deviations from an even coverage at all positions.
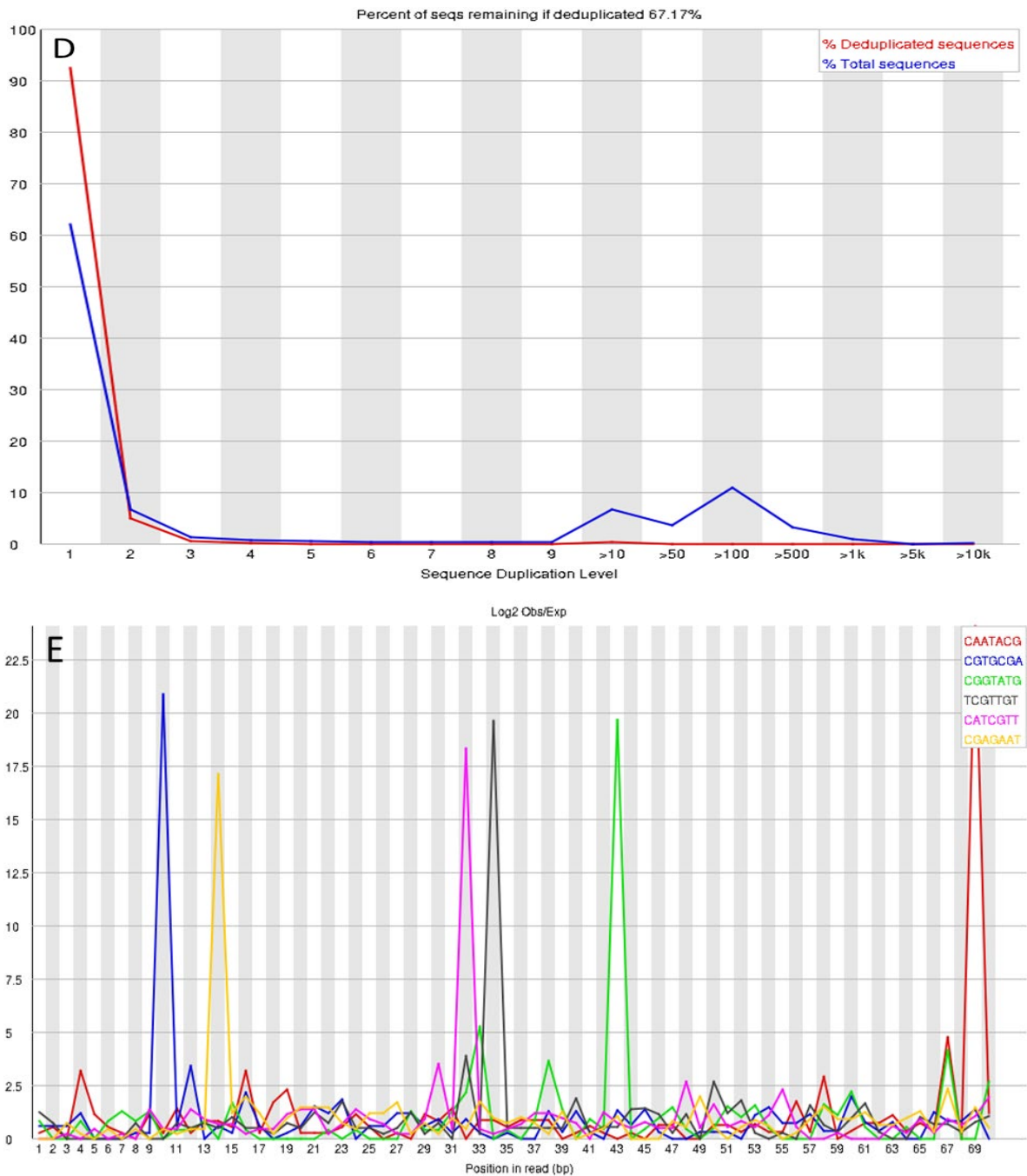
GC distribution over all sequences

Distribution of sequence lengths over all sequences

Figure 78. A characteristic example of a FastQC report of HCASMC based on ATAC-seq data. A) per base sequence content plot out the proportion of each base position in a file for which each of the four normal DNA bases has been called. B) per sequence GC content plot shows the GC content across the whole length of each sequence in a file and compares it to a modelled normal distribution of GC content. C) per base N content plot shows the percentage of base calls at each position for which an N was called. D) duplicate sequences plot shows the degree the degree of duplication for every sequence in a library and creates a plot showing the relative number of sequences with different degrees of duplication. E) kmer content plot shows the number of each 7-mer at each position in the library and then uses a binomial test to look for significant deviations from an even coverage at all positions.
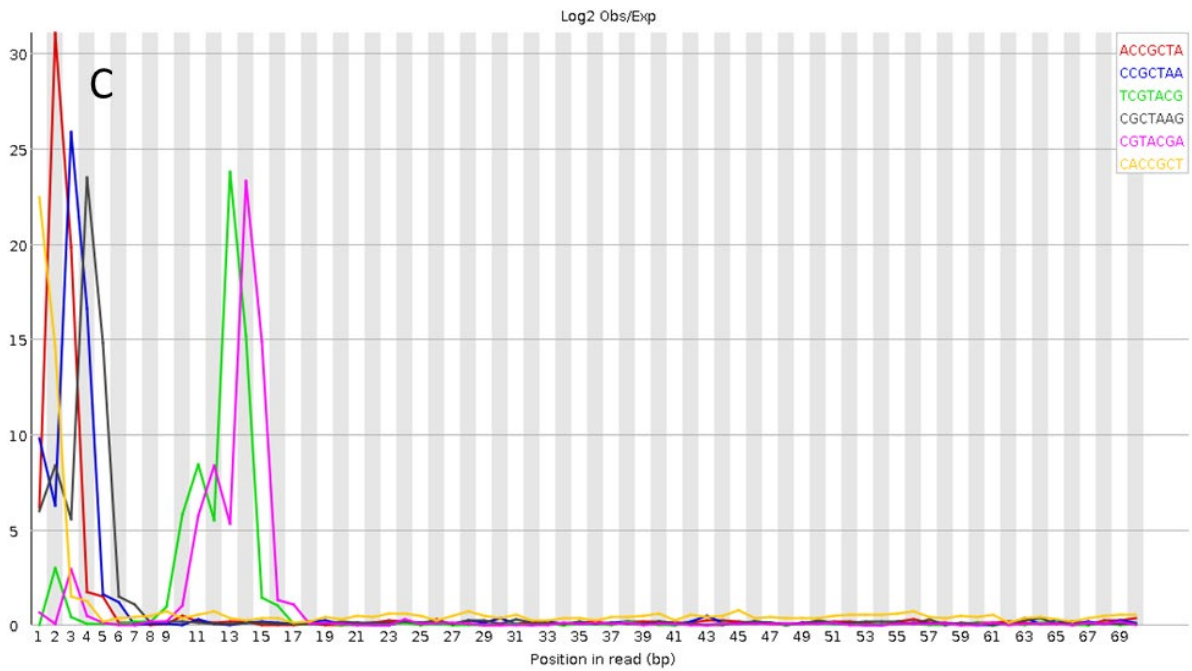
258

Sequence content across all bases

Percent of seqs remaining if deduplicated 36.98%
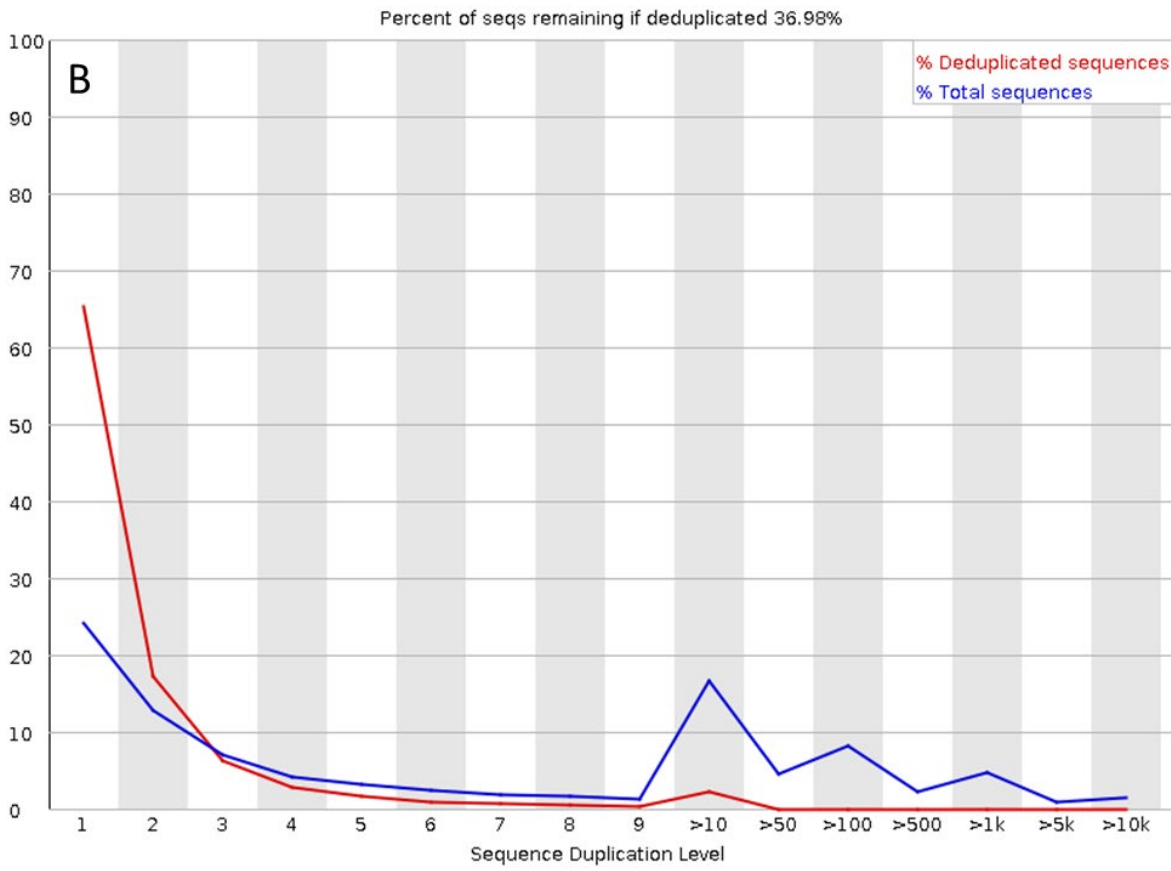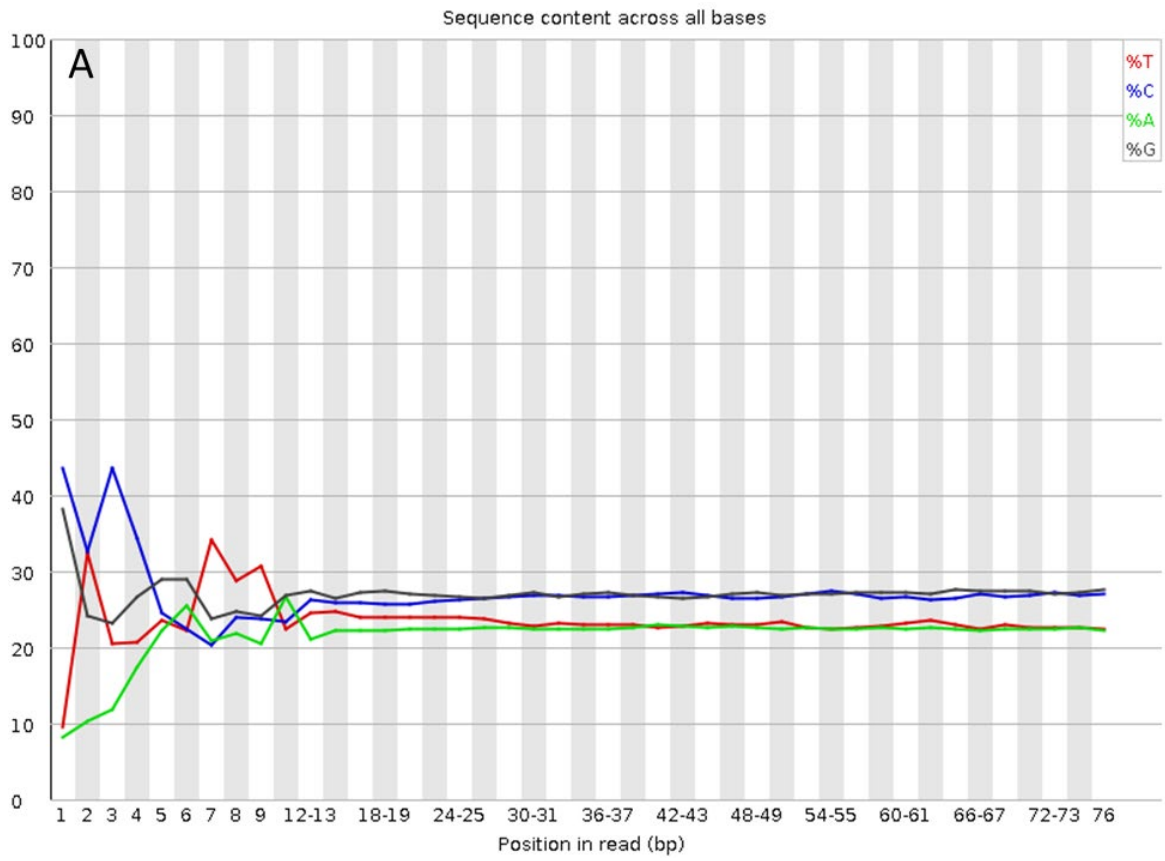
Figure 79. A characteristic example of a FastQC report of HCAEC based on RNA-seq data. A) per base sequence content plot out the proportion of each base position in a file for which each of the four normal DNA bases has been called. B) duplicate sequences plot shows the degree the degree of duplication for every sequence in a library and creates a plot showing the relative number of sequences with different degrees of duplication. C) kmer content plot shows the number of each 7-mer at each position in the library and then uses a binomial test to look for significant deviations from an even coverage at all positions.
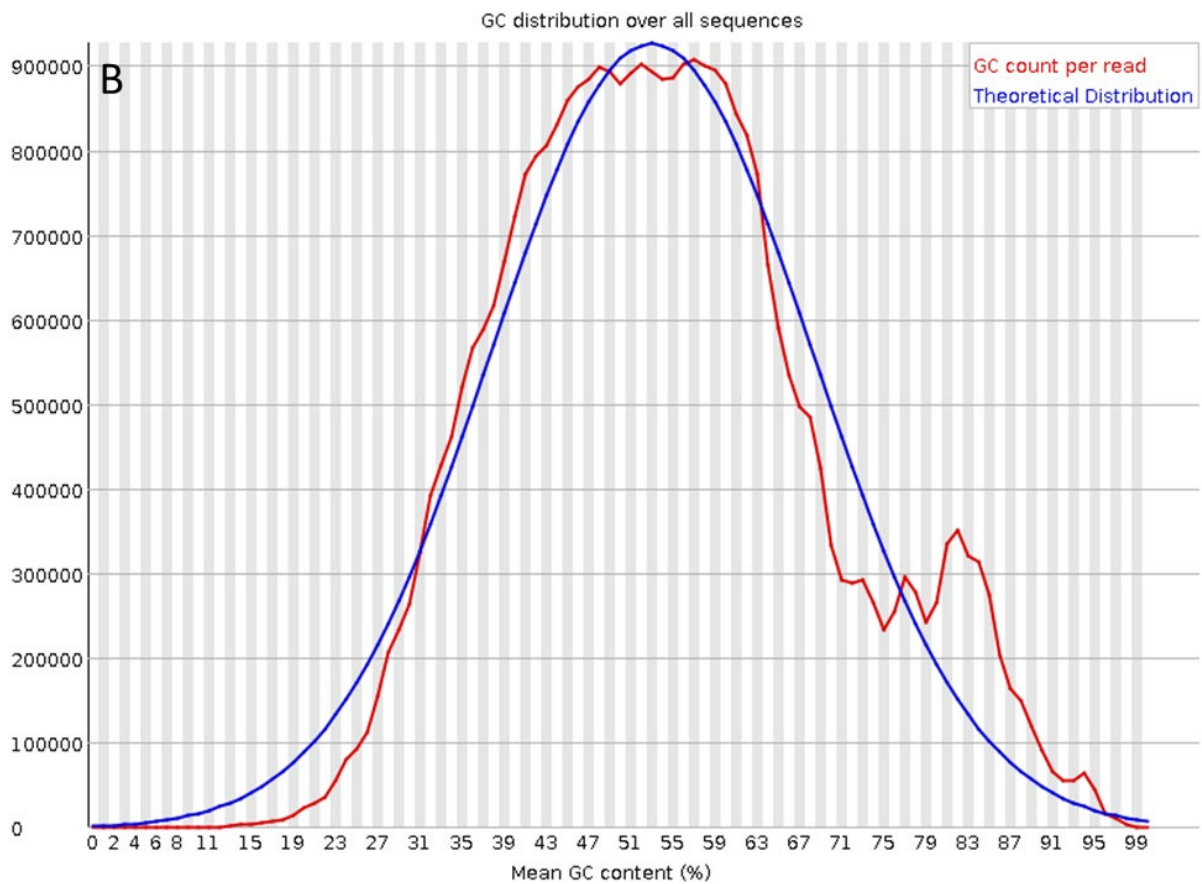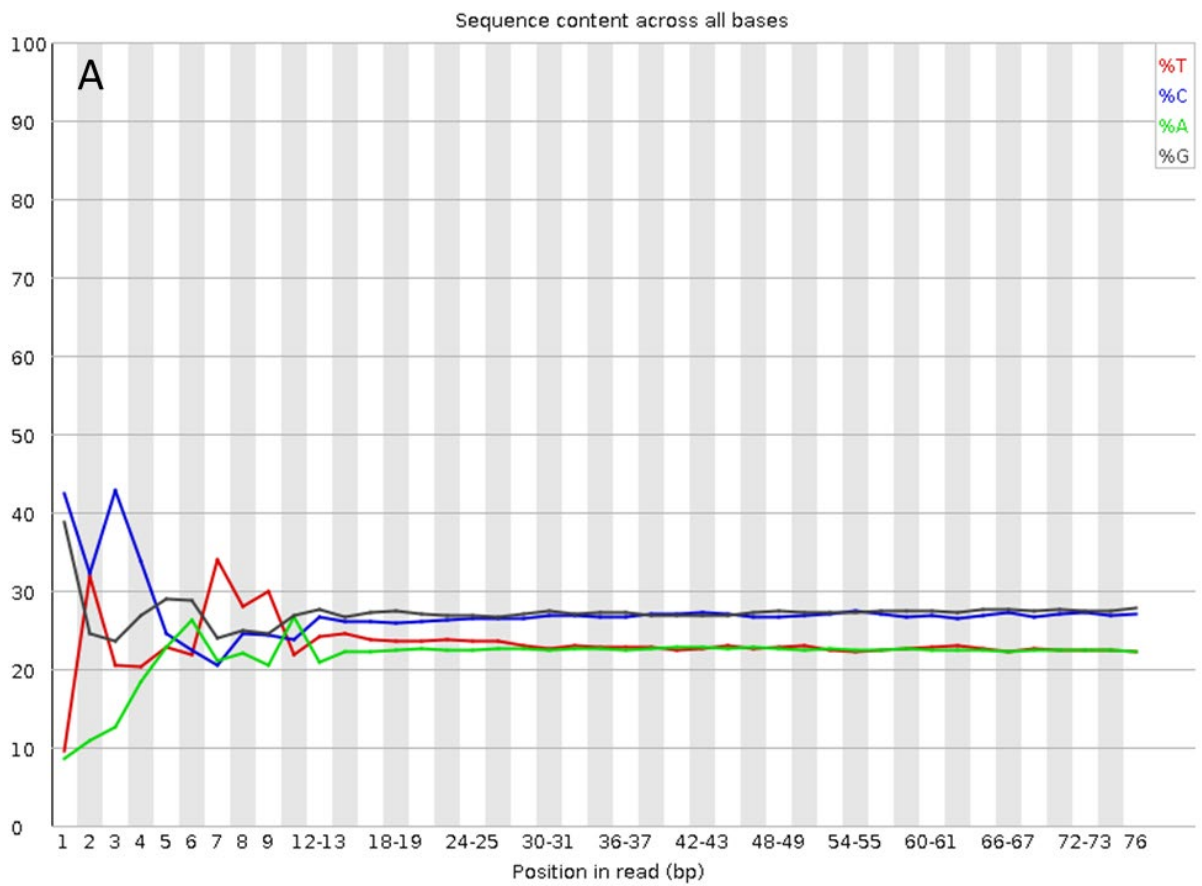
Sequence content across all bases

GC distribution over all sequences
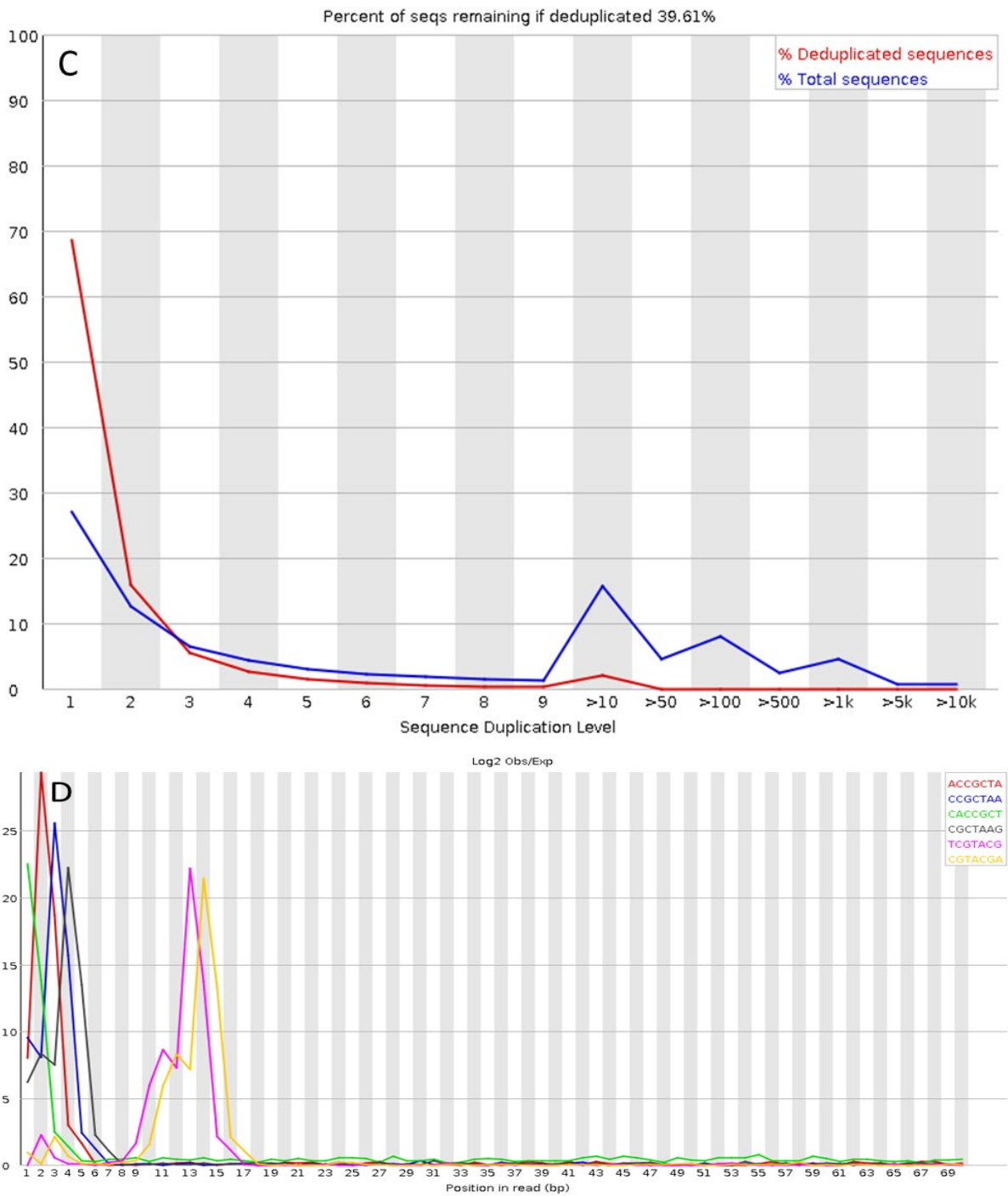
Figure 80. A characteristic example of a FastQC report of VEGFA-stimulated HCAEC based on RNA-seq data. A) per base sequence content plot out the proportion of each base position in a file for which each of the four normal DNA bases has been called. B) per sequence GC content plot shows the GC content across the whole length of each sequence in a file and compares it to a modelled normal distribution of GC content. C) duplicate sequences plot shows the degree the degree of duplication for every sequence in a library and creates a plot showing the relative number of sequences with different degrees of duplication. D) kmer content plot shows the number of each 7-mer at each position in the library and then uses a binomial test to look for significant deviations from an even coverage at all positions.

10. References

1. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases.* Science, 1996. **273**(5281): p. 1516-7.
2. Risch, N.J., *Searching for genetic determinants in the new millennium.* Nature, 2000. **405**(6788): p. 847-56.
3. Hardy, J. and A. Singleton, *Genomewide association studies and human disease.* N Engl J Med, 2009. **360**(17): p. 1759-68.
4. International HapMap, C., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-61.
5. Reich, D.E. and E.S. Lander, *On the allelic spectrum of human disease.* Trends Genet, 2001. **17**(9): p. 502-10.
6. Collins, F.S., M.S. Guyer, and A. Charkravarti, *Variations on a theme: cataloging human DNA sequence variation.* Science, 1997. **278**(5343): p. 1580-1.
7. Tennessen, J.A., et al., *Evolution and functional impact of rare coding variation from deep sequencing of human exomes.* Science, 2012. **337**(6090): p. 64-9.
8. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
9. Drmanac, R., et al., *Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays.* Science, 2010. **327**(5961): p. 78-81.
10. Mills, R.E., et al., *Mapping copy number variation by population-scale genome sequencing.* Nature, 2011. **470**(7332): p. 59-65.
11. Nelson, M.R., et al., *An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people.* Science, 2012. **337**(6090): p. 100-4.
12. Mathieson, I. and G. McVean, *Differential confounding of rare and common variants in spatially structured populations.* Nat Genet, 2012. **44**(3): p. 243-6.
13. Gravel, S., et al., *Demographic history and rare allele sharing among human populations.* Proc Natl Acad Sci U S A, 2011. **108**(29): p. 11983-8.
14. Wellcome Trust Case Control, C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
15. Easton, D.F., et al., *Genome-wide association study identifies novel breast cancer susceptibility loci.* Nature, 2007. **447**(7148): p. 1087-93.
16. Hunter, D.J., et al., *A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.* Nat Genet, 2007. **39**(7): p. 870-4.
17. Stacey, S.N., et al., *Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer.* Nat Genet, 2007. **39**(7): p. 865-9.
18. Yeager, M., et al., *Genome-wide association study of prostate cancer identifies a second risk locus at 8q24.* Nat Genet, 2007. **39**(5): p. 645-9.
19. Gudmundsson, J., et al., *Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes.* Nat Genet, 2007. **39**(8): p. 977-83.
20. Thomas, G., et al., *Multiple loci identified in a genome-wide association study of prostate cancer.* Nat Genet, 2008. **40**(3): p. 310-5.
21. Gudmundsson, J., et al., *Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer.* Nat Genet, 2008. **40**(3): p. 281-3.
22. Eeles, R.A., et al., *Multiple newly identified loci associated with prostate cancer susceptibility.* Nat Genet, 2008. **40**(3): p. 316-21.
23. Parkes, M., et al., *Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility.* Nat Genet, 2007. **39**(7): p. 830-2.
24. Duerr, R.H., et al., *A genome-wide association study identifies IL23R as an inflammatory bowel disease gene.* Science, 2006. **314**(5804): p. 1461-3.

25. Rioux, J.D., et al., *Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis.* Nat Genet, 2007. **39**(5): p. 596-604.

26. Libioulle, C., et al., *Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4.* PLoS Genet, 2007. **3**(4): p. e58.

27. Hampe, J., et al., *A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1.* Nat Genet, 2007. **39**(2): p. 207-11.

28. Todd, J.A., et al., *Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes.* Nat Genet, 2007. **39**(7): p. 857-64.

29. Hakonarson, H., et al., *A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene.* Nature, 2007. **448**(7153): p. 591-4.

30. Sladek, R., et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. **445**(7130): p. 881-5.

31. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. **316**(5829): p. 1336-41.

32. Scott, L.J., et al., *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* Science, 2007. **316**(5829): p. 1341-5.

33. Diabetes Genetics Initiative of Broad Institute of, H., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.

34. Steinthorsdottir, V., et al., *A variant in CDKAL1 influences insulin response and risk of type 2 diabetes.* Nat Genet, 2007. **39**(6): p. 770-5.

35. Zeggini, E., et al., *Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.* Nat Genet, 2008. **40**(5): p. 638-45.

36. Scuteri, A., et al., *Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits.* PLoS Genet, 2007. **3**(7): p. e115.

37. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. **316**(5826): p. 889-94.

38. Loos, R.J., et al., *Common variants near MC4R are associated with fat mass, weight and risk of obesity.* Nat Genet, 2008. **40**(6): p. 768-75.

39. Weedon, M.N., et al., *A common variant of HMGA2 is associated with adult and childhood height in the general population.* Nat Genet, 2007. **39**(10): p. 1245-50.

40. Sanna, S., et al., *Common variants in the GDF5-UQCC region are associated with variation in human height.* Nat Genet, 2008. **40**(2): p. 198-203.

41. Willer, C.J., et al., *Newly identified loci that influence lipid concentrations and risk of coronary artery disease.* Nat Genet, 2008. **40**(2): p. 161-9.

42. Kathiresan, S., et al., *Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans.* Nat Genet, 2008. **40**(2): p. 189-97.

43. Kooner, J.S., et al., *Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides.* Nat Genet, 2008. **40**(2): p. 149-51.

44. Consortium, E.P., *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-40.

45. Consortium, E.P., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.

46. Consortium, E.P., *A user's guide to the encyclopedia of DNA elements (ENCODE).* PLoS Biol, 2011. **9**(4): p. e1001046.

47. Mouse Genome Sequencing, C., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

48. Chiaromonte, F., et al., *The share of human genomic DNA under selection estimated from human-mouse genomic alignments.* Cold Spring Harb Symp Quant Biol, 2003. **68**: p. 245-54.

49. Cooper, G.M., et al., *Distribution and intensity of constraint in mammalian genomic sequence.* Genome Res, 2005. **15**(7): p. 901-13.

50. Parker, S.C., et al., *Local DNA topography correlates with functional noncoding regions of the human genome.* Science, 2009. **324**(5925): p. 389-92.

51. Lindblad-Toh, K., et al., *A high-resolution map of human evolutionary constraint using 29 mammals.* Nature, 2011. **478**(7370): p. 476-82.

52. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

53. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.

54. Tremethick, D.J., *Higher-order structures of chromatin: the elusive 30 nm fiber.* Cell, 2007. **128**(4): p. 651-4.

55. Woodcock, C.L. and S. Dimitrov, *Higher-order structure of chromatin and chromosomes.* Curr Opin Genet Dev, 2001. **11**(2): p. 130-5.

56. Tan, S. and C.A. Davey, *Nucleosome structural studies.* Curr Opin Struct Biol, 2011. **21**(1): p. 128-36.

57. Killian, J.L., et al., *Recent advances in single molecule studies of nucleosomes.* Curr Opin Struct Biol, 2012. **22**(1): p. 80-7.

58. Jiang, C. and B.F. Pugh, *Nucleosome positioning and gene regulation: advances through genomics.* Nat Rev Genet, 2009. **10**(3): p. 161-72.

59. Kaplan, N., et al., *Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology.* Genome Biol, 2010. **11**(11): p. 140.

60. Krebs, A.R., et al., *Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters.* Mol Cell, 2017. **67**(3): p. 411-422 e4.

61. Sung, M.H., et al., *DNase footprint signatures are dictated by factor dynamics and DNA sequence.* Mol Cell, 2014. **56**(2): p. 275-285.

62. Deal, R.B., J.G. Henikoff, and S. Henikoff, *Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones.* Science, 2010. **328**(5982): p. 1161-4.

63. John, S., et al., *Chromatin accessibility pre-determines glucocorticoid receptor binding patterns.* Nat Genet, 2011. **43**(3): p. 264-8.

64. John, S., et al., *Interaction of the glucocorticoid receptor with the chromatin landscape.* Mol Cell, 2008. **29**(5): p. 611-24.

65. He, H.H., et al., *Nucleosome dynamics define transcriptional enhancers.* Nat Genet, 2010. **42**(4): p. 343-7.

66. El Gazzar, M., et al., *Dynamic and selective nucleosome repositioning during endotoxin tolerance.* J Biol Chem, 2010. **285**(2): p. 1259-71.

67. Schones, D.E., et al., *Dynamic regulation of nucleosome positioning in the human genome.* Cell, 2008. **132**(5): p. 887-98.

68. Lever, M.A., et al., *Rapid exchange of histone H1.1 on chromatin in living human cells.* Nature, 2000. **408**(6814): p. 873-6.

69. Almer, A. and W. Horz, *Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast.* EMBO J, 1986. **5**(10): p. 2681-7.

70. Gilchrist, D.A., et al., *Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation.* Cell, 2010. **143**(4): p. 540-51.

71. Schep, A.N., et al., *Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions.* Genome Res, 2015. **25**(11): p. 1757-70.

72. Segal, E., et al., *A genomic code for nucleosome positioning.* Nature, 2006. **442**(7104): p. 772-8.

73. Kornberg, R., *The location of nucleosomes in chromatin: specific or statistical.* Nature, 1981. **292**(5824): p. 579-80.

74. Mavrich, T.N., et al., *A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome.* Genome Res, 2008. **18**(7): p. 1073-83.

75. Kaplan, N., et al., *The DNA-encoded nucleosome organization of a eukaryotic genome.* Nature, 2009. **458**(7236): p. 362-6.

76. Riposo, J. and J. Mozziconacci, *Nucleosome positioning and nucleosome stacking: two faces of the same coin.* Mol Biosyst, 2012. **8**(4): p. 1172-8.

77. Moyle-Heyrman, G., et al., *Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning.* Proc Natl Acad Sci U S A, 2013. **110**(50): p. 20158-63.

78. Muller, O., et al., *Changing chromatin fiber conformation by nucleosome repositioning.* Biophys J, 2014. **107**(9): p. 2141-50.

79. Valouev, A., et al., *Determinants of nucleosome organization in primary human cells.* Nature, 2011. **474**(7352): p. 516-20.

80. Raveh-Sadka, T., et al., *Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast.* Nat Genet, 2012. **44**(7): p. 743-50.

81. Valouev, A., et al., *A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning.* Genome Res, 2008. **18**(7): p. 1051-63.

82. Bednar, J., et al., *Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin.* Proc Natl Acad Sci U S A, 1998. **95**(24): p. 14173-8.

83. Fan, Y., et al., *Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation.* Cell, 2005. **123**(7): p. 1199-212.

84. Thoma, F., T. Koller, and A. Klug, *Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin.* J Cell Biol, 1979. **83**(2 Pt 1): p. 403-27.

85. Torres, C.M., et al., *The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity.* Science, 2016. **353**(6307).

86. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation.* Curr Opin Genet Dev, 2011. **21**(2): p. 175-86.

87. Robinson, P.J., et al., *30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction.* J Mol Biol, 2008. **381**(4): p. 816-25.

88. Wallrath, L.L. and S.C. Elgin, *Position effect variegation in Drosophila is associated with an altered chromatin structure.* Genes Dev, 1995. **9**(10): p. 1263-77.

89. Risca, V.I., et al., *Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping.* Nature, 2017. **541**(7636): p. 237-241.

90. Schalch, T., et al., *X-ray structure of a tetranucleosome and its implications for the chromatin fibre.* Nature, 2005. **436**(7047): p. 138-41.

91. Song, F., et al., *Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units.* Science, 2014. **344**(6182): p. 376-80.

92. Hsieh, T.H., et al., *Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C.* Cell, 2015. **162**(1): p. 108-19.

93. Ricci, M.A., et al., *Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo.* Cell, 2015. **160**(6): p. 1145-58.

94. Ou, H.D., et al., *ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells.* Science, 2017. **357**(6349).

95.   Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.

96.   Hewish, D.R. and L.A. Burgoyne, *Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease.* Biochem Biophys Res Commun, 1973. **52**(2): p. 504-10.

97.   Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science, 1974. **184**(4139): p. 868-71.

98.   Wu, C., Y.C. Wong, and S.C. Elgin, *The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity.* Cell, 1979. **16**(4): p. 807-14.

99.   Mueller, P.R. and B. Wold, *In vivo footprinting of a muscle specific enhancer by ligation mediated PCR.* Science, 1989. **246**(4931): p. 780-6.

100.   Rao, S., E. Procko, and M.F. Shannon, *Chromatin remodeling, measured by a novel real-time polymerase chain reaction assay, across the proximal promoter region of the IL-2 gene.* J Immunol, 2001. **167**(8): p. 4494-503.

101.   Lohr, D., R.T. Kovacic, and K.E. Van Holde, *Quantitative analysis of the digestion of yeast chromatin by staphylococcal nuclease.* Biochemistry, 1977. **16**(3): p. 463-71.

102.   Rizzo, J.M. and S. Sinha, *Analyzing the global chromatin structure of keratinocytes by MNase-seq.* Methods Mol Biol, 2014. **1195**: p. 49-59.

103.   Zhang, Z. and B.F. Pugh, *High-resolution genome-wide mapping of the primary structure of chromatin.* Cell, 2011. **144**(2): p. 175-86.

104.   Gaffney, D.J., et al., *Controls of nucleosome positioning in the human genome.* PLoS Genet, 2012. **8**(11): p. e1003036.

105.   Cui, K. and K. Zhao, *Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq.* Methods Mol Biol, 2012. **833**: p. 413-9.

106.   Weiner, A., et al., *High-resolution nucleosome mapping reveals transcription-dependent promoter packaging.* Genome Res, 2010. **20**(1): p. 90-100.

107.   Rizzo, J.M., J.E. Bard, and M.J. Buck, *Standardized collection of MNase-seq experiments enables unbiased dataset comparisons.* BMC Mol Biol, 2012. **13**: p. 15.

108.   Wallrath, L.L., et al., *Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures.* Bioessays, 1994. **16**(3): p. 165-70.

109.   Wu, C., *The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I.* Nature, 1980. **286**(5776): p. 854-60.

110.   Stalder, J., et al., *Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I.* Cell, 1980. **20**(2): p. 451-60.

111.   Henikoff, S., *Nucleosome destabilization in the epigenetic regulation of gene expression.* Nat Rev Genet, 2008. **9**(1): p. 15-26.

112.   Felsenfeld, G. and M. Groudine, *Controlling the double helix.* Nature, 2003. **421**(6921): p. 448-53.

113.   Crawford, G.E., et al., *Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites.* Proc Natl Acad Sci U S A, 2004. **101**(4): p. 992-7.

114.   Sabo, P.J., et al., *Discovery of functional noncoding elements by digital analysis of chromatin structure.* Proc Natl Acad Sci U S A, 2004. **101**(48): p. 16837-42.

115.   Dorschner, M.O., et al., *High-throughput localization of functional elements by quantitative chromatin profiling.* Nat Methods, 2004. **1**(3): p. 219-25.

116.   Song, L. and G.E. Crawford, *DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells.* Cold Spring Harb Protoc, 2010. **2010**(2): p. pdb prot5384.

117.   John, S., et al., *Genome-scale mapping of DNase I hypersensitivity.* Curr Protoc Mol Biol, 2013. **Chapter 27**: p. Unit 21 27.

118. Hesselberth, J.R., et al., *Global mapping of protein-DNA interactions in vivo by digital genomic footprinting.* Nat Methods, 2009. **6**(4): p. 283-9.

119. Boyle, A.P., et al., *High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells.* Genome Res, 2011. **21**(3): p. 456-64.

120. Neph, S., et al., *An expansive human regulatory lexicon encoded in transcription factor footprints.* Nature, 2012. **489**(7414): p. 83-90.

121. He, H.H., et al., *Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification.* Nat Methods, 2014. **11**(1): p. 73-78.

122. Giresi, P.G., et al., *FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin.* Genome Res, 2007. **17**(6): p. 877-85.

123. Simon, J.M., et al., *A detailed protocol for formaldehyde-assisted isolation of regulatory elements (FAIRE).* Curr Protoc Mol Biol, 2013. **Chapter 21**: p. Unit21 26.

124. Nagy, P.L., et al., *Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin.* Proc Natl Acad Sci U S A, 2003. **100**(11): p. 6364-9.

125. Gaulton, K.J., et al., *A map of open chromatin in human pancreatic islets.* Nat Genet, 2010. **42**(3): p. 255-9.

126. Ponts, N., et al., *Nucleosome landscape and control of transcription in the human malaria parasite.* Genome Res, 2010. **20**(2): p. 228-38.

127. Omidbakhshfard, M.A., et al., *A step-by-step protocol for formaldehyde-assisted isolation of regulatory elements from Arabidopsis thaliana.* J Integr Plant Biol, 2014. **56**(6): p. 527-38.

128. Buck, M.J., et al., *Alterations in chromatin accessibility and DNA methylation in clear cell renal cell carcinoma.* Oncogene, 2014. **33**(41): p. 4961-5.

129. Eeckhoute, J., et al., *Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers.* Genome Res, 2009. **19**(3): p. 372-80.

130. Cockell, M., D. Rhodes, and A. Klug, *Location of the primary sites of micrococcal nuclease cleavage on the nucleosome core.* J Mol Biol, 1983. **170**(2): p. 423-46.

131. Dingwall, C., G.P. Lomonossoff, and R.A. Laskey, *High sequence specificity of micrococcal nuclease.* Nucleic Acids Res, 1981. **9**(12): p. 2659-73.

132. Horz, W. and W. Altenburger, *Sequence specific cleavage of DNA by micrococcal nuclease.* Nucleic Acids Res, 1981. **9**(12): p. 2643-58.

133. Reznikoff, W.S., *Transposon Tn5.* Annu Rev Genet, 2008. **42**: p. 269-86.

134. Haniford, D.B. and M.J. Ellis, *Transposons Tn10 and Tn5.* Microbiol Spectr, 2015. **3**(1): p. MDNA3-0002-2014.

135. Buenrostro, J.D., et al., *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.* Curr Protoc Mol Biol, 2015. **109**: p. 21 29 1-21 29 9.

136. Picelli, S., et al., *Tn5 transposase and tagmentation procedures for massively scaled sequencing projects.* Genome Res, 2014. **24**(12): p. 2033-40.

137. Davie, K., et al., *Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling.* PLoS Genet, 2015. **11**(2): p. e1004994.

138. Bell, O., et al., *Determinants and dynamics of genome accessibility.* Nat Rev Genet, 2011. **12**(8): p. 554-64.

139. Mardis, E.R., *Next-generation DNA sequencing methods.* Annu Rev Genomics Hum Genet, 2008. **9**: p. 387-402.

140. Miskimen, K.L.S., E.R. Chan, and J.L. Haines, *Assay for Transposase-Accessible Chromatin Using Sequencing (ATAC-seq) Data Analysis.* Curr Protoc Hum Genet, 2017. **92**: p. 20 4 1-20 4 13.

141. Kumasaka, N., A.J. Knights, and D.J. Gaffney, *Fine-mapping cellular QTLs with RASQUAL and ATAC-seq.* Nat Genet, 2016. **48**(2): p. 206-13.

142. Ackermann, A.M., et al., *Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes.* Mol Metab, 2016. **5**(3): p. 233-244.

143. Quillien, A., et al., *Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq.* Cell Rep, 2017. **20**(3): p. 709-720.

144. Pranzatelli, T.J.F., D.G. Michael, and J.A. Chiorini, *ATAC2GRN: optimized ATAC-seq and DNase1-seq pipelines for rapid and accurate genome regulatory network inference.* BMC Genomics, 2018. **19**(1): p. 563.

145. Mu, J.C., et al., *Fast and accurate read alignment for resequencing.* Bioinformatics, 2012. **28**(18): p. 2366-73.

146. Maston, G.A., S.K. Evans, and M.R. Green, *Transcriptional regulatory elements in the human genome.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 29-59.

147. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter.* Annu Rev Biochem, 2003. **72**: p. 449-79.

148. Lim, C.Y., et al., *The MTE, a new core promoter element for transcription by RNA polymerase II.* Genes Dev, 2004. **18**(13): p. 1606-17.

149. Lee, M.P., et al., *ATG deserts define a novel core promoter subclass.* Genome Res, 2005. **15**(9): p. 1189-97.

150. Chen, Z. and J.L. Manley, *Core promoter elements and TAFs contribute to the diversity of transcriptional activation in vertebrates.* Mol Cell Biol, 2003. **23**(20): p. 7350-62.

151. Muller, F. and L. Tora, *The multicoloured world of promoter recognition complexes.* EMBO J, 2004. **23**(1): p. 2-8.

152. Morris, J.R., et al., *Enhancer choice in cis and in trans in Drosophila melanogaster: role of the promoter.* Genetics, 2004. **167**(4): p. 1739-47.

153. McKnight, S.L. and R. Kingsbury, *Transcriptional control signals of a eukaryotic protein-coding gene.* Science, 1982. **217**(4557): p. 316-24.

154. Ioshikhes, I.P. and M.Q. Zhang, *Large-scale human promoter mapping using CpG islands.* Nat Genet, 2000. **26**(1): p. 61-3.

155. Gershenzon, N.I. and I.P. Ioshikhes, *Synergy of human Pol II core promoter elements revealed by statistical sequence analysis.* Bioinformatics, 2005. **21**(8): p. 1295-300.

156. Jones, P.L., et al., *Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription.* Nat Genet, 1998. **19**(2): p. 187-91.

157. Andersson, R., A. Sandelin, and C.G. Danko, *A unified architecture of transcriptional regulatory elements.* Trends Genet, 2015. **31**(8): p. 426-33.

158. Osorio, J., *Gene regulation: Landscape and mechanisms of transcription factor cooperativity.* Nat Rev Genet, 2016. **17**(1): p. 5.

159. Visel, A., et al., *ChIP-seq accurately predicts tissue-specific activity of enhancers.* Nature, 2009. **457**(7231): p. 854-8.

160. Heintzman, N.D., et al., *Histone modifications at human enhancers reflect global cell-type-specific gene expression.* Nature, 2009. **459**(7243): p. 108-12.

161. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.

162. Li, W., D. Notani, and M.G. Rosenfeld, *Enhancers as non-coding RNA transcription units: recent insights and future perspectives.* Nat Rev Genet, 2016. **17**(4): p. 207-23.

163. Hu, Z. and W.W. Tee, *Enhancers and chromatin structures: regulatory hubs in gene expression and diseases.* Biosci Rep, 2017. **37**(2).

164. Ogbourne, S. and T.M. Antalis, *Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes.* Biochem J, 1998. **331 (Pt 1)**: p. 1-14.

165. Harris, M.B., J. Mostecki, and P.B. Rothman, *Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function.* J Biol Chem, 2005. **280**(13): p. 13114-21.

166. Li, L., et al., *Gene regulation by Sp1 and Sp3.* Biochem Cell Biol, 2004. **82**(4): p. 460-71.

167. Chen, L. and J. Widom, *Mechanism of transcriptional silencing in yeast.* Cell, 2005. **120**(1): p. 37-48.

168. Recillas-Targa, F., et al., *Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6883-8.

169. Fourel, G., F. Magdinier, and E. Gilson, *Insulator dynamics and the setting of chromatin domains.* Bioessays, 2004. **26**(5): p. 523-32.

170. Banerjee, S., et al., *Igf2/H19 imprinting control region (ICR): an insulator or a position-dependent silencer?* ScientificWorldJournal, 2001. **1**: p. 218-24.

171. Bell, A.C. and G. Felsenfeld, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.* Nature, 2000. **405**(6785): p. 482-5.

172. West, A.G. and P. Fraser, *Remote control of gene transcription.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R101-11.

173. Defossez, P.A., et al., *The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso.* J Biol Chem, 2005. **280**(52): p. 43017-23.

174. Chakalova, L., et al., *Developmental regulation of the beta-globin gene locus.* Prog Mol Subcell Biol, 2005. **38**: p. 183-206.

175. Tanimoto, K., et al., *Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice.* Nature, 1999. **398**(6725): p. 344-8.

176. Bank, A., *Regulation of human fetal hemoglobin: new players, new complexities.* Blood, 2006. **107**(2): p. 435-43.

177. Tolhuis, B., et al., *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.

178. Spilianakis, C.G., et al., *Interchromosomal associations between alternatively expressed loci.* Nature, 2005. **435**(7042): p. 637-45.

179. Roger, V.L., *Epidemiology of myocardial infarction.* Med Clin North Am, 2007. **91**(4): p. 537-52; ix.

180. Watkins, H. and M. Farrall, *Genetic susceptibility to coronary artery disease: from promise to progress.* Nat Rev Genet, 2006. **7**(3): p. 163-73.

181. Go, A.S., et al., *Heart disease and stroke statistics--2014 update: a report from the American Heart Association.* Circulation, 2014. **129**(3): p. e28-e292.

182. Wong, N.D., *Epidemiological studies of CHD and the evolution of preventive cardiology.* Nat Rev Cardiol, 2014. **11**(5): p. 276-89.

183. *The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators.* J Clin Epidemiol, 1988. **41**(2): p. 105-14.

184. Yusuf, S., et al., *Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study.* Lancet, 2004. **364**(9438): p. 937-52.

185. Worth, R.M., et al., *Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: mortality.* Am J Epidemiol, 1975. **102**(6): p. 481-90.

186. Sekikawa, A., et al., *A "natural experiment" in cardiovascular epidemiology in the early 21st century.* Heart, 2003. **89**(3): p. 255-7.

187. *Infectious Disease/CDC Update: Update on emerging infections: news from the Centers for Disease Control and Prevention. Evaluation of 11 commercially available rapid influenza diagnostic tests-United States, 2011-2012.* Ann Emerg Med, 2013. **61**(5): p. 573-7.

188. Writing Group, M., et al., *Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association.* Circulation, 2016. **133**(4): p. e38-360.

189. Chadha, S.L., et al., *Epidemiological study of coronary heart disease in urban population of Delhi.* Indian J Med Res, 1990. **92**: p. 424-30.

190. Mulcahy, R., *The health benefits of smoking cessation.* Ir Med J, 1990. **83**(2): p. 45-6.

191. Stamler, J., et al., *Diabetes, other risk factors, and 12-yr cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial.* Diabetes Care, 1993. **16**(2): p. 434-44.

192. Verschuren, W.M., et al., *Serum total cholesterol and long-term coronary heart disease mortality in different cultures. Twenty-five-year follow-up of the seven countries study.* JAMA, 1995. **274**(2): p. 131-6.

193. Haffner, S.M., *Diabetes, hyperlipidemia, and coronary artery disease.* Am J Cardiol, 1999. **83**(9B): p. 17F-21F.

194. Kannel, W.B. and D.L. McGee, *Diabetes and cardiovascular disease. The Framingham study.* JAMA, 1979. **241**(19): p. 2035-8.

195. Assmann, G. and H. Schulte, *The Prospective Cardiovascular Munster (PROCAM) study: prevalence of hyperlipidemia in persons with hypertension and/or diabetes mellitus and the relationship to coronary heart disease.* Am Heart J, 1988. **116**(6 Pt 2): p. 1713-24.

196. Strauer, B.E., *The concept of coronary flow reserve.* J Cardiovasc Pharmacol, 1992. **19 Suppl 5**: p. S67-80.

197. DeFronzo, R.A. and E. Ferrannini, *Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease.* Diabetes Care, 1991. **14**(3): p. 173-94.

198. Matsuzawa, Y., et al., *Visceral fat accumulation and cardiovascular disease.* Obes Res, 1995. **3 Suppl 5**: p. 645S-647S.

199. Kumada, M., et al., *Association of hypoadiponectinemia with coronary artery disease in men.* Arterioscler Thromb Vasc Biol, 2003. **23**(1): p. 85-9.

200. Becker, B.F., *Towards the physiological function of uric acid.* Free Radic Biol Med, 1993. **14**(6): p. 615-31.

201. Dawson, J., T. Quinn, and M. Walters, *Uric acid reduction: a new paradigm in the management of cardiovascular risk?* Curr Med Chem, 2007. **14**(17): p. 1879-86.

202. Steptoe, A. and M. Kivimaki, *Stress and cardiovascular disease.* Nat Rev Cardiol, 2012. **9**(6): p. 360-70.

203. Brotman, D.J., S.H. Golden, and I.S. Wittstein, *The cardiovascular toll of stress.* Lancet, 2007. **370**(9592): p. 1089-100.

204. Kivimaki, M., et al., *Job strain and ischaemic disease: does the inclusion of older employees in the cohort dilute the association? The WOLF Stockholm Study.* J Epidemiol Community Health, 2008. **62**(4): p. 372-4.

205. Carmeliet, P., *Fibroblast growth factor-1 stimulates branching and survival of myocardial arteries: a goal for therapeutic angiogenesis?* Circ Res, 2000. **87**(3): p. 176-8.

206. Parisi, A.F., E.D. Folland, and P. Hartigan, *A comparison of angioplasty with medical therapy in the treatment of single-vessel coronary artery disease. Veterans Affairs ACME Investigators.* N Engl J Med, 1992. **326**(1): p. 10-6.

207. Serruys, P.W., M.J. Kutryk, and A.T. Ong, *Coronary-artery stents.* N Engl J Med, 2006. **354**(5): p. 483-95.

208. Bush, M.A., et al., *Pharmacokinetics and pharmacodynamics of recombinant FGF-2 in a phase I trial in coronary artery disease.* J Clin Pharmacol, 2001. **41**(4): p. 378-85.

209. Schrader, B.J. and S.I. Berk, *Antiplatelet agents in coronary artery disease.* Clin Pharm, 1990. **9**(2): p. 118-24.

210. Clappers, N., M.A. Brouwer, and F.W. Verheugt, *Antiplatelet treatment for coronary heart disease.* Heart, 2007. **93**(2): p. 258-65.

211. Quyyumi, A.A., et al., *Medical treatment of patients with severe exertional and rest angina: double blind comparison of beta blocker, calcium antagonist, and nitrate.* Br Heart J, 1987. **57**(6): p. 505-11.

212. Hwang, J.K., et al., *Association of beta-blocker therapy with long-term clinical outcomes in patients with coronary chronic total occlusion.* Medicine (Baltimore), 2016. **95**(30): p. e4300.

213. Tadamura, E., et al., *The effect of nitroglycerin on myocardial blood flow in various segments characterized by rest-redistribution thallium SPECT.* J Nucl Med, 2003. **44**(5): p. 745-51.

214. Akhras, F. and G. Jackson, *Efficacy of nifedipine and isosorbide mononitrate in combination with atenolol in stable angina.* Lancet, 1991. **338**(8774): p. 1036-9.

215. Marenberg, M.E., et al., *Genetic susceptibility to death from coronary heart disease in a study of twins.* N Engl J Med, 1994. **330**(15): p. 1041-6.

216. Morgan, T.M., et al., *Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study.* JAMA, 2007. **297**(14): p. 1551-61.

217. Samani, N.J., et al., *Genomewide association analysis of coronary artery disease.* N Engl J Med, 2007. **357**(5): p. 443-53.

218. Chen, Z., et al., *A common variant on chromosome 9p21 affects the risk of early-onset coronary artery disease.* Mol Biol Rep, 2009. **36**(5): p. 889-93.

219. McPherson, R., et al., *A common allele on chromosome 9 associated with coronary heart disease.* Science, 2007. **316**(5830): p. 1488-91.

220. Schunkert, H., et al., *Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease.* Circulation, 2008. **117**(13): p. 1675-84.

221. Helgadottir, A., et al., *The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm.* Nat Genet, 2008. **40**(2): p. 217-24.

222. O'Donnell, C.J., et al., *Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction.* Circulation, 2011. **124**(25): p. 2855-64.

223. Preuss, M., et al., *Design of the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Study: A Genome-wide association meta-analysis involving more than 22 000 cases and 60 000 controls.* Circ Cardiovasc Genet, 2010. **3**(5): p. 475-83.

224. Coronary Artery Disease Genetics, C., *A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease.* Nat Genet, 2011. **43**(4): p. 339-44.

225. Erdmann, J., et al., *A decade of genome-wide association studies for coronary artery disease: the challenges ahead.* Cardiovasc Res, 2018. **114**(9): p. 1241-1257.

226. Schunkert, H., et al., *Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease.* Nat Genet, 2011. **43**(4): p. 333-8.

227. Consortium, C.A.D., et al., *Large-scale association analysis identifies new risk loci for coronary artery disease.* Nat Genet, 2013. **45**(1): p. 25-33.

228. Nelson, C.P., et al., *Association analyses based on false discovery rate implicate new loci for coronary artery disease.* Nat Genet, 2017. **49**(9): p. 1385-1391.

229. Voight, B.F., et al., *The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits.* PLoS Genet, 2012. **8**(8): p. e1002793.

230. *Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease.* N Engl J Med, 2016. **374**(19): p. 1898.

231. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

232. Iglesias, A.I., et al., *Haplotype reference consortium panel: Practical implications of imputations with large reference panels.* Hum Mutat, 2017. **38**(8): p. 1025-1032.

233. Loh, P.R., et al., *Reference-based phasing using the Haplotype Reference Consortium panel.* Nat Genet, 2016. **48**(11): p. 1443-1448.

234. Klarin, D., et al., *Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease.* Nat Genet, 2017. **49**(9): p. 1392-1397.

235. Verweij, N., et al., *Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure.* Sci Rep, 2017. **7**(1): p. 2761.

236. van der Harst, P. and N. Verweij, *Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease.* Circ Res, 2018. **122**(3): p. 433-443.

237. Christofidou, P., et al., *Runs of Homozygosity: Association with Coronary Artery Disease and Gene Expression in Monocytes and Macrophages.* Am J Hum Genet, 2015. **97**(2): p. 228-37.

238. Loley, C., et al., *No Association of Coronary Artery Disease with X-Chromosomal Variants in Comprehensive International Meta-Analysis.* Sci Rep, 2016. **6**: p. 35278.

239. Nikpay, M., et al., *A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease.* Nat Genet, 2015. **47**(10): p. 1121-1130.

240. Myocardial Infarction, G., et al., *Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease.* N Engl J Med, 2016. **374**(12): p. 1134-44.

241. Webb, T.R., et al., *Systematic Evaluation of Pleiotropy Identifies 6 Further Loci Associated With Coronary Artery Disease.* J Am Coll Cardiol, 2017. **69**(7): p. 823-836.

242. Olsson, A.K., et al., *VEGF receptor signalling - in control of vascular function.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 359-71.

243. Muller, Y.A., et al., *Vascular endothelial growth factor: crystal structure and functional mapping of the kinase domain receptor binding site.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7192-7.

244. De Falco, S., B. Gigante, and M.G. Persico, *Structure and function of placental growth factor.* Trends Cardiovasc Med, 2002. **12**(6): p. 241-6.

245. Woolard, J., et al., *VEGF165b, an inhibitory vascular endothelial growth factor splice variant: mechanism of action, in vivo effect on angiogenesis and endogenous protein expression.* Cancer Res, 2004. **64**(21): p. 7822-35.

246. Lee, S., et al., *Processing of VEGF-A by matrix metalloproteinases regulates bioavailability and vascular patterning in tumors.* J Cell Biol, 2005. **169**(4): p. 681-91.

247. Christinger, H.W., et al., *The crystal structure of placental growth factor in complex with domain 2 of vascular endothelial growth factor receptor-1.* J Biol Chem, 2004. **279**(11): p. 10382-8.

248. Fuh, G., et al., *Requirements for binding and signalling of the kinase domain receptor for vascular endothelial growth factor.* J Biol Chem, 1998. **273**(18): p. 11197-204.

249. Kendall, R.L. and K.A. Thomas, *Inhibition of vascular endothelial cell growth factor activity by an endogenously encoded soluble receptor.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10705-9.

250. Ebos, J.M., et al., *A naturally occurring soluble form of vascular endothelial growth factor receptor 2 detected in mouse and human plasma.* Mol Cancer Res, 2004. **2**(6): p. 315-26.

251. Hughes, D.C., *Alternative splicing of the human VEGFGR-3/FLT4 gene as a consequence of an integrated human endogenous retrovirus.* J Mol Evol, 2001. **53**(2): p. 77-9.

252. Rahimi, N., V. Dayanir, and K. Lashkari, *Receptor chimeras indicate that the vascular endothelial growth factor receptor-1 (VEGFR-1) modulates mitogenic activity of VEGFR-2 in endothelial cells.* J Biol Chem, 2000. **275**(22): p. 16986-92.

253. Zeng, H., H.F. Dvorak, and D. Mukhopadhyay, *Vascular permeability factor (VPF)/vascular endothelial growth factor (VEGF) peceptor-1 down-modulates VPF/VEGF receptor-2-mediated endothelial cell proliferation, but not migration, through phosphatidylinositol 3-kinase-dependent pathways.* J Biol Chem, 2001. **276**(29): p. 26969-79.

254. Carmeliet, P., et al., *Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions.* Nat Med, 2001. **7**(5): p. 575-83.

255. Autiero, M., et al., *Role of PlGF in the intra- and intermolecular cross talk between the VEGF receptors Flt1 and Flk1.* Nat Med, 2003. **9**(7): p. 936-43.

256. Takahashi, T., et al., *A single autophosphorylation site on KDR/Flk-1 is essential for VEGF-A-dependent activation of PLC-gamma and DNA synthesis in vascular endothelial cells.* EMBO J, 2001. **20**(11): p. 2768-78.

257. Sakurai, Y., et al., *Essential role of Flk-1 (VEGF receptor 2) tyrosine residue 1173 in vasculogenesis in mice.* Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1076-81.

258. Holmqvist, K., et al., *The adaptor protein shb binds to tyrosine 1175 in vascular endothelial growth factor (VEGF) receptor-2 and regulates VEGF-dependent cellular migration.* J Biol Chem, 2004. **279**(21): p. 22267-75.

259. Fujio, Y. and K. Walsh, *Akt mediates cytoprotection of endothelial cells by vascular endothelial growth factor in an anchorage-dependent manner.* J Biol Chem, 1999. **274**(23): p. 16349-54.

260. Warner, A.J., et al., *The Shc-related adaptor protein, Sck, forms a complex with the vascular-endothelial-growth-factor receptor KDR in transfected cells.* Biochem J, 2000. **347**(Pt 2): p. 501-9.

261. Sakai, R., et al., *The mammalian ShcB and ShcC phosphotyrosine docking proteins function in the maturation of sensory and sympathetic neurons.* Neuron, 2000. **28**(3): p. 819-33.

262. Shu, X., et al., *Sphingosine kinase mediates vascular endothelial growth factor-induced activation of ras and mitogen-activated protein kinases.* Mol Cell Biol, 2002. **22**(22): p. 7758-68.

263. Meadows, K.N., P. Bryant, and K. Pumiglia, *Vascular endothelial growth factor induction of the angiogenic phenotype requires Ras activation.* J Biol Chem, 2001. **276**(52): p. 49289-98.

264. Takahashi, T., H. Ueno, and M. Shibuya, *VEGF activates protein kinase C-dependent, but Ras-independent Raf-MEK-MAP kinase pathway for DNA synthesis in primary endothelial cells.* Oncogene, 1999. **18**(13): p. 2221-30.

265. Kroll, J. and J. Waltenberger, *The vascular endothelial growth factor receptor KDR activates multiple signal transduction pathways in porcine aortic endothelial cells.* J Biol Chem, 1997. **272**(51): p. 32521-7.

266. Matsumoto, T., et al., *VEGF receptor-2 Y951 signalling and a role for the adapter molecule TSAd in tumor angiogenesis.* EMBO J, 2005. **24**(13): p. 2342-53.

267. Zeng, H., S. Sanyal, and D. Mukhopadhyay, *Tyrosine residues 951 and 1059 of vascular endothelial growth factor receptor-2 (KDR) are essential for vascular permeability factor/vascular endothelial growth factor-induced endothelium migration and proliferation, respectively.* J Biol Chem, 2001. **276**(35): p. 32714-9.

268. Issbrucker, K., et al., *p38 MAP kinase--a molecular switch between VEGF-induced angiogenesis and vascular hyperpermeability.* FASEB J, 2003. **17**(2): p. 262-4.

269. Matsumoto, T., et al., *p38 MAP kinase negatively regulates endothelial cell survival, proliferation, and differentiation in FGF-2-stimulated angiogenesis.* J Cell Biol, 2002. **156**(1): p. 149-60.

270. McMullen, M.E., et al., *Activation of p38 has opposing effects on the proliferation and migration of endothelial cells.* J Biol Chem, 2005. **280**(22): p. 20995-1003.

271. Rousseau, S., et al., *p38 MAP kinase activation by vascular endothelial growth factor mediates actin reorganization and cell migration in human endothelial cells.* Oncogene, 1997. **15**(18): p. 2169-77.

272. Hart, M.J., et al., *IQGAP1, a calmodulin-binding protein with a rasGAP-related domain, is a potential effector for cdc42Hs.* EMBO J, 1996. **15**(12): p. 2997-3005.

273.   Yamaoka-Tojo, M., et al., *IQGAP1, a novel vascular endothelial growth factor receptor binding protein, is involved in reactive oxygen species--dependent endothelial migration and proliferation.* Circ Res, 2004. **95**(3): p. 276-83.

274.   Fournier, E., et al., *Mutation at tyrosine residue 1337 abrogates ligand-dependent transforming capacity of the FLT4 receptor.* Oncogene, 1995. **11**(5): p. 921-31.

275.   Dixelius, J., et al., *Ligand-induced vascular endothelial growth factor receptor-3 (VEGFR-3) heterodimerization with VEGFR-2 in primary lymphatic endothelial cells regulates tyrosine phosphorylation sites.* J Biol Chem, 2003. **278**(42): p. 40973-9.

276.   Saharinen, P., et al., *Lymphatic vasculature: development, molecular regulation and role in tumor metastasis and inflammation.* Trends Immunol, 2004. **25**(7): p. 387-95.

277.   Makinen, T., et al., *Isolated lymphatic endothelial cells transduce growth, survival and migratory signals via the VEGF-C/D receptor VEGFR-3.* EMBO J, 2001. **20**(17): p. 4762-73.

278.   Karkkainen, M.J., et al., *Vascular endothelial growth factor C is required for sprouting of the first lymphatic vessels from embryonic veins.* Nat Immunol, 2004. **5**(1): p. 74-80.

279.   Wang, J.F., X. Zhang, and J.E. Groopman, *Activation of vascular endothelial growth factor receptor-3 and its downstream signalling promote cell survival under oxidative stress.* J Biol Chem, 2004. **279**(26): p. 27088-97.

280.   Korpelainen, E.I., et al., *Endothelial receptor tyrosine kinases activate the STAT signalling pathway: mutant Tie-2 causing venous malformations signals a distinct STAT activation response.* Oncogene, 1999. **18**(1): p. 1-8.

281.   Yuan, L., et al., *Abnormal lymphatic vessel development in neuropilin 2 mutant mice.* Development, 2002. **129**(20): p. 4797-806.

282.   Siemann, D.W., D.J. Chaplin, and M.R. Horsman, *Realizing the Potential of Vascular Targeted Therapy: The Rationale for Combining Vascular Disrupting Agents and Anti-Angiogenic Agents to Treat Cancer.* Cancer Invest, 2017. **35**(8): p. 519-534.

283.   Moyerbrailean, G.A., et al., *Which Genetics Variants in DNase-Seq Footprints Are More Likely to Alter Binding?* PLoS Genet, 2016. **12**(2): p. e1005875.

284.   Bysani, M., et al., *ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes.* Sci Rep, 2019. **9**(1): p. 7785.

285.   Wingett, S.W. and S. Andrews, *FastQ Screen: A tool for multi-genome mapping and quality control.* F1000Res, 2018. **7**: p. 1338.

286.   Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.

287.   Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

288.   Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

289.   Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.

290.   Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

291.   Yu, G., L.G. Wang, and Q.Y. He, *ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization.* Bioinformatics, 2015. **31**(14): p. 2382-3.

292.   Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.

293.   Ramirez, F., et al., *deepTools: a flexible platform for exploring deep-sequencing data.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W187-91.

294.   Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nat Biotechnol, 2010. **28**(8): p. 817-25.

295. Ernst, J., et al., *Mapping and analysis of chromatin state dynamics in nine human cell types.* Nature, 2011. **473**(7345): p. 43-9.

296. Cui, P., et al., *Comparative analyses of H3K4 and H3K27 trimethylations between the mouse cerebrum and testis.* Genomics Proteomics Bioinformatics, 2012. **10**(2): p. 82-93.

297. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.

298. Guttman, M., et al., *Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs.* Nat Biotechnol, 2010. **28**(5): p. 503-10.

299. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.

300. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

301. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

302. Jassal, B., et al., *The reactome pathway knowledgebase.* Nucleic Acids Res, 2020. **48**(D1): p. D498-D503.

303. Krishna G Aragam., e.a., *Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants.* medRxiv, 2021.

304. Evangelou, E., et al., *Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits.* Nat Genet, 2018. **50**(10): p. 1412-1425.

305. Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB.* Genome Res, 2012. **22**(9): p. 1790-7.

306. Liao, X., et al., *Exploration and detection of potential regulatory variants in refractive error GWAS.* Sci Rep, 2016. **6**: p. 33090.

307. Ward, L.D. and M. Kellis, *HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease.* Nucleic Acids Res, 2016. **44**(D1): p. D877-81.

308. Ward, L.D. and M. Kellis, *HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.* Nucleic Acids Res, 2012. **40**(Database issue): p. D930-4.

309. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.

310. Bailey, T.L. and P. Machanick, *Inferring direct DNA binding from ChIP-seq.* Nucleic Acids Res, 2012. **40**(17): p. e128.

311. Richmond, T.J. and C.A. Davey, *The structure of DNA in the nucleosome core.* Nature, 2003. **423**(6936): p. 145-50.

312. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

313. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

314. Henikoff, S. and K. Ahmad, *Assembly of variant histones into chromatin.* Annu Rev Cell Dev Biol, 2005. **21**: p. 133-53.

315. Szenker, E., D. Ray-Gallet, and G. Almouzni, *The double face of the histone variant H3.3.* Cell Res, 2011. **21**(3): p. 421-34.

316. Hake, S.B. and C.D. Allis, *Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis".* Proc Natl Acad Sci U S A, 2006. **103**(17): p. 6428-35.

317. Radman-Livaja, M. and O.J. Rando, *Nucleosome positioning: how is it established, and why does it matter?* Dev Biol, 2010. **339**(2): p. 258-66.

318. Gross, D.S. and W.T. Garrard, *Nuclease hypersensitive sites in chromatin.* Annu Rev Biochem, 1988. **57**: p. 159-97.

319. Gaspar-Maia, A., et al., *Chd1 regulates open chromatin and pluripotency of embryonic stem cells.* Nature, 2009. **460**(7257): p. 863-8.

320. Hargreaves, D.C. and G.R. Crabtree, *ATP-dependent chromatin remodeling: genetics, genomics and mechanisms.* Cell Res, 2011. **21**(3): p. 396-420.

321. Schwartzentruber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma.* Nature, 2012. **482**(7384): p. 226-31.

322. Tsompana, M. and M.J. Buck, *Chromatin accessibility: a window into the genome.* Epigenetics Chromatin, 2014. **7**(1): p. 33.

323. Xi, H., et al., *Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome.* PLoS Genet, 2007. **3**(8): p. e136.

324. Boyle, A.P., et al., *High-resolution mapping and characterization of open chromatin across the genome.* Cell, 2008. **132**(2): p. 311-22.

325. Stitzel, M.L., et al., *Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci.* Cell Metab, 2010. **12**(5): p. 443-55.

326. Buenrostro, J.D., et al., *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.* Curr Protoc Mol Biol, 2015. **109**: p. 21 29 1-9.

327. Buenrostro, J.D., et al., *Single-cell chromatin accessibility reveals principles of regulatory variation.* Nature, 2015. **523**(7561): p. 486-90.

328. Adey, A., et al., *Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition.* Genome Biol, 2010. **11**(12): p. R119.

329. Goryshin, I.Y. and W.S. Reznikoff, *Tn5 in vitro transposition.* J Biol Chem, 1998. **273**(13): p. 7367-74.

330. Koues, O.I., et al., *Distinct Gene Regulatory Pathways for Human Innate versus Adaptive Lymphoid Cells.* Cell, 2016. **165**(5): p. 1134-1146.

331. Shih, H.Y., et al., *Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality.* Cell, 2016. **165**(5): p. 1120-1133.

332. Wu, J., et al., *The landscape of accessible chromatin in mammalian preimplantation embryos.* Nature, 2016. **534**(7609): p. 652-7.

333. Gray, L.T., et al., *Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex.* Elife, 2017. **6**.

334. Scott, L.J., et al., *The genetic regulatory signature of type 2 diabetes in human skeletal muscle.* Nat Commun, 2016. **7**: p. 11764.

335. Xu, J., et al., *Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells.* Nat Genet, 2017. **49**(3): p. 377-386.

336. Bao, X., et al., *A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63.* Genome Biol, 2015. **16**: p. 284.

337. Corces, M.R., et al., *Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution.* Nat Genet, 2016. **48**(10): p. 1193-203.

338. Corces, M.R., et al., *An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues.* Nat Methods, 2017. **14**(10): p. 959-962.

339. Rouillard, A.D., et al., *The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins.* Database (Oxford), 2016. **2016**.

340. Lilly, B., *We have contact: endothelial cell-smooth muscle cell interactions.* Physiology (Bethesda), 2014. **29**(4): p. 234-41.

341. Doran, A.C., N. Meller, and C.A. McNamara, *Role of smooth muscle cells in the initiation and early progression of atherosclerosis.* Arterioscler Thromb Vasc Biol, 2008. **28**(5): p. 812-9.

342. Lusis, A.J., *Atherosclerosis.* Nature, 2000. **407**(6801): p. 233-41.

343. Hirase, T. and K. Node, *Endothelial dysfunction as a cellular mechanism for vascular failure.* Am J Physiol Heart Circ Physiol, 2012. **302**(3): p. H499-505.

344. Mombouli, J.V. and P.M. Vanhoutte, *Endothelial dysfunction: from physiology to therapy.* J Mol Cell Cardiol, 1999. **31**(1): p. 61-74.

345. Xu, J. and M.H. Zou, *Molecular insights and therapeutic targets for diabetic endothelial dysfunction.* Circulation, 2009. **120**(13): p. 1266-86.

346. Bennett, M.R., S. Sinha, and G.K. Owens, *Vascular Smooth Muscle Cells in Atherosclerosis.* Circ Res, 2016. **118**(4): p. 692-702.

347. Weber, C. and H. Noels, *Atherosclerosis: current pathogenesis and therapeutic options.* Nat Med, 2011. **17**(11): p. 1410-22.

348. Fearon, I.M., M.D. Gaca, and B.K. Nordskog, *In vitro models for assessing the potential cardiovascular disease risk associated with cigarette smoking.* Toxicol In Vitro, 2013. **27**(1): p. 513-22.

349. Bouis, D., et al., *Endothelium in vitro: a review of human vascular endothelial cell lines for blood vessel-related research.* Angiogenesis, 2001. **4**(2): p. 91-102.

350. Bernstein, B.E., et al., *Genomic maps and comparative analysis of histone modifications in human and mouse.* Cell, 2005. **120**(2): p. 169-81.

351. Garces de Los Fayos Alonso, I., et al., *The Role of Activator Protein-1 (AP-1) Family Members in CD30-Positive Lymphomas.* Cancers (Basel), 2018. **10**(4).

352. Yoshitomi, Y., et al., *Emerging Role of AP-1 Transcription Factor JunB in Angiogenesis and Vascular Development.* Int J Mol Sci, 2021. **22**(6).

353. Landt, S.G., et al., *ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia.* Genome Res, 2012. **22**(9): p. 1813-31.

354. Lalonde, S., et al., *Integrative analysis of vascular endothelial cell genomic features identifies AIDA as a coronary artery disease candidate gene.* Genome Biol, 2019. **20**(1): p. 133.

355. Miller, C.L., et al., *Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci.* Nat Commun, 2016. **7**: p. 12092.

356. Song, L., et al., *Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity.* Genome Res, 2011. **21**(10): p. 1757-67.

357. Ghisletti, S., et al., *Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages.* Immunity, 2010. **32**(3): p. 317-28.

358. Chrzanowska-Wodnicka, M., *Distinct functions for Rap1 signalling in vascular morphogenesis and dysfunction.* Exp Cell Res, 2013. **319**(15): p. 2350-9.

359. Geng, K., et al., *Electrical stimulation facilitates the angiogenesis of human umbilical vein endothelial cells through MAPK/ERK signalling pathway by stimulating FGF2 secretion.* Am J Physiol Cell Physiol, 2019. **317**(2): p. C277-C286.

360. Boopathy, G.T.K. and W. Hong, *Role of Hippo Pathway-YAP/TAZ Signalling in Angiogenesis.* Front Cell Dev Biol, 2019. **7**: p. 49.

361. Gaengel, K. and C. Betsholtz, *Endocytosis regulates VEGF signalling during angiogenesis.* Nat Cell Biol, 2013. **15**(3): p. 233-5.

362. Cattaneo, M.G., G. Lucci, and L.M. Vicentini, *Oxytocin stimulates in vitro angiogenesis via a Pyk-2/Src-dependent mechanism.* Exp Cell Res, 2009. **315**(18): p. 3210-9.

363. Jernigan, P.L., et al., *The role of sphingolipids in endothelial barrier function.* Biol Chem, 2015. **396**(6-7): p. 681-91.

364. Serban, D., J. Leng, and D. Cheresh, *H-ras regulates angiogenesis and vascular permeability by activation of distinct downstream effectors.* Circ Res, 2008. **102**(11): p. 1350-8.

365. Dubey, R.K., et al., *Cyclic AMP-adenosine pathway inhibits vascular smooth muscle cell growth.* Hypertension, 1996. **28**(5): p. 765-71.

366. Liu, Z. and R.A. Khalil, *Evolving mechanisms of vascular smooth muscle contraction highlight key targets in vascular disease.* Biochem Pharmacol, 2018. **153**: p. 91-122.

367. Reuwer, A.Q., et al., *Functional consequences of prolactin signalling in endothelial cells: a potential link with angiogenesis in pathophysiology?* J Cell Mol Med, 2012. **16**(9): p. 2035-48.

368. Goumans, M.J., Z. Liu, and P. ten Dijke, *TGF-beta signalling in vascular biology and dysfunction.* Cell Res, 2009. **19**(1): p. 116-27.

369. Dejana, E., *The role of wnt signalling in physiological and pathological angiogenesis.* Circ Res, 2010. **107**(8): p. 943-52.

370. Roy, A.R., et al., *The transcriptional regulator CCCTC-binding factor limits oxidative stress in endothelial cells.* J Biol Chem, 2018. **293**(22): p. 8449-8461.

371. Tang, M., et al., *Restraint of angiogenesis by zinc finger transcription factor CTCF-dependent chromatin insulation.* Proc Natl Acad Sci U S A, 2011. **108**(37): p. 15231-6.

372. Lu, J. and M. Tang, *CTCF-dependent chromatin insulator as a built-in attenuator of angiogenesis.* Transcription, 2012. **3**(2): p. 73-7.

373. Loukinov, D.I., et al., *BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6806-11.

374. Hore, T.A., J.E. Deakin, and J.A. Marshall Graves, *The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes.* PLoS Genet, 2008. **4**(8): p. e1000169.

375. Suzuki, T., et al., *Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS.* Mol Cell Biol, 2010. **30**(10): p. 2473-84.

376. Sleutels, F., et al., *The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner.* Epigenetics Chromatin, 2012. **5**(1): p. 8.

377. Pugacheva, E.M., et al., *The structural complexity of the human BORIS gene in gametogenesis and cancer.* PLoS One, 2010. **5**(11): p. e13872.

378. Kosaka-Suzuki, N., et al., *Transcription factor BORIS (Brother of the Regulator of Imprinted Sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter.* J Biol Chem, 2011. **286**(31): p. 27378-88.

379. Shaulian, E. and M. Karin, *AP-1 as a regulator of cell life and death.* Nat Cell Biol, 2002. **4**(5): p. E131-6.

380. Eferl, R. and E.F. Wagner, *AP-1: a double-edged sword in tumorigenesis.* Nat Rev Cancer, 2003. **3**(11): p. 859-68.

381. Glover, J.N. and S.C. Harrison, *Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA.* Nature, 1995. **373**(6511): p. 257-61.

382. Hicklin, D.J. and L.M. Ellis, *Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis.* J Clin Oncol, 2005. **23**(5): p. 1011-27.

383. Soker, S., et al., *Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor.* Cell, 1998. **92**(6): p. 735-45.

384. Matsumoto, T. and L. Claesson-Welsh, *VEGF receptor signal transduction.* Sci STKE, 2001. **2001**(112): p. re21.

385. Nilsson, M. and J.V. Heymach, *Vascular endothelial growth factor (VEGF) pathway.* J Thorac Oncol, 2006. **1**(8): p. 768-70.

386. Carmeliet, P., et al., *Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis.* Cell, 1999. **98**(2): p. 147-57.

387. Schlessinger, J., *New roles for Src kinases in control of cell survival and angiogenesis.* Cell, 2000. **100**(3): p. 293-6.

388. Mac Gabhann, F. and A.S. Popel, *Systems biology of vascular endothelial growth factors.* Microcirculation, 2008. **15**(8): p. 715-38.

389. Robinson, C.J. and S.E. Stringer, *The splice variants of vascular endothelial growth factor (VEGF) and their receptors.* J Cell Sci, 2001. **114**(Pt 5): p. 853-65.

390.   Koch, S. and L. Claesson-Welsh, *Signal transduction by vascular endothelial growth factor receptors.* Cold Spring Harb Perspect Med, 2012. **2**(7): p. a006502.

391.   Finley, S.D. and A.S. Popel, *Predicting the effects of anti-angiogenic agents targeting specific VEGF isoforms.* AAPS J, 2012. **14**(3): p. 500-9.

392.   Boucher, J.M. and V.L. Bautch, *Antiangiogenic VEGF-A in peripheral artery disease.* Nat Med, 2014. **20**(12): p. 1383-5.

393.   Chu, L.H., et al., *A multiscale computational model predicts distribution of anti-angiogenic isoform VEGF165b in peripheral arterial disease in human and mouse.* Sci Rep, 2016. **6**: p. 37030.

394.   Harper, S.J. and D.O. Bates, *VEGF-A splicing: the key to anti-angiogenic therapeutics?* Nat Rev Cancer, 2008. **8**(11): p. 880-7.

395.   Kim-Hellmuth, S., et al., *Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations.* Nat Commun, 2017. **8**(1): p. 266.

396.   Alasoo, K., et al., *Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response.* Nat Genet, 2018. **50**(3): p. 424-431.

397.   Ostuni, R., et al., *Latent enhancers activated by stimulation in differentiated cells.* Cell, 2013. **152**(1-2): p. 157-71.

398.   Gate, R.E., et al., *Genetic determinants of co-accessible chromatin regions in activated T cells across humans.* Nat Genet, 2018. **50**(8): p. 1140-1150.

399.   Shibuya, M., *Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signalling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies.* Genes Cancer, 2011. **2**(12): p. 1097-105.

400.   Liu, L., J.C. Tsai, and W.C. Aird, *Egr-1 gene is induced by the systemic administration of the vascular endothelial growth factor and the epidermal growth factor.* Blood, 2000. **96**(5): p. 1772-81.

401.   O'Donovan, K.J., et al., *The EGR family of transcription-regulatory factors: progress at the interface of molecular and systems neuroscience.* Trends Neurosci, 1999. **22**(4): p. 167-73.

402.   Fahmy, R.G., et al., *Transcription factor Egr-1 supports FGF-dependent angiogenesis during neovascularization and tumor growth.* Nat Med, 2003. **9**(8): p. 1026-32.

403.   Zeng, H., et al., *Orphan nuclear receptor TR3/Nur77 regulates VEGF-A-induced angiogenesis through its transcriptional activity.* J Exp Med, 2006. **203**(3): p. 719-29.

404.   Arkenbout, E.K., et al., *TR3 orphan receptor is expressed in vascular endothelial cells and mediates cell cycle arrest.* Arterioscler Thromb Vasc Biol, 2003. **23**(9): p. 1535-40.

405.   Abe, M. and Y. Sato, *cDNA microarray analysis of the gene expression profile of VEGF-activated human umbilical vein endothelial cells.* Angiogenesis, 2001. **4**(4): p. 289-98.

406.   Xi, H., et al., *Interplay between RORgammat, Egr3, and E proteins controls proliferation in response to pre-TCR signals.* Immunity, 2006. **24**(6): p. 813-826.

407.   Carter, J.H. and W.G. Tourellotte, *Early growth response transcriptional regulators are dispensable for macrophage differentiation.* J Immunol, 2007. **178**(5): p. 3038-47.

408.   Xi, H. and G.J. Kersh, *Early growth response gene 3 regulates thymocyte proliferation during the transition from CD4-CD8- to CD4+CD8+.* J Immunol, 2004. **172**(2): p. 964-71.

409.   Xi, H. and G.J. Kersh, *Sustained early growth response gene 3 expression inhibits the survival of CD4/CD8 double-positive thymocytes.* J Immunol, 2004. **173**(1): p. 340-8.

410.   Liu, D., et al., *Vascular endothelial growth factor-regulated gene expression in endothelial cells: KDR-mediated induction of Egr3 and the related nuclear receptors Nur77, Nurr1, and Nor1.* Arterioscler Thromb Vasc Biol, 2003. **23**(11): p. 2002-7.

411.   Qin, L., H. Zeng, and D. Zhao, *Requirement of protein kinase D tyrosine phosphorylation for VEGF-A165-induced angiogenesis through its interaction and regulation of phospholipase Cgamma phosphorylation.* J Biol Chem, 2006. **281**(43): p. 32550-8.

412.	Rey, O., et al., *Rapid protein kinase D translocation in response to G protein-coupled receptor activation. Dependence on protein kinase C.* J Biol Chem, 2001. **276**(35): p. 32616-26.

413.	Ye, J., et al., *Apelin and vascular endothelial growth factor are associated with mobilization of endothelial progenitor cells after acute myocardial infarction.* J Biomed Res, 2012. **26**(6): p. 400-9.

414.	Lee, C.S., et al., *Loss of phospholipase D2 impairs VEGF-induced angiogenesis.* BMB Rep, 2016. **49**(3): p. 191-6.

415.	Jia, J., et al., *AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation.* Microvasc Res, 2016. **105**: p. 103-8.

416.	Li, X. and U. Eriksson, *Novel VEGF family members: VEGF-B, VEGF-C and VEGF-D.* Int J Biochem Cell Biol, 2001. **33**(4): p. 421-6.

417.	Forouzanfar, M.H., et al., *Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015.* JAMA, 2017. **317**(2): p. 165-182.

418.	Munoz, M., et al., *Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank.* Nat Genet, 2016. **48**(9): p. 980-3.

419.	Feinleib, M., et al., *The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results.* Am J Epidemiol, 1977. **106**(4): p. 284-5.

420.	Cabrera, C.P., et al., *Exploring hypertension genome-wide association studies findings and impact on pathophysiology, pathways, and pharmacogenetics.* Wiley Interdiscip Rev Syst Biol Med, 2015. **7**(2): p. 73-90.

421.	Ehret, G.B., et al., *The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals.* Nat Genet, 2016. **48**(10): p. 1171-1184.

422.	Surendran, P., et al., *Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension.* Nat Genet, 2016. **48**(10): p. 1151-1161.

423.	Liu, C., et al., *Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci.* Nat Genet, 2016. **48**(10): p. 1162-70.

424.	Hoffmann, T.J., et al., *Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation.* Nat Genet, 2017. **49**(1): p. 54-64.

425.	Warren, H.R., et al., *Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk.* Nat Genet, 2017. **49**(3): p. 403-415.

426.	Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.

427.	International Consortium for Blood Pressure Genome-Wide Association, S., et al., *Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk.* Nature, 2011. **478**(7367): p. 103-9.

428.	Gaziano, J.M., et al., *Million Veteran Program: A mega-biobank to study genetic influences on health and disease.* J Clin Epidemiol, 2016. **70**: p. 214-23.

429.	Leitsalu, L., et al., *Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu.* Int J Epidemiol, 2015. **44**(4): p. 1137-47.

430.	Gimbrone, M.A., Jr. and G. Garcia-Cardena, *Endothelial Cell Dysfunction and the Pathobiology of Atherosclerosis.* Circ Res, 2016. **118**(4): p. 620-36.

431.	Liu, B., et al., *Genetic Regulatory Mechanisms of Smooth Muscle Cells Map to Coronary Artery Disease Risk Loci.* Am J Hum Genet, 2018. **103**(3): p. 377-388.

432.	Kheradpour, P. and M. Kellis, *Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments.* Nucleic Acids Res, 2014. **42**(5): p. 2976-87.

433.	Xu, Y., et al., *Regulation of endothelial intracellular adenosine via adenosine kinase epigenetically modulates vascular inflammation.* Nat Commun, 2017. **8**(1): p. 943.

434.	Zhang, F., et al., *Tetraspanin CD151 maintains vascular stability by balancing the forces of cell adhesion and cytoskeletal tension.* Blood, 2011. **118**(15): p. 4274-84.

435. Verrecchia, F., et al., *Smad3/AP-1 interactions control transcriptional responses to TGF-beta in a promoter-specific manner.* Oncogene, 2001. **20**(26): p. 3332-40.

436. Liu, D., et al., *The zinc-finger transcription factor, early growth response 3, mediates VEGF-induced angiogenesis.* Oncogene, 2008. **27**(21): p. 2989-98.

437. Baeten, J.T. and B. Lilly, *Notch Signalling in Vascular Smooth Muscle Cells.* Adv Pharmacol, 2017. **78**: p. 351-382.

438. Yang, W., et al., *Exploring the Mechanism of the miRNA-145/Paxillin Axis in Cell Metabolism During VEGF-A-Induced Corneal Angiogenesis.* Invest Ophthalmol Vis Sci, 2021. **62**(10): p. 25.

439. Morlon, A. and P. Sassone-Corsi, *The LIM-only protein FHL2 is a serum-inducible transcriptional coactivator of AP-1.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3977-82.

440. Hayashi, H., et al., *FHL-2 suppresses VEGF-induced phosphatidylinositol 3-kinase/Akt activation via interaction with sphingosine kinase-1.* Arterioscler Thromb Vasc Biol, 2009. **29**(6): p. 909-14.

441. Simons, K.H., et al., *IRF3 and IRF7 mediate neovascularization via inflammatory cytokines.* J Cell Mol Med, 2019. **23**(6): p. 3888-3896.

442. Yordy, J.S., et al., *SP100 inhibits ETS1 activity in primary endothelial cells.* Oncogene, 2005. **24**(5): p. 916-31.

443. Wasylyk, C., et al., *Sp100 interacts with ETS-1 and stimulates its transcriptional activity.* Mol Cell Biol, 2002. **22**(8): p. 2687-702.

444. Liu, C.L., et al., *Cysteine protease cathepsins in cardiovascular disease: from basic research to clinical trials.* Nat Rev Cardiol, 2018. **15**(6): p. 351-370.

445. Sukhova, G.K., et al., *Expression of the elastolytic cathepsins S and K in human atheroma and regulation of their production in smooth muscle cells.* J Clin Invest, 1998. **102**(3): p. 576-83.

446. Shi, G.P., et al., *Cystatin C deficiency in human atherosclerosis and aortic aneurysms.* J Clin Invest, 1999. **104**(9): p. 1191-7.

447. Shi, G.P., et al., *Role for cathepsin F in invariant chain processing and major histocompatibility complex class II peptide loading by macrophages.* J Exp Med, 2000. **191**(7): p. 1177-86.

448. Shi, G.P., et al., *Cathepsin S required for normal MHC class II peptide loading and germinal center development.* Immunity, 1999. **10**(2): p. 197-206.

449. Nakagawa, T., et al., *Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus.* Science, 1998. **280**(5362): p. 450-3.

450. Sun, J., et al., *Deficiency of antigen-presenting cell invariant chain reduces atherosclerosis in mice.* Circulation, 2010. **122**(8): p. 808-20.

451. Ewald, S.E., et al., *The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor.* Nature, 2008. **456**(7222): p. 658-62.

452. Zhou, Y., et al., *Cathepsin K Deficiency Ameliorates Systemic Lupus Erythematosus-like Manifestations in Fas(lpr) Mice.* J Immunol, 2017. **198**(5): p. 1846-1854.

453. Taleb, S., A. Tedgui, and Z. Mallat, *Regulatory T-cell immunity and its relevance to atherosclerosis.* J Intern Med, 2008. **263**(5): p. 489-99.

454. Sharir, R., et al., *Experimental myocardial infarction induces altered regulatory T cell hemostasis, and adoptive transfer attenuates subsequent remodeling.* PLoS One, 2014. **9**(12): p. e113653.

455. Zhou, Y., et al., *Regulatory T cells in human and angiotensin II-induced mouse abdominal aortic aneurysms.* Cardiovasc Res, 2015. **107**(1): p. 98-107.

456. Westenskow, P.D., et al., *Ras pathway inhibition prevents neovascularization by repressing endothelial cell sprouting.* J Clin Invest, 2013. **123**(11): p. 4900-8.

457. Zheng, Q., et al., *The Integrated Landscape of Biological Candidate Causal Genes in Coronary Artery Disease.* Front Genet, 2020. **11**: p. 320.

458. van Buul, J.D., et al., *RhoG regulates endothelial apical cup assembly downstream from ICAM1 engagement and is involved in leukocyte trans-endothelial migration.* J Cell Biol, 2007. **178**(7): p. 1279-93.

459. Themistocles Assimes., e.a., *A large-scale multi-ethnic genome-wide association study of coronary artery disease.* Research Square, 2021.

460. Muerdter, F., L.M. Boryn, and C.D. Arnold, *STARR-seq - principles and applications.* Genomics, 2015. **106**(3): p. 145-150.

461. Santiago-Algarra, D., et al., *Recent advances in high-throughput approaches to dissect enhancer function.* F1000Res, 2017. **6**: p. 939.

462. Wang, J., et al., *HACER: an atlas of human active enhancers to interpret regulatory variants.* Nucleic Acids Res, 2019. **47**(D1): p. D106-D112.

463. Selvarajan, I., et al., *Integrative analysis of liver-specific non-coding regulatory SNPs associated with the risk of coronary artery disease.* Am J Hum Genet, 2021. **108**(3): p. 411-430.

464. Sun, L., B.M. Lutz, and Y.X. Tao, *The CRISPR/Cas9 system for gene editing and its potential application in pain research.* Transl Perioper Pain Med, 2016. **1**(3): p. 22-33.

465. Rees, H.A. and D.R. Liu, *Base editing: precision chemistry on the genome and transcriptome of living cells.* Nat Rev Genet, 2018. **19**(12): p. 770-788.

466. Goodwin, A.M., *In vitro assays of angiogenesis for assessment of angiogenic and anti-angiogenic agents.* Microvasc Res, 2007. **74**(2-3): p. 172-83.

467. Harman, J.L., et al., *Emerging Roles for Neuropilin-2 in Cardiovascular Disease.* Int J Mol Sci, 2020. **21**(14).

468. Caunt, M., et al., *Blocking neuropilin-2 function inhibits tumor cell metastasis.* Cancer Cell, 2008. **13**(4): p. 331-42.

469. Parker, M.W., et al., *Structural basis for VEGF-C binding to neuropilin-2 and sequestration by a soluble splice form.* Structure, 2015. **23**(4): p. 677-87.

470. Said, A.M., M.W. Parker, and C.W. Vander Kooi, *Design, synthesis, and evaluation of a novel benzamidine-based inhibitor of VEGF-C binding to Neuropilin-2.* Bioorg Chem, 2020. **100**: p. 103856.