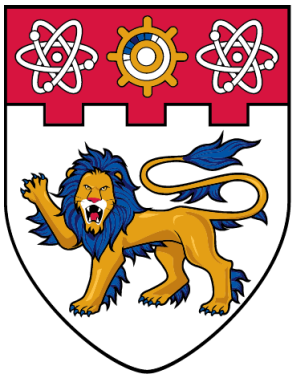


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

FUNCTIONAL GENETICS OF CORONARY ARTERY DISEASE SUSCEPTIBILITY LOCUS 6P24.1 IN ENDOTHELIAL DISEASE BIOLOGY

TAY KAI YI

LEE KONG CHIAN SCHOOL OF MEDICINE

2025

FUNCTIONAL GENETICS OF CORONARY ARTERY DISEASE SUSCEPTIBILITY LOCUS 6P24.1 IN ENDOTHELIAL DISEASE BIOLOGY

TAY KAI YI

Lee Kong Chian School of Medicine

A thesis submitted to the Nanyang Technological University

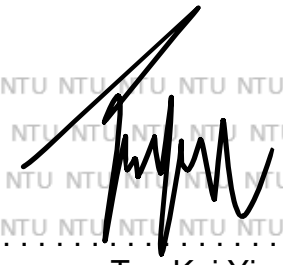
in partial fulfilment of the requirement for the degree of

Doctor of Philosophy

2025

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.



10/01/2025

. . . . . .
Date

. . . . . .
Tay Kai Yi

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

10/01/2025

. . . . . .
Date

. . . . . .
Assoc Prof Christine Cheung

Authorship Attribution Statement

This thesis contains material from 2 papers published in the following peer-reviewed journal(s) in which I am listed as an author.

Chapters 3 and 4 are published as Tay KY, Wu KX, Chioh FWJ, Autio MI, Pek NMQ, Narmada BC, Tan SH, Low AF, Lian MM, Chew EGY, Lau HH, Kao SL, Teo AKK, Foo JN, Foo RSY, Heng CK, Chan MYY, Cheung C. Trans-interaction of risk loci 6p24.1 and 10q11.21 is associated with endothelial damage in coronary artery disease. Atherosclerosis. 2022 Dec;362:11-22. doi: 10.1016/j.atherosclerosis.2022.10.012.

The contributions of the co-authors are as follows:

- A/Prof Christine Cheung provided the initial project direction and edited the manuscript drafts.
- I prepared the initial manuscript draft, co-designed the study with A/Prof Christine Cheung, and performed all the laboratory work at the Lee Kong Chian School of Medicine, School of Biological Sciences, and Yong Loo Lin School of Medicine. My specific contributions included performing CRISPR-Cas9 NHEJ editing, iPSC differentiation into endothelial cells, quantitative RT-PCR, 3C-droplet digital PCR and profiling of circulating endothelial cells. Additionally, I conducted ChIP Sequencing, RNA Sequencing and Hi-C data analysis.
- Dr Kanxing Wu assisted with the CRISPR-Cas9 NHEJ editing and prepared the samples for RNA-Sequencing and Hi-C.
- Ms Florence Chioh assisted in profiling circulating endothelial cells.
- Dr Matias Ilmari Autio and Prof Roger Sik Yin Foo provided guidance on the CRISPR-Cas9 NHEJ editing.
- Ms Nicole Min Qian Pek and Dr Balakrishnan Chakrapani Narmada were involved in patient sample recruitment, generation of iPSCs, and endothelial characterizations.
- Ms Michelle Mulan Lian and A/P Jia Nee Foo provided guidance on RNA Sequencing analysis.

- Dr Sock-Hwee Tan, A/P Adrian Fatt-Hoe Low, Dr Elaine Guo Yan Chew, Dr Hwee Hui Law, Dr Shih Ling Kao, A/P Adrian Kee Keong Teo, A/P Chew Kiat Heng and A/P Mark Yan Yee Chan provided clinical patient samples and demographics.

Chapter 5 is currently pending revision and available as a preprint on bioRxiv as Tay KY, Cheung C, Wee HSA, Nguyen N, Autio MI, Wazny VK, Lee KL, Tay D, Wang Y, Gao X, Heng CK, Chan MYY, Foo RSY, Lee J, Loh M. Coronary Artery Disease Risk Variant rs6903956 Links to Endothelial Dysfunction via *PHACTR1* Regulation. bioRxiv. 2025; doi: 10.1101/2025.05.11.653298

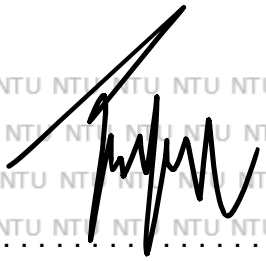
The contributions of the co-authors are as follows:

- A/Prof Christine Cheung provided the initial project direction and edited the manuscript drafts.
- I prepared the initial manuscript draft, co-designed the study with A/Prof Christine Cheung, and performed all laboratory experiments conducted at the Lee Kong Chian School of Medicine and School of Biological Sciences. My specific contributions included CRISPR-Cas9 base editing, iPSC differentiation into arterial endothelial cells, quantitative RT-PCR, ChIP-qPCR, siRNA knockdown and dual luciferase assay. I also conducted MESA phenotype association and HELIOS eQTL rs6903956 data analysis.
- Ms Hannah Su-Ann Wee and Ms Khang Leng Lee performed the Ibidi flow assays.
- Ms Hannah Su-Ann Wee and Ms Vanessa Kristina Wazny conducted the immunofluorescence staining experiments.
- Ms Nhi Nguyen performed the AlphaFold 3 structural representation.
- Dr Matias Ilmari Autio and Prof Roger Sik Yin Foo were responsible for the generation of the pMIA20 construct.

- Dr Darwin Tay, Dr Jimmy Lee and A/P Marie Loh conducted the HELIOS expression QTL.
- Ms Yuting Wang and Dr Xu Gao performed analyses for Guilin Cohort Study.
- A/P Chew Kiat Heng and A/P Mark Yan Yee Chan provided clinical patient samples and associated demographic data.

10/01/2025

. . . . . .
Date



. . . . . .
Tay Kai Yi

Acknowledgment

I am profoundly grateful to my supervisor, Assoc Prof Christine Cheung, for her unparalleled guidance and support from my URECA-FYP through to my PhD. Her nurturing mentorship created a secure environment that enabled me to explore my scientific aspirations freely and independently. Her exceptional problem-solving skills and critical thinking have shaped me into a capable and confident scientist. I am truly fortunate to have met such an amazing, compassionate, and brilliant mentor at this early stage of my career. Her influence has set a high standard for all my future endeavors, and the love and respect I have gained under her mentorship are immeasurable.

In addition, I extend heartfelt gratitude to my previous post-doctoral mentor, Dr Wu Kanxing, who has been pivotal in teaching me the essential laboratory skills and scientific thought processes that form the backbone of my work. His methodical approach and thorough emphasis on foundational skills have been key in shaping my approach to experiment planning and scientific writing. I attribute my discipline, thoroughness, and precision as a scientist today to his invaluable mentorship.

Special mention goes to my best friend in the lab, Vanessa Wazny, for brightening every day with delightful lunches and endless inside jokes. You have been the light in my life these past four years. I am also thankful to my past and present lab members—Florence Chioh, Dr Natalie Yeo, Lee Khang Leng, Nguyen Le Uyen Nhi, Hannah Wee, and Dr See Jia Xiang—for their companionship on this journey. Every scientific discussion and forged friendship have been invaluable.

My gratitude extends to my thesis advisory committee members—Prof John Chambers, Assoc Prof Rinkoo Dalan, and Asst Prof Franklin Zhong—for their generous scientific advice and the selfless sharing of their resources, which have been instrumental in my research.

I acknowledge the support of the Academic Research Fund Tier 1 grant (2018-T1-001-030) from the Ministry of Education, Singapore, and the Human Frontier Science Program Research Grant (RGY0069/2019) for funding my research. My PhD journey was also supported by the NTU Research Scholarship, and the LKCMedicine RASS and Graduate Programmes Office for their administrative assistance.

Lastly, the deepest appreciation goes to my family, who has been my sanctuary and greatest source of encouragement throughout this journey. Papa and Mama, your understanding and support have been my pillars of strength. I could not have reached this milestone without your unwavering love.

Table of Contents

Summary

Coronary artery disease (CAD) continues to be the primary cause of mortality worldwide. The single nucleotide polymorphism (SNP) rs6903956, located within the non-coding region of the *ADTRP* gene on chromosome 6p24.1, is recognized as a genetic determinant for CAD risk in East Asian populations. However, the specific role of rs6903956 in the development of CAD is still not fully understood. The objective of this dissertation is to delineate the specific molecular interactions through which rs6903956 contributes to vascular dysfunction in CAD.

Data from the Multi-Ethnic Study of Atherosclerosis (MESA) cohort support the role of rs6903956 in endothelial cell function. Induced pluripotent stem cells (iPSCs) were generated from CAD patients with the AA risk genotype and non-CAD subjects with the GG non-risk genotype at rs6903956. Using CRISPR-Cas9 editing, we generated isogenic iPSC-derived endothelial cells, converting the rs6903956 AA risk genotype in CAD patient-derived iPSCs to the GG non-risk genotype.

Our findings revealed that the A allele at rs6903956 enhances *PHACTR1* expression, a gene situated within the same topologically associating domain (TAD) as rs6903956, with known vascular effects in endothelial cells. This regulatory effect is supported by eQTL data from the Health for Life in Singapore (HELIOS) Study. Additionally, we demonstrated that transcription factor complexes from the HOX and TALE families differentially bind to the rs6903956 enhancer region, regulating *PHACTR1* transcription. RNA sequencing and chromatin conformation capture assays (Hi-C and 3C) identified inter-chromosomal interactions between the 6p24.1 risk locus and a weak promoter of *CXCL12*, a CXC chemokine ligand on chromosome 10q11.21. Circulating endothelial cells (CECs) from CAD patients with the AA/AG genotypes exhibited elevated *CXCL12* expression and increased markers of endothelial injury.

These findings highlight the regulatory role of rs6903956 in CAD, suggesting a cis-acting mechanism affecting *PHACTR1* and a trans-acting mechanism involving interactions between the 6p24.1 and 10q11.21 loci. By elucidating these molecular pathways, our study advances the understanding of genetic determinants in cardiovascular disease and provides insights into potential personalized medicine therapeutic targets.

Chapter 1: Introduction and background

Chapter 1: Introduction and background

1.1 Structure of thesis

This dissertation is organized into five chapters, each addressing a specific part of the study with background, specific aims, methodology, results, and discussion sections.

Chapter 1 provides a comprehensive background and literature review, outlining the current knowledge on CAD genetics, the significance of East Asian-specific loci, and the regulatory role of rs6903956 as the focus of this study.

Chapter 2 focuses on the *in silico* and population-based analysis of rs6903956. This chapter begins with statistical fine-mapping to identify the causal variant at the 6p24.1 locus. It then integrates population cohort analyses to determine cardiovascular phenotypes associated with rs6903956 and identify a specific disease-relevant cell type. Finally, it uses genomic annotations to predict the regulatory role of rs6903956 and its impact on gene expression.

Chapter 3 describes the generation of a disease-relevant cell model using CRISPR-mediated non-homologous end joining (NHEJ) on iPSCs, derived from CAD patient and control samples. By deleting the region around rs6903956, this chapter aims to establish a cell line that enables the functional interrogation of the SNP in a relevant disease context.

Chapter 4 utilizes the disease-relevant cell model to explore the trans-impact of rs6903956, focusing on identifying the causal gene(s) regulated by this SNP and linking it to CAD pathogenesis.

Chapter 5 experimentally validates the precise regulatory mechanism of rs6903956, including the identification of its binding transcription factor and the downstream cis-regulated gene(s) contributing to CAD pathogenesis.

Finally, we conclude the dissertation with a summary of key findings, a discussion of their significance, and potential directions for future research. This includes further dissection of the role of rs6903956 in CAD pathogenesis and the application of the methods developed here for studying other GWAS-identified SNPs.

1.2 Purpose and significance

CAD remains the leading cause of mortality worldwide[1], emphasizing the critical need to unravel its molecular underpinnings to improve diagnostic accuracy, risk assessment, and treatment strategies. The etiology of CAD is multifactorial, involving both genetic predispositions and environmental factors. GWAS have identified over 163 risk loci for CAD[2], which include genetic variants that influence critical physiological processes including blood pressure regulation and vascular structural integrity. However, many of these loci, particularly non-coding variants, remain functionally uncharacterized and are likely to influence gene expression through cis-regulatory elements (CREs) such as enhancers and promoters[3-5].

These regulatory variants may alter transcription factor binding, modify chromatin accessibility, and affect enhancer-promoter interactions, leading to genotype-specific effects on gene expression. While it is commonly inferred that the causal gene is the one directly proximal to the variant, many studies suggest this is not always the case. For instance, one of the pioneers in GWAS characterization showed that rs12740374[6] enhances the binding of C/EBP at a regulatory site, altering levels of SORT1 and contributing to CAD risk. Notably, *SORT1* was identified as the causal gene, rather than the physically closer genes *CELSR2* and *PSRC1*. Similarly, the variant rs9349379[7] located in a vascular-specific enhancer in *PHACTR1* regulates *EDN1*, a gene positioned more than 500kb away. Further validation efforts at loci such as 1p32.2, *HDAC9*, *LMOD1*, and *SMAD3*[8-11] have demonstrated the enhancer-driven regulatory effects of CAD-associated SNPs, highlighting the need for functional genomics to uncover causal mechanisms underlying GWAS variants.

Advancements in integrative functional genomics, bolstered by resources like ENCODE[12] and Roadmap Epigenomics[4], has demonstrated that SNPs associated with CAD often occur in cell-type-specific CREs. Large-scale datasets have shown that variants linked to cardiac traits predominantly localize to heart enhancers involved in vascular morphogenesis[13]. Conversely, enhancers within liver hepatocytes frequently relate to lipid metabolism[14]. EpiMap[3], which integrates 10,000 epigenomic maps across 800 samples, has further refined tissue-specific pathways, linking CAD-associated variants to vessel morphogenesis in heart tissue, lipid metabolism in the liver, and endothelial function in vascular tissues. These observations highlight the crucial role of cell type and tissue specificity in understanding the functional implications of loci identified by GWAS.

Despite these advances, majority of regulatory functional evidence stems from European cohorts, leaving non-European populations, particularly East Asians, underrepresented. Ethnic-specific differences in

allele frequencies and gene-environment interactions highlight the critical need for studies in diverse populations[15-17]. These differences can reveal population-specific regulatory mechanisms, offering insights into disease pathogenesis that may not be observed in European-centric studies.

This dissertation addresses this gap by focusing on rs6903956, an East Asian-specific variant located within the first intron of *ADTRP* at the 6p24.1 locus. Identified in Chinese Han populations[18] and validated in Singaporean and Japanese cohorts[19-22], rs6903956 has not been implicated in European populations, suggesting a regulatory mechanism unique to East Asian ancestry.

To establish the regulatory role of rs6903956, this study integrates existing genomic datasets, including ENCODE and Roadmap Epigenomics, to decipher enhancer activity through chromatin accessibility and transcriptional regulation. Our work shows the utility of these datasets in pinpointing functional elements at GWAS loci, demonstrating their applicability in interpreting non-coding variants associated with CAD.

In addition, our study employs two CRISPR technologies—NHEJ and base editing—to develop a CRISPR-edited CAD patient-derived induced pluripotent stem cell (iPSC) model. By differentiating these iPSCs into disease-relevant endothelial cells, we were able to directly interrogate rs6903956 in a physiologically relevant context. This model enabled us to bridge the gap between regulatory SNPs and their downstream biological effects by elucidating the precise transcription factor-mediated mechanism through which rs6903956 influences *PHACTR1*, a cis-causal gene located approximately 900 kb away, as well as a trans-regulated gene, *CXCL12*, a known CAD risk locus. This dual-gene interaction provides novel insights into the potential synergy between CAD risk loci, advancing our understanding of the mechanisms underlying CAD pathogenesis within an East Asian-specific genetic framework.

This dissertation demonstrates the power of integrative functional genomics combined with experimental validation in disease-specific models. By focusing on rs6903956, an East Asian-specific variant, this dissertation delivers insights into ancestry-specific regulatory mechanisms that contribute to CAD pathogenesis. Furthermore, it establishes a robust framework for investigating non-coding variants, paving the way for precision medicine approaches tailored to underrepresented populations, particularly East Asians. The combination of human genomic datasets and patient relevant experimental models ensures that this work not only deepens our understanding of CAD genetics but also sets a benchmark for future research in deciphering the role of GWAS variants in complex diseases.

1.3 Literature review

In this section, I provide an overview of the genetic influence of non-coding variants on complex diseases and outline methods to elucidate the mechanisms of CAD-associated variants identified through GWAS. I also highlight the role of endothelial vasculature in CAD, focusing on how non-coding variants impact endothelial cell function and disease progression.

1.3.1 Hereditary influence on disease susceptibility

Chronic multifactorial diseases are the primary drivers of healthcare demand[23-25] and have a significant impact on mortality and productivity in developed countries[26,27]. Unlike single gene disorders such as phenylketonuria or sickle cell anemia, these diseases tend to cluster within families but do not follow patterns of Mendelian single-gene disorders (Figure 1). Deciphering the genetic underpinnings of these hereditary conditions is challenging due to the complex variability of the human genome and the numerous genetic variables influencing disease susceptibility. The latest complete sequencing of the human genome in 2022[28], coupled with 1000 genome project published in 2015 aimed at characterizing human genetic variation[29], has provided us with extensive genetic information for identifying disease causative genes.

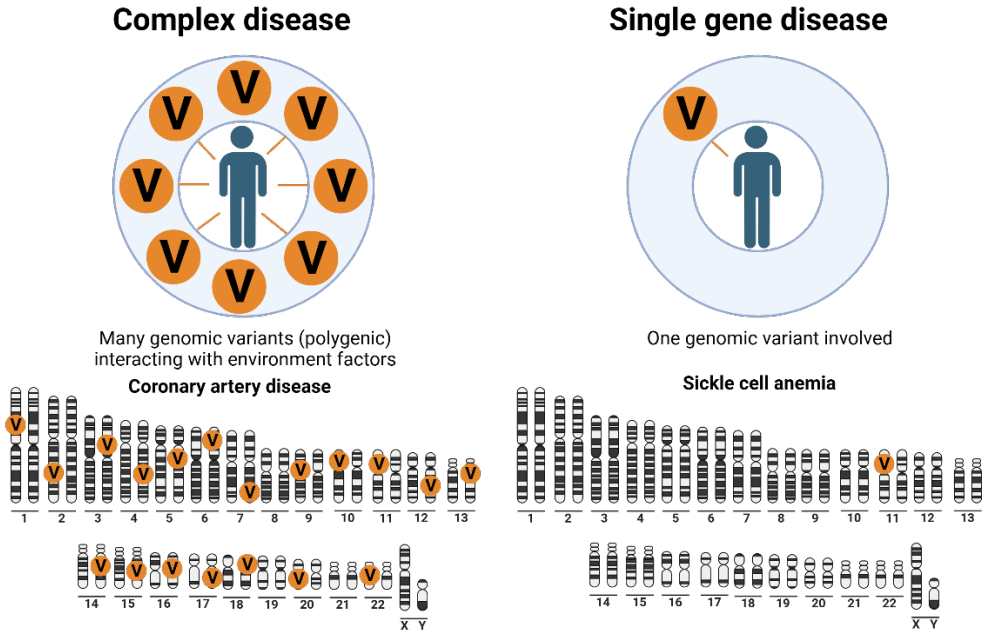


Figure 1. Complex vs single gene disease. Left: Complex diseases, such as CAD, are affected by multiple genomic variants, environmental, and lifestyle contributions. These diseases typically result from the combined effects of numerous small genetic changes. The chromosomal distribution of potential genetic variants associated with CAD is represented by orange markers (V) across various chromosomes, illustrating the

polygenic nature of complex diseases. Right: Single-gene diseases are a result of a single genetic variant with a direct and significant impact on the phenotype, exemplifying their monogenic nature. Figure created using BioRender.

1.3.1.1 Genetic inheritance of complex diseases

Nearly all conditions and diseases have a genetic component to them – including common health problems such as asthma, diabetes, dementia, CAD and depression. These are referred to as complex diseases, arising from the interplay between genetic predispositions and diverse environmental factors. These interactions, governed by dynamic, epigenetic, and regulatory mechanisms, eventually integrate to manifest the disease phenotype[30-33]. Complex diseases commonly display a polygenic nature, where disease manifestation arises from the combined effects of numerous genetic variants[34]. The concept of polygenic inheritance emerged based on observations that, although these disorders frequently cluster in families, they do not adhere to traditional Mendelian inheritance patterns. Specific risk values for related family members have been examined in genetic architecture models incorporating environmental factors, and in many cases, are consistent with polygenic inheritance[35-37].

1.3.1.2 Genetic variation in complex disease

Identifying contributors to complex diseases poses significant challenges due to the intricate interactions between genetic and environmental variables. The Human Genome Project has significantly altered our approach to studying complex diseases by illuminating the genetic mechanisms of various conditions[28]. The human genome contains 3.2 billion base pairs and approximately twenty thousand protein-coding genes. A pivotal finding from the Human Genome Project is the extensive genetic diversity present in the human genome, with SNPs being the most common form of genetic variation.

SNPs are minute variations in the genome, occurring as single base pair changes scattered across the DNA sequence. These variations, known as alleles, represent different versions of a DNA sequence found within a population. The occurrence of these alleles is measured by the minor allele frequency (MAF), with a variant classified as a polymorphism if it appears with a MAF of at least 1%. The human genome is estimated to contain at least 11 million SNPs, introducing substantial variation as approximately 3 million SNPs typically differ between any given individual and the reference genome. It's notable that not every gene requires full functionality—many operate with just one working copy, and

some display functional redundancy or plasticity, enhancing genetic flexibility.

The relevance of SNPs extends beyond mere markers of genetic diversity; they are pivotal in disease research. SNPs can highlight an increased susceptibility to common diseases, especially when they disrupt the normal function of a gene involved in disease pathways, resulting in discernible phenotypic effects. Moreover, SNPs tend to group into 'haplotype blocks' along chromosomes, areas where recombination is rare. This grouping results in linkage disequilibrium (LD), where alleles at adjacent SNP positions are inherited together more often than would be randomly expected. Within these haplotype blocks, 'tag' SNPs in high LD can uniquely mark these blocks, offering a valuable tool for genetic mapping and identifying disease associations (Figure 2).

The highest degree of genetic variation is observed within populations of African ancestry, which aligns with the migration patterns out of Africa throughout history. As different groups migrated, they carried subsets of genetic variants with them, leading to unique genetic profiles in various populations. While there are common variants shared among populations worldwide, rare variants tend to be more population specific. An example of population-specific variation is the presence of a polymorphism on the GRK5 gene, which is more prevalent in individuals of African American descent[38]. This polymorphism enhances the desensitization of beta-adrenergic receptors, resulting in a more efficient regulation of excess catecholamine signaling. This genetic variation has been associated with improved survival rates in African Americans with heart failure. However, they also respond less favorably to β-blocker medications compared to individuals of European descent. This underscores the importance of population-specific genetic variations and their implications on disease susceptibility, treatment response, and overall health outcomes.

To address the need for a comprehensive understanding of polymorphisms in the human genome, the 1000 Genomes Project[29] was launched to map genetic variation at an unprecedented scale. This project cataloged over 88 million genetic variants, including SNPs, insertions, deletions, across a diverse set of global populations. By analyzing genomes from multiple ancestry groups, the project provides insights into allele frequencies, LD, and population-specific variants. Similar to previous projects such as HapMap[39], the 1000 Genomes Project also utilized tag SNPs to efficiently capture genomic variation patterns.

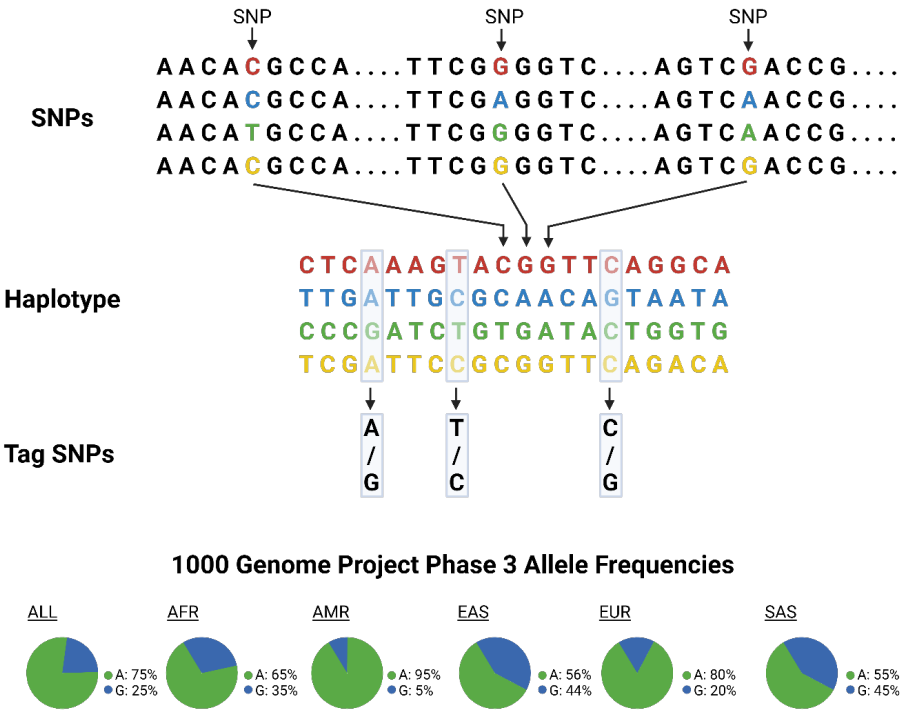


Figure 2. Genetic variation across populations: SNPs, haplotypes, and tag SNPs. A short stretch of DNA is shown to illustrate SNPs, where each SNP has two possible alleles (e.g., C and T). Haplotypes, defined as combinations of alleles at nearby SNPs, are depicted using 20 SNPs spanning a 6,000-base region, with only the most variable bases shown. Tag SNPs are highlighted as representative markers sufficient to uniquely identify these haplotypes. Examples of genotype frequencies output obtained from 1000 Genomes Project Phase 3 demonstrate significant variation among populations, including African (AFR), European (EUR), East Asian (EAS), South Asian (SAS), and American (AMR) populations, highlighting the genetic diversity and population-specific allele distributions across global groups. Figure created using BioRender.

1.3.1.3 Genome wide association studies

The recognition that numerous diseases have a polygenic nature, coupled with the advances of HapMap and 1000 Genome Project, has provided a foundation for GWAS[39]. GWAS is designed to identify correlations between genetic markers and disease phenotypes by analyzing how allele frequencies of genetic variants differ among individuals from similar ancestral backgrounds but with different phenotypes[40,41]. This method involves comparing SNP frequencies between diseased individuals (cases) and healthy controls to compute an odds ratio (OR) for each SNP, which quantifies the risk of disease. The steps for conducting GWAS is illustrated in figure 3. It is noteworthy that GWAS utilize tag SNPs to thoroughly assess genomic variation, therefore aiming not directly to discover causal alleles but to pinpoint genomic regions likely harboring them.

GWAS has been instrumental in pinpointing candidate loci linked to a variety of complex diseases. Notably, GWAS have made significant contributions to our understanding of diseases such as type 1[42] and type 2 diabetes[43], Crohn's disease[44], CAD[45], among others. For instance, GWAS facilitated the discovery of the PCSK9 locus in CAD, illustrating that mutations inactivating this gene correlate with lower LDL cholesterol levels and a decreased risk of CAD[46,47]. This discovery led to the development of FDA-approved monoclonal antibodies, such as Alirocumab[48] and Evolucumab[49]. These findings have also provided insights into the underlying biochemical pathways and processes involved in various disorders. For example, GWAS shed light on the mechanisms underlying beta-cell dysfunction and insulin resistance in type 2 diabetes[50], nitric oxide signaling in cardiovascular disease[51] and immune system dysregulation in schizophrenia[52]. Moreover, GWAS has also uncovered shared genetic risk factors between certain disorders, such as psychiatric disorders and autoimmune diseases[53]. This discovery highlighted the intricate interplay between genetic factors and the manifestation of diverse disease phenotypes.

The NHGRI-EBI GWAS Catalog currently summarizes 537,476 statistical associations from 6459 publications[54]. However, a prevailing trend observed across these studies is that the identified SNPs have a relatively low effect size on disease risk, typically ranging from ORs of less than 1.2[55]. The phenomenon of 'missing heritability' is thought to arise from several factors: elusive variants with minor effects, challenges in detecting rare variants, limitations in identifying structural variants, insufficient power to discern gene-gene interactions, and overlooked environmental factors among relatives[56]. Consequently, integrating data into polygenic scores has become essential, as multiple genes with minor effects can significantly influence the heritability of traits. While individual variants typically exert minor effects, their aggregate impact can be profound, potentially resulting in risk levels comparable with those associated with strong monogenic variants[57,58]. These findings align with the expectations under a polygenic model, which posits that individual variants typically exert minimal effects on diseases. Many studies have demonstrated the efficacy of PRS in predicting disease status within controlled case-control studies[58-60]. Moreover, the utility of PRS extends beyond these settings, proving effective in predicting disease in population-based cohorts and studies that utilize electronic health records[61,62].

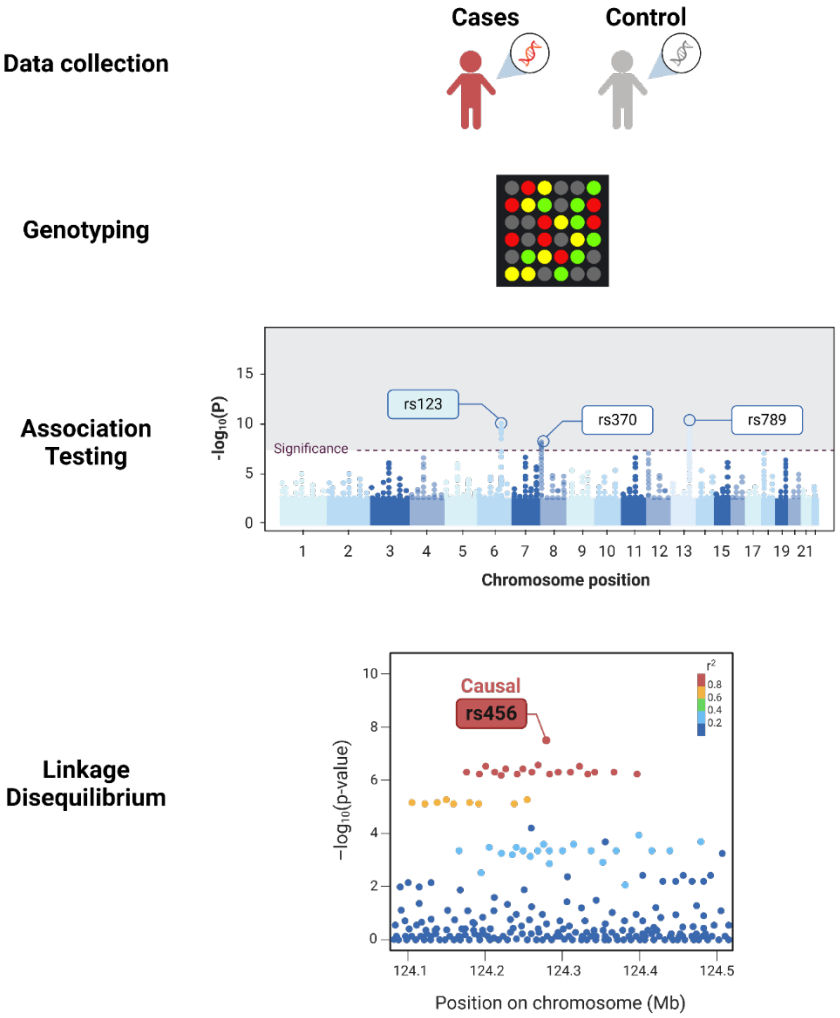


Figure 3. Principle of a GWAS. Illustrating the key steps in a GWAS to identify genetic variants associated with a specific phenotype. In the top panel, genetic data from individuals with and without the trait of interest are analyzed using high-throughput genotyping arrays to identify SNPs. The middle panel presents a Manhattan plot, where each point represents a SNP and the significance of its association (-log10(p-value)) with the trait. SNPs surpassing the genome-wide significance threshold are highlighted, with loci such as rs123, rs370, and rs789 annotated as significant. The bottom panel shows fine-mapping results for a locus of interest, where the likely causal SNP (rs456) is identified based on its high statistical significance and LD (r^2) with nearby variants, represented by the color gradient. Figure created using BioRender.

1.3.1.4 Non-coding genetic variants

More than 90% of the genetic variants identified through GWAS are located in the non-coding regions of the genome. While it is possible that these risk variants are associated with nearby causative coding variants, it is more likely that they reside within regulatory regions and affect the function of a neighboring gene.

Regulatory elements are crucial for the precise control of gene expression, responding adaptively to environmental changes or developmental cues. These elements interact with a variety of proteins, including transcription factors, which are broadly categorized into 'activators' and 'repressors'[63]. Transcription factors engage with cis-regulatory elements not only at transcription start sites, but also at distal regulatory elements that may be located thousands of base pairs away from these sites[64].

Key regulatory elements include:

1) Promoters: Promoters are DNA sequences typically found upstream of genes that recruit RNA Polymerase II to initiate transcription. In the presence of specific transcription factors, RNA Polymerase II recognizes and binds to core promoter elements, including the TATA box and other proximal elements to facilitate the assembly of the transcription initiation complex[65].
2) Enhancers: Regulatory DNA sequences that significantly increase the transcription of target genes. Bind to transcriptional activators and, despite often being located far from the TSS, can interact directly with the core promoter through the looping of DNA. Enhancers can exert their effect regardless of their position relative to the gene they regulate[66].
3) Silencers: Short DNA sequences that attract repressive proteins, known as transcriptional repressors, to downregulate gene expression. By binding these proteins, silencers can directly or indirectly interfere with the activity of promoters and other regulatory elements to decrease transcription levels[67].
4) Insulators: Function as genomic boundary elements. The CCCTC-binding factor (CTCF) is a well-known protein that binds to these sequences, organizing the genome into TADs to modulate gene expression through chromatin architecture[68].

Non-coding variants situated within enhancer regions are a major consideration for functional interpretation of GWAS loci. For instance, by examining regulatory variants within DNase I hypersensitivity sites, researchers identified rs7539120 – an enhancer variant on chromosome 1q. This variant was determined to alter cardiac function by promoting the expression of the downstream *NOS1AP* gene in cardiomyocyte-intercalated discs, elevating the risk of cardiac arrhythmias[69]. This is one of several instances where functional enhancer SNPs have been identified to regulate gene expression in critical biological pathways. In the liver, the expression of *GCKR* is controlled by a transcription enhancer regulated by FOXA2, located within a type 2 intronic locus[70].

Additionally, the *PHF11* gene, which is relevant to asthma, is transcriptionally regulated in an allele-specific manner by Oct-1[71].

Figure 4 illustrates how enhancers and silencers recruit transcription factors to upregulate or downregulate mRNA production by modulating transcriptional activity at the promoter.

A comprehensive exploration of other forms of regulatory elements, including non-coding RNAs (ncRNAs) falls outside the purview of this dissertation.

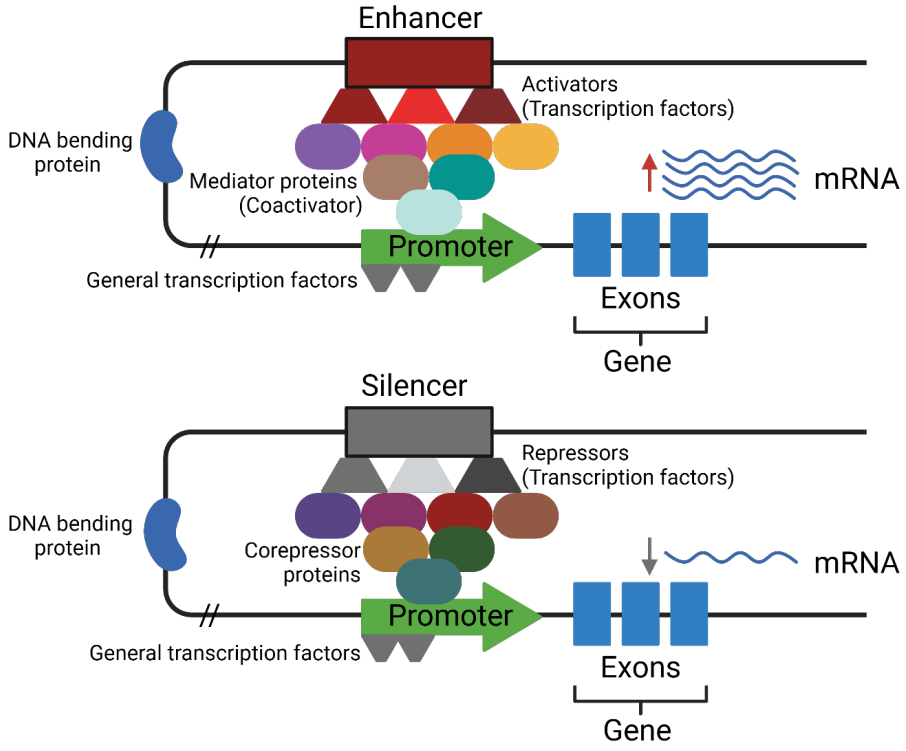


Figure 4. Mechanisms of transcriptional activation and repression via enhancers and silencers. Top: Enhancers recruit activators (transcription factors) and coactivator complexes, such as mediator proteins, to facilitate transcriptional activation. DNA bending proteins bring the enhancer into proximity with the promoter, where general transcription factors assemble with RNA polymerase II to initiate mRNA synthesis. Bottom: Silencers recruit repressors (transcription factors) and corepressor complexes, which inhibit transcriptional initiation. DNA bending proteins similarly bring silencers into proximity with the promoter, but repressor-corepressor interactions prevent the formation or activity of the transcription initiation complex, reducing mRNA synthesis. Double slashes (//) represent the spatial separation between regulatory elements and the core promoter. Figure created using BioRender.

1.3.1.5 Functional genomics to identify non-coding SNP mechanisms

Rao and colleagues[72] devised a structural approach for the functional interpretation of disease-associated non-coding variants identified

through GWAS. The procedure they proposed can be categorized into three main stages:

1) Prioritization of Candidate Variants: This is achieved through the application of statistical fine-mapping and incorporation of genomic annotations.
2) Prioritization of Causal Genes: This process leans heavily on chromatin interactions and analyzing expression quantitative trait loci (eQTL) data.
3) Experimental Validation and Functional Study: Techniques such as protein binding assays, reporter assays, and genome editing are used to confirm and explore the functional implications of the prioritized genes.

The upcoming sections will delve deeper into each of these three stages.

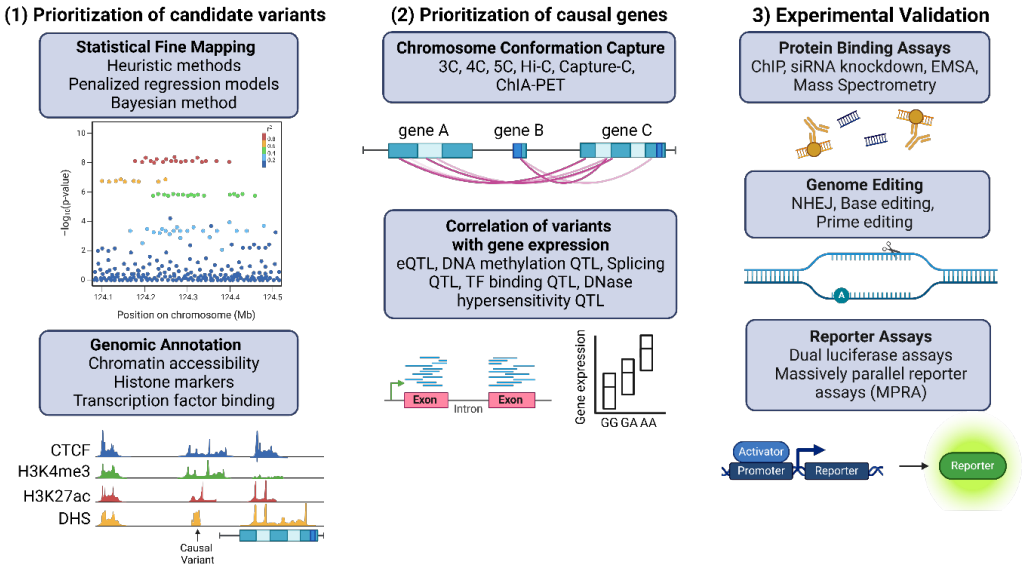


Figure 5. A typical workflow from GWAS to functional dissection of regulatory variants. This figure is adapted from Rao et al.[72]. Figure created using BioRender.

1.3.1.5.1 Prioritization of Candidate Variants

The association of a specific tag SNP with a particular trait, as identified through GWAS, may not be directly causal. Instead, it could be attributed to an indirect LD with another SNP that truly harbors the functional consequence. Fine-mapping methods assist in pinpointing causal variants by selecting and prioritizing those most likely responsible for complex traits. This is achieved by employing sequencing or dense genotyping arrays to clarify the LD patterns of SNPs. Heuristic methods are commonly used in fine-mapping, where it filters SNPs based on their LD (r^2) score with the index SNP, retaining only those with an LD score

exceeding a predefined threshold. The statistical associations can be observed using LocusZoom plots[73], which visually map the linkage strength of each SNP to the lead SNP. Additionally, the pairwise LD among SNPs within haplotype blocks can be examined using tools like Haploview[74]. This approach helps prioritize SNPs for subsequent functional studies.

While fine-mapping is instrumental in narrowing down the list of causal variants, it often yields multiple candidates that require further elucidation. GWAS variants are mostly situated within functional regulatory regions, which can be identified by mapping to enhancer markers and DNase I hypersensitivity sites, among others[12,13,75]. Consortiums such as ENCODE and the Roadmap Epigenomics Mapping Consortium (REMC) have generated diverse genome-wide datasets to explore the epigenetic landscape of non-coding elements within the human genome across various cell types. Table 1 details strategies for identifying regulatory regions through the examination of chromatin accessibility and specific histone modifications:

Histone Marks (ChIP-Seq)		
Histone	**Description**	**Regulatory element**
H3K4me1	Aids in the recruitment of chromatin remodelling complexes, promoting accessibility for transcription machinery.	Active and primer enhancers
H3K27Ac	Marks active enhancer regions and, less commonly, promoters, decreasing histone-DNA interaction and recruiting transcription machinery[76,77].	Active enhancers
H3K4me3	Acts as a recognition site for chromatin remodelling, rendering DNA more accessible and facilitating transcription initiation[78].	Active gene promoters
Chromatin Accessibility		
Assay	**Description**	**Regulatory element**
DNase I hotspot	Areas of open chromatin containing TF binding motifs[79] which are hypersensitive to DNase I.	Promoter, distal enhancer, TFBS
ATAC-seq	Measures chromatin accessibility using engineered Tn5 transposase[80].	Promoter, distal enhancer, TFBS

Table 1. Genomic annotation. Key histone markers and chromatin accessibility assays for regulatory element identification.

Single nucleotide changes can significantly influence gene regulation by altering transcription factor binding sites (TFBS). These changes can create, modify, or disrupt TFBS, thereby affecting the binding affinity of transcription factors. Such alterations can have downstream effects on

gene expression, potentially leading to changes in cellular function or contributing to disease. For instance, the non-coding variant at the 1q13 locus associated with LDL cholesterol levels and myocardial infarction lead to a new TFBS, leading to increased expression of the nearby gene *SORT1*[6]. *SORT1*, which regulates plasma lipoprotein levels, lowers LDL levels, making it a potential drug target for myocardial infarction.

To assess the impact of SNPs on TFBS, TFBS computational prediction tools can be employed. This involves scanning DNA sequences for known motifs—short, conserved sequences recognized by transcription factors. Tools such as FIMO[81] and JASPAR[82] utilize position weight matrices (PWMs) derived from experimental data to predict where transcription factors are likely to bind across the genome. By integrating SNP data, these tools can identify variants that may disrupt or create new binding sites. An example is MotifbreakR[83], which leverages ENCODE's ChIP-Seq data to identify the impact of SNPs on transcription factor motifs. These predictions are important for prioritizing SNPs that might play a regulatory role, although their accuracy hinges on the availability of comprehensive transcription factor binding data across relevant cell types.

1.3.1.5.2 Prioritize causal genes

Once candidate variants have been prioritized, the next step is to identify the genes they may affect. While coding variants can often be linked to their target genes based on proximity, non-coding variants present a greater challenge. Cis-regulatory elements like enhancers exert influence on gene transcription from a distance, utilizing chromatin looping mechanisms to interact with promoters. Consequently, chromatin interaction data is informative for identifying the physical linkages between non-coding variants and gene promoters.

Hi-C, a high-throughput technique introduced by Lieberman-Aiden et al. in 2009, captures genome-wide chromatin interactions in the nucleus[84]. Hi-C builds upon earlier chromatin conformation capture methods, including 3C, 4C, and 5C, by crosslinking DNA and proteins, digesting the DNA with restriction enzymes, re-ligating interacting fragments, and sequencing the ligation products. This method enables the mapping of long-range regulatory interactions and provides insight into how distant regulatory elements might influence gene expression. Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) further refines this approach by identifying specific chromatin interactions, particularly at TFBSs[85]. The ENCODE project provides extensive Hi-C and ChIA-PET datasets across various cell types, allowing for the prediction of regulatory relationships between distant enhancers and

promoters. However, one limitation of these datasets is that they often originate from bulk tissues, which can obscure important cell-type-specific interactions crucial for understanding disease mechanisms.

Chromatin interaction data provides insights into target genes affected by non-coding variants, revealing potential mechanisms by which these variants influence gene expression. eQTL analysis complements chromatin interaction data by linking SNPs to gene expression changes, identifying SNPs that correlate with expression differences in either nearby genes (cis-eQTLs) or distant genes (trans-eQTLs). This method is typically conducted by genotyping a large population and measuring gene expression using RNA sequencing. The Genotype-Tissue Expression (GTEx) project[86], one of the most extensive resources for eQTL data, has profiled gene expression across 53 tissues from nearly 1,000 individuals, offering insights into tissue-specific gene regulation.

The integration of chromatin interaction data with eQTL analysis is particularly powerful for identifying potential target genes influenced by GWAS-identified SNPs. However, one of the challenges remains the tissue specificity of eQTL data, as bulk tissues may not accurately represent gene expression patterns in the most relevant disease cell type.

While epigenetic modifications, chromatin interactions, and eQTL analyses are commonly used to prioritize causal variants and genes, they are not the only available tools. Advances in predictive technologies have increasingly focused on integrating multiple approaches to improve the identification of causal genes from GWAS data. For example, transcriptome-wide association studies (TWAS) combine gene expression data with GWAS findings[87], providing a more direct link between SNPs and gene function. High-resolution 3D genome mapping techniques, such as Capture-C and HiChIP[88], offer further precision by combining chromatin interaction mapping with protein binding data. By incorporating these approaches, the process of identifying the causal genes of noncoding variants in complex disease studies can be significantly enhanced.

1.3.1.5.3 Functional approaches

Protein binding assays

Most noncoding variants identified in GWAS are found within CREs, which frequently interact with DNA-binding proteins, including transcription factors. To explore how these noncoding variants affect the binding affinities of such regulatory proteins, the most commonly

employed techniques are Electrophoretic Mobility Shift Assay (EMSA)[89] and Chromatin Immunoprecipitation (ChIP)[90].

Electrophoretic Mobility Shift Assay (EMSA)

The Electrophoretic Mobility Shift Assay (EMSA) is a widely used *in vitro* technique for studying protein-DNA interactions. In this assay, a radiolabelled or fluorescently labelled DNA probe containing the sequence of interest, such as a region harbouring a SNP, is incubated with nuclear extracts or purified transcription factors. When proteins bind to DNA, they form a complex that moves more slowly during gel electrophoresis compared to unbound DNA probes. This reduction in electrophoretic mobility creates a noticeable shift in the gel's banding pattern, known as a 'shift.' This shift enables detection of specific interactions between proteins and DNA at SNP locations.

A common theme among unbiased approaches for identifying proteins interacting with noncoding variants is the use of mass spectrometry to detect unknown binding partners. One such approach is DNA-affinity pull-down assays, where oligonucleotides corresponding to alternative alleles are used to capture proteins binding to specific SNPs, which are then identified using mass spectrometry. Another technique is proteome-wide analysis, such as Stable Isotope Labeling by/with Amino Acids in Cell Culture (SILAC) first described by Ong SE et al.[91], where metabolically labeled proteins are distinguished by mass spectrometry based on isotope incorporation, allowing for identification of specific protein interactors that may be affected by noncoding variants. Additionally, mass spectrometry can be applied to analyze protein-DNA complexes from EMSA, providing insights into specific transcription factors bound to DNA. However, since peptides from different proteins—especially those within the same transcription factor family—can share similarities, these methods often require additional validation through techniques like EMSA with specific antibodies or immunoblotting for increased specificity.

Chromatin Immunoprecipitation (ChIP)

However, the methods discussed above are strictly *in vitro* techniques using DNA probes. To verify transcription factor binding *in vivo*, ChIP is a powerful method that investigates protein-DNA interactions within the chromatin context of living cells. ChIP involves crosslinking proteins to DNA using formaldehyde to preserve their interactions, followed by fragmentation of chromatin into smaller pieces. Antibodies specific to the protein of interest, such as transcription factors, histone modifications, or other chromatin-associated proteins, are used to immunoprecipitate the

protein-DNA complexes. The associated DNA fragments are purified and analyzed, allowing for the identification of genomic regions bound by the protein.

ChIP-qPCR[92] is a targeted approach where the immunoprecipitated DNA is analyzed using quantitative PCR (qPCR) to assess protein binding at specific genomic regions. This method is particularly useful for validating transcription factor binding to regions containing SNPs identified through GWAS. In contrast, ChIP-seq[93] combines ChIP with high-throughput sequencing, enabling genome-wide identification of binding sites for a given protein. This approach is invaluable for understanding how SNPs in non-coding regions affect the broader binding landscape of transcription factors. Additionally, several variants of ChIP assays have been developed for more precise target validation and higher resolution. For example, ChIP-exo[94] offers single base-pair resolution, while CUT&RUN[95] requires significantly fewer cells for analysis, making it suitable for low-input samples.

Protein binding assays like EMSA and ChIP are invaluable for elucidating the impact of SNPs on transcription factor binding and gene regulation. By directly comparing the binding of proteins to DNA sequences containing different SNP alleles, these assays can reveal whether a SNP disrupts, enhances, or creates new binding sites for regulatory proteins.

Reporter assays

Reporter assays are the gold standard experimental approaches to determine the transcriptional regulatory activity of noncoding variants. They are employed to validate the regulatory potential of sequences identified by techniques such as ChIP-Seq and to assess enhancer-promoter interactions. By introducing specific DNA sequences containing SNPs into a reporter construct—typically linked to a luciferase or fluorescent protein gene—changes in reporter gene expression can be measured to reflect the regulatory activity of the SNP-containing sequence.

In conventional reporter assays, region surrounding the SNP is cloned into a plasmid upstream of a reporter gene. The construct is transfected into a relevant cell line, and the expression of the reporter gene quantified. Differences in reporter activity between constructs containing different SNP alleles can reveal the regulatory potential of the variant. For example, a SNP within an enhancer region may enhance or suppress the activity of the enhancer, as reflected by changes in luciferase expression.

The biggest limitation of using reporter assays is that they are limited by the lack of native chromatin context, as they typically rely on episomal plasmids that may not fully replicate the chromatin environment of the endogenous locus. This can result in discrepancies between reporter assay outcomes and the actual in vivo regulatory activity of the variant. Additionally, the use of artificial constructs may not account for long-range chromatin interactions or the combined influence of multiple regulatory elements. Furthermore, conventional reporter assays often focus on single variants or short sequences, potentially missing the synergistic effects of co-inherited variants. This limitation is important in complex diseases, as regulatory elements operate within larger, interconnected networks.

Despite these challenges, reporter assays remain a critical tool for the initial functional characterization of non-coding variants, offering valuable insights into the regulatory mechanisms by which SNPs influence gene expression.

CRISPR/Cas9 genome editing

Traditional experimental approaches often fail to account for the native genomic context in which variants exist, potentially leading to misleading results due to the exclusion of chromatin architecture and cellular regulatory complexity. Genome editing, particularly the CRISPR/Cas9 system, offers a more accurate alternative by enabling targeted genetic modifications within their natural environment. This technique relies on a guide RNA (gRNA) to direct the Cas9 endonuclease to a specific DNA sequence, where it induces a double-strand break (DSB).

Following the introduction of a DSB, the cell employs its innate repair pathways to restore the DNA. The most frequently utilized repair mechanism, NHEJ, often results in random insertions or deletions (indels) that disrupt gene function. Alternatively, homology-directed repair (HDR) can be harnessed for precise edits if a donor DNA template is supplied, allowing the introduction of specific sequences at the break site.

CRISPR/Cas9 has become a powerful tool for exploring the functional roles of non-coding SNPs, particularly those that influence gene regulation. By creating targeted mutations or deletions in regulatory regions containing SNPs, researchers can investigate their effects on enhancers, promoters, or other cis-regulatory elements. A typical application involves substituting a reference allele with a disease-associated allele to examine its impact on chromatin accessibility,

transcription factor binding, or gene expression, thereby clarifying its role in gene regulation.

For instance, Gao et al.[96] applied CRISPR-Cas9 to study the non-coding SNP rs11672691, located at the 19q13 locus and linked to aggressive prostate cancer. Their edits in prostate cancer cell lines revealed that the risk allele G enhanced binding of the transcription factor HOXA2, which upregulated the lncRNA PCAT19 and CEACAM21, promoting cell proliferation and invasiveness. This demonstrates how a non-coding SNP can influence transcription factor interactions and gene activity, establishing its contribution to disease progression.

However, the CRISPR-Cas9 system has inherent limitations, including off-target editing, where Cas9 induces unintended mutations at genomic sites similar to the intended target[97]. Such off-target effects must be carefully assessed using bioinformatic prediction tools or genome-wide sequencing approaches. Additionally, the efficiency of HDR significantly depends on the length and complexity of donor DNA templates[98,99]. Longer nucleotide changes or insertions typically reduce editing efficiency, as extended templates are more prone to multimerization and can hinder the cellular repair machinery's ability to efficiently incorporate the donor sequence.

Recent advancements in CRISPR technology, such as base editing, enable precise single-nucleotide changes without inducing DSBs. This approach uses a modified Cas9 enzyme fused with a cytidine or adenosine deaminase to convert C•G to T•A or A•T to G•C base pairs, respectively, at specific genomic sites[100]. Base editing is especially valuable for modeling disease-associated SNPs, as it allows for highly accurate point mutations that closely mimic natural genetic variants.

1.3.2 Coronary Artery Disease

1.3.2.1 Disease pathogenesis

CAD remains one of the most significant causes of morbidity and mortality globally, with its progression closely linked to atherosclerosis. Atherosclerosis is a chronic inflammatory disorder characterized by the accumulation of lipids and the formation of plaques within coronary arteries, leading to compromised blood flow to the myocardium. The development of CAD involves complex processes, including endothelial dysfunction, lipid deposition, smooth muscle cell activity, and inflammation, which collectively contribute to plaque formation and destabilization[101].

The initiation of atherosclerosis often begins with endothelial dysfunction, which serves as a key trigger in disease progression[102-104]. Under normal conditions, endothelial cells regulate vascular homeostasis by promoting vasodilation, inhibiting platelet aggregation, and preventing leukocyte adhesion[105]. However, exposure to risk factors such as elevated cholesterol levels, hypertension, smoking, and diabetes causes endothelial activation. This transition shifts the endothelium from an anti-inflammatory state to one that promotes atherogenesis. A hallmark of this transition is the reduction in nitric oxide (NO) availability due to endothelial nitric oxide synthase dysfunction, alongside an increase in reactive oxygen species (ROS)[106]. Elevated ROS levels exacerbate oxidative stress, leading to the upregulation of adhesion molecules such as ICAM-1 and VCAM-1 facilitating monocyte adhesion and migration.

Once monocytes infiltrate the subendothelial space, they differentiate into macrophages, which internalize oxidized low-density lipoprotein (Ox-LDL) to form foam cells[107]. The accumulation of foam cells creates fatty streaks, recognized as the earliest detectable lesions in atherosclerosis. These streaks are further aggravated by the release of inflammatory mediators, including interleukins such as IL-1β, IL-3, and IL-8, which attract additional immune cells and amplify the inflammatory response.

Endothelial cells also undergo phenotypic changes that contribute to disease progression. Endothelial-to-mesenchymal transition (EndMT)[108] , driven by factors like TGF-β, results in endothelial cells adopting mesenchymal characteristics, promoting extracellular matrix deposition and facilitating vascular smooth muscle cell (VSMC) migration and proliferation. Additionally, endothelial-to-immune-like cell transition (EndICLT)[109] enhances inflammation through the expression of immune markers, further recruiting immune cells to the site of injury.

As atherosclerosis progresses, VSMCs move into the intima layer of the vessel wall and undergo proliferation[110]. These cells contribute to the synthesis of extracellular matrix components, including collagen and elastin, which collectively form a fibrous cap over the atherosclerotic plaque, temporarily stabilizing it. However, persistent inflammation and the breakdown of the extracellular matrix weaken the fibrous cap over time. When the cap ruptures, pro-thrombotic substances are exposed to the bloodstream, leading to the formation of a clot. This thrombus can obstruct blood flow, potentially causing acute coronary events such as myocardial infarction.

Figure 6 summarizes the multifaceted pathogenesis of CAD, which involves an interplay of endothelial dysfunction, lipid accumulation, chronic inflammation, and cellular transitions. These processes culminate in the formation and destabilization of atherosclerotic plaques, highlighting the complexity of this disease.

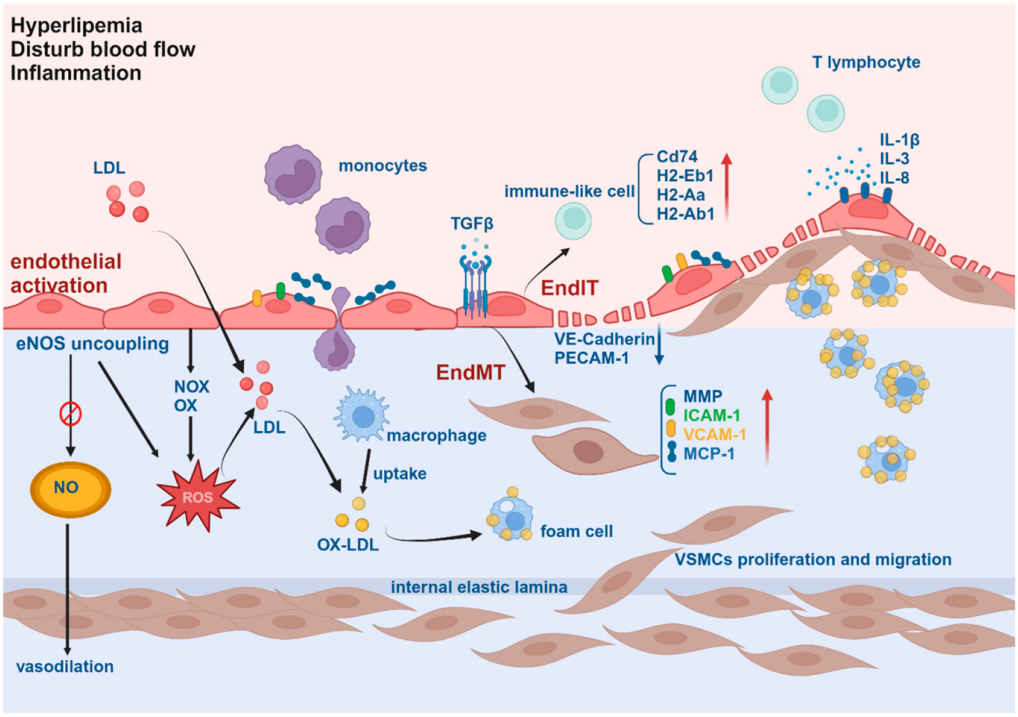


Figure 6. Pathogenesis of atherosclerosis driven by endothelial dysfunction and inflammation. Hyperlipidemia, disrupted blood flow, or chronic inflammation can activate endothelial cells. This reduces nitric oxide (NO) production, increases reactive oxygen species (ROS), and upregulates adhesion molecules like ICAM-1 and VCAM-1, promoting monocyte adhesion. Monocytes infiltrate the subendothelial layer, differentiate into macrophages, and internalize oxidized LDL, forming foam cells that create fatty streaks, the earliest atherosclerotic lesions. EndMT and EndICLT, driven by factors like TGF-β and inflammatory cytokines (e.g., IL-1β, IL-3, IL-8), further promote inflammation, smooth muscle cell migration, and fibrous cap formation. The diagram includes arrows indicating activation (black), upregulation (red), and downregulation (blue). Figure from Zhang et al.[101].

1.3.2.2 Vascular endothelial function and activation in CAD

The vascular endothelium is central to cardiovascular homeostasis, regulating processes such as vasodilation, haemostasis, inflammation, and vascular permeability[111]. In CAD, endothelial dysfunction and activation are pivotal in initiating and driving atherosclerosis.

Endothelial Dysfunction

Endothelial dysfunction refers to the impaired ability of the endothelium to maintain vascular function, marked by reduced NO production and

bioavailability[106,112]. NO is a crucial vasodilator that helps regulate vascular tone, inhibit platelet aggregation, and prevent leukocyte adhesion. In endothelial dysfunction, the endothelium loses its ability to promote vasodilation effectively, contributing to altered coronary blood flow and increased vascular resistance[113].

Endothelial dysfunction is not only a functional impairment but also an early marker of atherogenesis. It is one of the earliest measurable deteriorations in the vasculature and precedes the appearance of visible atherosclerotic plaques. When the endothelium becomes dysfunctional, it shifts from its normal anti-inflammatory, antithrombotic state to one that favours vasoconstriction, inflammation, thrombosis, and increased vascular permeability[113,114]. This process affects both the microvascular arterioles[115], leading to impaired coronary flow and myocardial ischemia, and the larger epicardial coronary arteries[116], setting the stage for plaque formation.

Furthermore, endothelial dysfunction disrupts the endothelium's ability to maintain vascular homeostasis[117]. Endothelial dysfunction also disrupts vascular homeostasis by increasing leukocyte adhesion and lipid infiltration, which promotes lesion formation[118]. Common risk factors such as hyperlipidemia, hypertension, smoking, and diabetes exacerbate endothelial dysfunction, linking it to CAD progression.

Endothelial Activation

Endothelial activation involves a transition from a quiescent to an inflammatory state[119], characterized by the upregulation of adhesion molecules (e.g., ICAM-1, VCAM-1, selectins) [120], proinflammatory cytokines (e.g., IL-6, IL-8)[121], and prothrombotic factors (e.g., von Willebrand factor, tissue factor)[122]. These changes enhance leukocyte adhesion, amplify inflammatory pathways, and contribute to plaque formation and instability.

A notable marker of endothelial activation in CAD is the presence of CECs in the bloodstream[123,124]. CECs are shed from the vascular endothelium in response to injury or activation and serve as a real-time indicator of endothelial health. Increased levels of CECs have been observed in CAD patients and correlate with disease severity[125,126]. The detection of CECs provides insight into the extent of endothelial injury and ongoing vascular inflammation, making them a potential biomarker for CAD progression.

The interplay between endothelial dysfunction and activation establishes a proinflammatory and prothrombotic environment within the coronary arteries, accelerating atherogenesis and plaque development.

1.3.2.3 Genetic architecture of CAD

GWAS has emerged as a pivotal approach for identifying genetic variants linked to CAD. By analyzing SNPs across the genome in large populations, this method has revealed numerous genetic loci associated with CAD, offering valuable insights into its complex genetic basis.

To date, over 163 risk loci for CAD have been identified[2], encompassing both common and rare variants. Notably, majority of lead-associated variants are located in non-coding regions[127], indicating a significant role for regulatory elements in CAD pathogenesis. These non-coding variants are believed to influence gene expression through mechanisms such as altering TFBS, modifying chromatin architecture, or affecting enhancer and promoter activities, ultimately impacting pathways implicated in atherogenesis.

Among the loci identified, 9p21 stands out as one of the most well-established GWAS signals for CAD. This locus has the strongest genetic contribution to coronary atherosclerosis, with variants in this region being present in approximately 75% of the global population, excluding certain African populations[128]. 9p21 variants are specifically associated with the risk of a first coronary heart disease (CHD) event (hazard ratio [HR] = 1.19; 95% confidence interval [CI]: 1.17–1.22) rather than subsequent events, suggesting their primary contribution to the initiation of atherosclerosis[128,129]. The CAD-associated region on 9p21 overlaps with the *ANRIL (antisense non-coding RNA)*. Lo Sardo et al.[130] showed that deletion of the entire 9p21 ~60 kb risk locus in iPSC-derived VSMCs resulted in decreased *ANRIL* levels. Additionally, the locus is situated near *CDKN2A* and *CDKN2B*, but a direct functional mechanism linking these genes to CAD has not been clearly established. However, chromatin conformation studies suggest that long-range interactions between 9p21 and the *CDKN2A/B* region may regulate distant gene expression[131].

Another important locus identified in GWAS for CAD is 1p13, which has been strongly associated with plasma LDL-C levels[132]. Variants at this locus influence the expression of the *SORT1* gene in the liver, affecting lipoprotein metabolism and modulating myocardial infarction (MI)[6]. Functional studies integrating eQTL and protein QTL data have further demonstrated how these variants alter lipid levels and contribute to CAD.

The *PHACTR1* locus on chromosome 6p24 represents another well-studied region. Variants within this locus are associated with multiple vascular disorders, including CAD, hypertension, and arterial

dissection[133]. The lead SNP in this region has been linked to *endothelin-1 (EDN1)* expression, a distal gene in the same 6p24 region, which is a potent vasoconstrictor involved in atherogenesis and vascular tone regulation[7]. This suggests that variants within the 6p24 locus may affect vascular biology primarily through the regulation of *EDN1*. Experiments using iPSC-derived endothelial cells have shown that the SNP at this locus acts as a vascular-specific non-coding regulatory element influencing *EDN1* levels, providing a potential mechanistic link between genetic variation and endothelial vasculature.

While GWAS has successfully identified numerous loci associated with CAD, the effect sizes of individual variants are typically modest, highlighting the polygenic nature of the disease. CAD appears to result from the cumulative influence of multiple genetic variants, each contributing incrementally to disease risk. These findings emphasize the heterogeneity of CAD, with genetic variants affecting diverse biological pathways such as lipid metabolism, inflammation, vascular remodeling, and endothelial function. Further exploration of gene-environment and gene-lifestyle interactions will be essential for unraveling the complex pathogenesis of CAD and identifying novel therapeutic targets.

1.3.2.4 Asian-specific risk loci for CAD

Genetic variants hold promise as biomarkers for disease risk, as they can be quantified at birth, long before clinical risk factors emerge. PRS for CAD, which aggregate multiple risk alleles, have shown potential in enhancing risk prediction, particularly in populations of European descent[134]. Some of the most notable European CAD cohorts include CARDIoGRAMplusC4D[135], which consists of meta-analyses from over 60,000 cases and 120,000 controls, and the UK Biobank[136], a large-scale biomedical database with genetic data from 500,000 participants across the UK. These European-based studies have identified numerous risk loci for CAD, uncovering both common and rare genetic variants.

However, PRS tools exhibit significantly lower predictive power in non-European populations[15-17]. This discrepancy arises from differences in variant frequencies, effect sizes, and LD patterns across ancestries. As European cohorts dominate GWAS datasets, they fail to fully represent the genetic diversity of global populations. This limitation is particularly significant given that nearly half of the world's cardiovascular disease burden is estimated to occur in Asia. Addressing this gap necessitates the development of ancestry-specific genetic tools and precision medicine strategies to improve CAD prediction and treatment in underrepresented populations.

The BioBank Japan (BBJ) project[137] has provided critical insights into CAD genetic risk factors specific to East Asian populations, including several novel loci not observed in European cohorts. BBJ, one of the largest biobanks in the world, includes genomic data from over 200,000 individuals. Through GWAS conducted using BBJ data, variants like rs564427867 (*PCSK9*), rs13306206 (*APOB*), rs879255211 (*LDLR*), rs746959386 (*LDLR*) and rs778408161 (*LDLR*) have been identified as new CAD loci in the Japanese population but are not significantly replicated in the European-based CAD Genome-Wide Replication and Meta-Analysis (C4D) and UK Biobank (UKBB) cohorts[138]. These findings highlight the genetic differences that may underlie CAD susceptibility between populations.

A particularly notable variant identified in the BBJ cohort is rs112735431 (RNF213) on chromosome 17q25. This nonsynonymous substitution (p.Arg4810Lys) in *RNF213* is significantly associated with CAD in the Japanese population (OR = 2.11, p = 0.017)[139]. Interestingly, this variant is also a known susceptibility locus for moyamoya disease[140], a rare cerebrovascular condition prevalent in East Asia, illustrating the pleiotropic effects of certain variants in this population.

In contrast, loci such as *9p21*[141,142] and *PHACTR1 (6p24)*[47,133] have been identified in both European and East Asian populations, demonstrating that some CAD-associated variants are shared across ancestries. However, the discovery of population-specific loci, as seen in the BBJ study, highlights the need to expand research in diverse populations to create more accurate prediction tools and improve precision medicine approaches. The findings from the BBJ project, therefore, are invaluable in understanding the unique genetic architecture of CAD in Asian populations and contribute to the global effort to improve risk prediction and disease prevention.

1.3.2.5 Rs6903956 identified as a Chinese-Han specific GWAS risk locus for CAD

Given the underrepresentation of East Asian populations in CAD genetic research, identifying East Asian-specific loci associated with CAD is a critical starting point for understanding ancestry-specific disease mechanisms.

One such variant of particular interest is rs6903956, located within the first intron of *ADTRP* at the 6p24.1 locus. First published in 2011 by Wang and colleagues[18], this variant was identified in a GWAS conducted on the Chinese Han population, where it showed a significant association with CAD risk but was not observed in European cohorts.

The association was highlighted in a three-stage GWAS study ($p = 4.87 \times 10^{-12}$, OR = 1.51 for the A allele) in a combined population of 3,240 cases and 4,353 controls. Interestingly, despite a higher MAF of the A allele in Europeans (37.8%) compared to East Asians (7.3%), it was not associated with CAD in European populations[143]. This highlights the potential ancestry-specific role of rs6903956 in CAD susceptibility.

Subsequent studies in other Asian populations, including Singaporean and Japanese cohorts, have further validated the association between rs6903956 and CAD[19-22]. The locus at 6p24.1 itself is particularly noteworthy as it has been previously identified as one of the most implicated CAD risk loci[47,135], the *PHACTR1* risk locus which lies approximately 500 kb downstream.

Wang et al.[18] also provided evidence for the functional relevance of rs6903956, showing that the minor risk allele A is associated with decreased *ADTRP* mRNA expression in leukocytes. This eQTL analysis was performed in two independent groups: group 1 (GG: n=50, AG/AA: n=54; p=0.004) and group 2 (GG: n=344, AG/AA: n=67; $p=3.56 \times 10^{-6}$), supporting a potential functional mechanism linking rs6903956 to *ADTRP*, the closest gene to this variant. More recently, a study demonstrated that binding of the transcription factor GATA2 to a 519 bp region containing the non-risk allele G enhances *ADTRP* expression levels in HeLa cells[144]. This suggests that rs6903956 may act as an enhancer for *ADTRP*. However, there remains limited data regarding the functional role and mechanisms of rs6903956 in disease-relevant cell types.

SNP rs6903956 lies within the first intron of the *ADTRP*, which plays a key role in endothelial cell protection by regulating tissue factor pathway inhibitor (TFPI), thereby contributing to anti-coagulant activity[145]. Studies in ADTRP-deficient mouse and zebrafish models exhibit hallmarks of atherosclerosis such as vascular leakage, extracellular matrix degradation, and localized inflammation[146]. Additionally, *ADTRP* has been shown to hydrolyzing fatty acid esters of hydroxy fatty acids (FAHFAs), signaling lipids with known anti-inflammatory effects[147,148]. In humans, lower plasma levels of ADTRP have been reported in CAD patients compared to control subjects[149].

Despite evidence linking rs6903956 to *ADTRP* and its potential role as an enhancer, a significant gap remains in understanding its functional mechanisms in disease-relevant cell types. Further studies are needed to determine how rs6903956 influences *ADTRP* expression or that of other genes implicated in CAD, and how this contributes to CAD pathogenesis. Such investigations will provide deeper insights into the

regulatory mechanisms at the 6p24.1 locus. Addressing this knowledge gap is essential for validating rs6903956 as a functional variant and uncovering its role in East Asian-specific CAD risk.

1.4 Aims

Considering the critical gaps in understanding the functional mechanisms of rs6903956 and its role in CAD pathogenesis, this dissertation aims to uncover the mechanistic underpinnings of this regulatory variant. By adopting a structured, multi-disciplinary approach inspired by the techniques outlined by Rao and colleagues[72], this study seeks to bridge the connection between regulatory SNPs, transcription factor interactions, causal gene expression, and disease phenotypes.

The overarching aim of my work is to **dissect the precise mechanistic pathway** linking rs6903956 to CAD development, ultimately identifying the causal regulatory variant, its downstream targets, and its contribution to disease pathogenesis. The central hypothesis is that the CAD-associated variant rs6903956 contributes to vascular dysfunction by disrupting a non-coding regulatory element, thereby altering gene expression in vascular cell types involved in atherogenesis.

This dissertation is structured into five key aims:

1. **To delineate the regulatory potential of rs6903956** through statistical fine-mapping and comprehensive genomic annotations, elucidating its regulatory potential within the 6p24.1 locus. This aim is addressed in Chapter 2.
2. **To establish a disease-relevant model** for interrogating the functional role of rs6903956 using CRISPR-mediated gene editing in iPSC-derived CAD-relevant cell types, such as endothelial and smooth muscle cells. This aim is addressed in Chapter 3.
3. **To identify and validate the causal gene(s)** influenced by rs6903956, providing insights into the downstream pathways that drive CAD pathogenesis. This aim is addressed in Chapters 4 and 5.
4. **To identify regulatory factors mediating the effects of rs6903956 on potential effector genes** and experimentally validate its role in transcriptional regulation, establishing a direct link between the variant and its impact on gene expression. This aim is addressed in Chapter 5.
5. **Establish cellular phenotypes associated with rs6903956** in disease-relevant cells. This aim is addressed in Chapter 5.

By integrating genomic techniques, functional modelling, and precise validation in disease-relevant systems, this dissertation will offer a comprehensive mechanistic understanding of rs6903956. These

findings will provide critical insights into East Asian-specific CAD risk and inform future precision medicine strategies.

Chapter 2: Endothelial-specific vascular effect of rs6903956

Chapter 2: Endothelial-specific vascular effect of rs6903956

2.1 Chapter background

In this chapter, we investigate the regulatory role of rs6903956 in CAD using a combination of *in silico* genomic and population-based approaches. Specifically, we hypothesize that rs6903956, a non-coding variant within the 6p24 locus, acts as a regulatory element influencing gene expression in cell types critical to CAD pathogenesis.

To test this hypothesis, we integrate fine-mapping, ENCODE data analysis, and population-based cardiovascular cohort studies to comprehensively characterize the role of rs6903956. Fine-mapping will be employed to pinpoint the causal variants at the 6p24 locus, while ENCODE data will provide insights into the regulatory mechanisms of rs6903956. Furthermore, population-specific biobank resources will be used to investigate the mechanistic pathways linking rs6903956 to vascular phenotypes, including endothelial function (e.g., flow-mediated dilation) and arterial stiffness in smooth muscle cells.

By combining these approaches, we aim to establish a detailed understanding of the regulatory function of rs6903956, laying the groundwork for future experimental validation and mechanistic studies.

2.1.1 Fine-mapping

GWAS identifies significant associations through "tag" SNPs which act as markers for genetic variation. However, these tag SNPs are not necessarily the causal variants responsible for the observed phenotypic effects. This complexity arises from LD, a phenomenon where neighboring genetic variants are inherited together due to co-segregation during recombination[37]. As a result, multiple variants within a genomic region may exhibit correlation, making it challenging to pinpoint the specific causal variant driving the association.

To address this challenge, we employ statistical fine-mapping methods in post-GWAS analyses to visualize the LD structure at 6p24.1 locus. Fine-mapping prioritizes variants that are more likely to have a direct biological role by integrating genetic association signals and specific LD patterns[150]. Fine-mapping can help determine whether rs6903956 plays a causal role in CAD or merely reflects a correlation with other functional variants.

A critical component of fine-mapping is the consideration of population-specific LD structures, which can vary markedly among different ancestral groups[151]. Databases such as HapMap Phase II[39] and the

1000 Genomes Project[47] allow for exploring LD patterns specific to populations, including Europeans, East Asians, and Africans.

Given that rs6903956 was initially identified in a GWAS conducted on the Chinese Han population[18], analysing LD patterns unique to East Asian populations enables us to investigate whether the regulatory mechanisms of rs6903956 are influenced by distinct LD structures that differ from those in other populations. For instance, LD patterns in the East Asian population may reveal stronger correlations between rs6903956 and functional variants at the locus, suggesting that one of these tightly linked variants could be the actual causal variant. This population-specific focus could provide insights into the genetic architecture underlying CAD in different ancestral groups.

2.1.2 Large-scale biobank resources

Since rs6903956 was discovered in a GWAS conducted on the Chinese Han population, it is crucial to investigate its clinical phenotypes in populations with East Asian-specific ancestry. To achieve this, we utilize large-scale biobank resources such as MESA[152]. Cardiovascular cohort studies provide a valuable opportunity to study genotype-to-phenotype relationships within populations that include significant representation of individuals with East Asian ancestry.

MESA is a multi-ethnic cohort study involving over 6,000 participants, with approximately 12% identifying as Chinese American. The primary goal of MESA is to understand the development and progression of subclinical atherosclerosis and cardiovascular disease across diverse populations. Key data collected in MESA include imaging markers of subclinical atherosclerosis, endothelial dysfunction biomarkers, inflammatory markers, and longitudinal clinical outcomes.

By analyzing data from MESA, we aim to identify phenotypes associated with rs6903956 that are specific to East Asian populations. These phenotypes focus on arterial-specific traits, including markers of endothelial dysfunction such as ICAM-1, VCAM-1, E-selectin, and flow-mediated dilation (FMD), as well as markers of arterial stiffness, including pulse wave velocity (PWV) and carotid intima-media thickness (cIMT).

Identifying specific phenotypes is crucial for designing focused experimental models, particularly when selecting the appropriate cell types to study variants like rs6903956. The relevance of cell-type specificity is important for understanding complex diseases like CAD[153]. Each cell type harbours a unique chromatin landscape and transcription

factor repertoire[154], which are instrumental in determining how non-coding variants interact with regulatory elements to influence gene expression. For CAD, key cell types include endothelial cells[155], smooth muscle cells[156], and mononuclear cells, which play direct roles in vascular function, inflammation, and remodeling.

The utility of population biobanks in confirming the relevant cell type for CAD risk variants has been demonstrated in numerous studies. For instance, rs17114036 at 1p32.2, located in an endothelial-specific enhancer, has been linked to mechanotransduction mechanisms[8], initially identified from an eQTL study showing decreased *PPAP2B* expression in human aortic endothelial cells (HAECs). Similarly, rs9349379 in the 6p24 locus, within the *PHACTR1* gene, has demonstrated cis-regulation of *EDN1* expression in endothelial cells[7]. The endothelial-specific role of rs9349379 was identified through a significant association between this SNP and arterial stiffness in the UK Biobank, with the G allele linked to reduced flow-mediated vasodilation in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. Such approaches provide a basis for determining the appropriate cell types to study when investigating the regulatory effects of rs6903956.

2.1.3 Genomic annotations from ENCODE project resources

The identification of a SNP's regulatory role typically starts with data mining tools before progressing to experimental models. The ENCODE project[12], an international initiative funded by the National Human Genome Research Institute (NHGRI), offers a comprehensive catalog of functional elements within the human genome. These elements include enhancers, promoters, and TFBS functioning at both protein and RNA levels. As an open-access resource, ENCODE aids in pinpointing non-coding SNPs that could affect gene expression and regulatory mechanisms, thereby facilitating the prioritization of candidates for experimental validation to delineate their roles in disease pathways.

Non-coding variants, including rs6903956, often influence gene expression through mechanisms involving enhancers, silencers, or insulators. These regulatory elements may disrupt transcription factor binding, chromatin looping, or splicing patterns, significantly altering gene regulation. Recent studies suggest that rs6903956 may act as an enhancer. Luo et al. transfected a 519bp region spanning rs6903956 and a 1513bp promoter/regulatory region of *ADTRP* into HeLa cells and found higher *ADTRP* promoter activity with preferential binding of transcription factor GATA2 to rs6903956 with common allele G[144]. They further showed in human umbilical vein endothelial cells (HUVECs) that

GATA2 siRNA knock down resulted in lower ADTRP protein levels, but a direct enhancer-promoter interaction remained to be proven. This regulatory function can be further investigated by analysing histone modification markers, such as H3K4me1 and H3K27ac, which indicate active enhancer regions.

By investigating regulatory information from ENCODE, we can better understand the regulatory roles of non-coding variants like rs6903956 and their impact on CAD-related phenotypes. This approach establishes a foundation for identifying causal regulatory elements and guiding targeted experimental models.

2.2 Methodology

2.2.1 Fine-mapping

Regional association plots were generated for a ~1 kb region centered on the query variant rs6903956, using the LDproxy tool (GRCh37). Proxy variants with R^2>0.2 were identified as in LD with rs6903956 and highlighted accordingly. Calculations were performed on 1000 Genomes Phase 3 populations, including East Asian, European, African, American, and South Asian cohorts, ensuring consistency with LDlink methodology.

2.2.2 Phenome-Wide Association Studies (PheWAS)

To study whether rs6903956 is associated with cardiovascular or metabolic phenotypes in a population setting, we utilized large biobank PheWAS study: UK Biobank (http://www.ukbiobank.ac.uk/). Data was accessed using web-based PheWAS sites (https://pheweb.org/UKB-TOPMed/), which reports genome-wide associations for 1,419 ICD-based phenotypes in white British individuals, imputed using the TOPMed reference panel.

We downloaded all available phenotype-level summary statistics (n=1,419) directly from the API for rs6903956. We applied Bonferroni and Benjamini-Hochberg False Discovery Rate (FDR) corrections across all 1,419 phenotypes to account for multiple testing.

2.2.3 MESA study population and phenotypic characterization

MESA (dbGAP study accession phs000420.v6.p3) is a cohort study aimed at identifying factors contributing to subclinical CVD in adults aged 45–84 years. Details on the study design are available from Bild et al.[152].

FMD, a measure of vascular function, was assessed using ultrasound imaging[157]. A standard protocol included inflating a blood pressure cuff to 50 mmHg above systolic pressure for 5 minutes, followed by imaging to capture the vasodilatory response. Absolute FMD (mm) was calculated as the difference between maximum and baseline diameters, while relative FMD (%) was determined as ((maximum diameter − baseline diameter)/baseline diameter) × 100%. We reported associations with both mm FMD and %FMD. mm FMD is not correlated with baseline diameter, which might confound results. However, %FMD was also included as it is a common parameter shown to predict cardiovascular events.

2.2.4 Retrieval of individual level genotypic information from MESA cohort and statistical analysis

Genotypic data were obtained from the database of Genotypes and Phenotypes (dbGaP) as the MESA SNP Health Association Resource (SHARe) project (study accession phs00209). Briefly, genotype calls were downloaded in matrix format and extracted from dbGAP using SRA toolkit. Affymetrix gene annotation files were used to map for rs6903956 and individual genotypes extracted using PLINK[158].

Univariate analysis of covariance (adjusted for age and gender) was performed to evaluate the association between rs6903956 genotypes with mm FMD and %FMD. Analysis was done using IBM SPSS Statistics Version 28.0.1.0.

2.2.5 Epigenetic prediction using ENCODE database

Histone marks and CTCF ChIP-seq datasets in the 6p24 locus were visualized using the Integrative Genomics Viewer (IGV). H3K27ac and H3K4me3 ChIP-seq data were downloaded from GEO accession (GSE131681)[159], while HUVEC CTCF ChIP-seq data was obtained from the ENCODE Project (accession ENCSR000DLW). All signal tracks represent IGV-scaled coverage data, in which signal intensities are normalized to the dynamic range within the specific genomic window to enable local visualization of enrichment around rs6903956.

HUVEC DNase I footprinting data was similarly visualized in IGV using data from GEO accession GSM1014528.

2.3 Results

2.3.1 Fine-mapping identifies a tandem repeat, rs140361069, as the strongest LD signal with rs6903956 in East Asian populations

GWAS for CAD have implicated the variant rs9349379 at the PHACTR1 locus on chromosome 6p24[7]. To determine whether rs6903956 is in LD with rs9349379, we performed genetic fine-mapping using SNP associations from the East Asian population in 1000 Genomes Project (Figure 7). There was no significant LD with the PHACTR1 lead variant. Instead, genetic fine-mapping using the LDlink web tool[160] identified six variants nearby (± 600bp) with strong LD for rs6903956 ($R^2 > 0.8$), including a short tandem repeat, rs140361069. The 'A' risk allele at rs6903956 was associated with a deletion of $(AAAT)_3$ at rs140361069 (Table 2). Importantly, this association exhibited ethnicity-specific effects, as LD between rs6903956 and rs140361069 was observed only in East Asian populations, while no significant LD was detected in European populations.

These findings highlight the potential role of rs140361069 in mediating the observed effects of rs6903956 in East Asian cohorts.

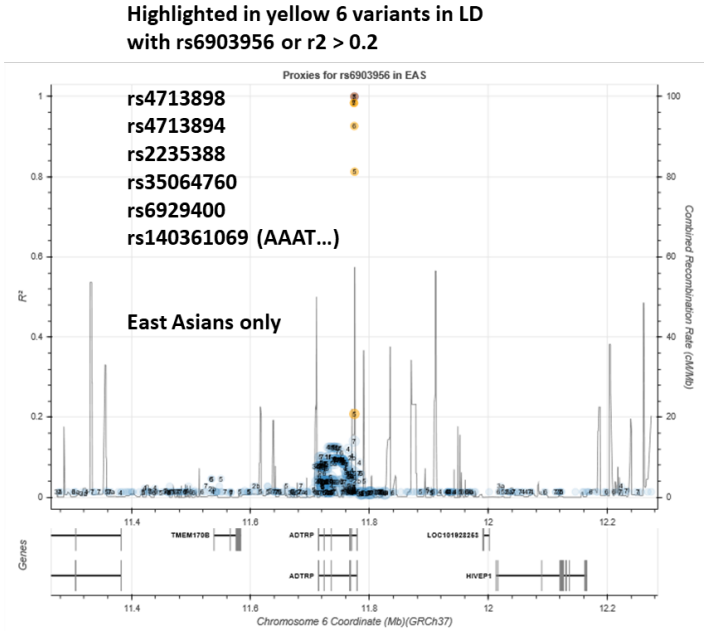


Figure 7: Regional association plot for ~1 kb region centered on query variant rs6903956 in CAD GWAS of East Asian Population from LDproxy tool[160]. Proxy variants with $R^2 > 0.2$ were highlighted and identified as in LD with rs6903956.

Populations		Variants in LD with rs6903956 in East Asian ($R^2 > 0.2$)						
	RS	**rs6903956**	rs4713898	rs4713894	rs2235388	rs35064760	rs6929400	**rs140361069**
	Distance	0	-133	-554	-873	-198	209	-39
	Alleles	A=A,G=G	A=G,G=C	A=T,G=C	A=G,G=C	A=-,G=A	A=C,G=T	A=-,G=AAATAAATAAAT
East Asian	**R2**	1	1	0.9842	0.9842	0.9266	0.8126	0.2081
	MAF	0.0675	0.0665	0.0665	0.066	0.0724	0.0794	0.2579
European	**R2**	1	0.9539	0.7135	0.9539	0.9787	0.4916	-
	MAF	0.3767	0.3738	0.3101	0.3738	0.3748	0.3867	-
American	**R2**	1	0.9624	0.7315	0.9499	0.9624	0.6069	-
	MAF	0.3487	0.3487	0.2939	0.3487	0.3516	0.33	-
African	**R2**	1	0.987	0.3044	0.9454	0.9774	-	0.3271
	MAF	0.3661	0.3661	0.1536	0.3578	0.3699	-	0.4236
South Asian	**R2**	1	0.9649	0.6867	0.9649	0.8991	0.4997	-
	MAF	0.1759	0.1708	0.1278	0.1708	0.183	0.1697	-

Table 2. Proxy Variants in LD with rs6903956 ($R^2 > 0.2$) Segregated by Populations. The tandem repeat rs140361069, which is in LD with rs6903956, is specific to East Asian and African populations, highlighting population-specific genetic variation.

2.3.2 PheWAS reveals association of rs6903956 with hypertension

Phenome-wide association studies (PheWAS) offer an unbiased approach to exploring relationships between genetic variants and a wide spectrum of disease phenotypes. We searched for rs6903956 in the UK Biobank PheWAS, a large prospective cohort of ~500,000 individuals aged 40-69 across the United Kingdom[136]. Filtering for circulatory system phenotypes revealed associations with essential hypertension (p=0.023) and hypertension (p=0.025) (Table 3). However, following correction for multiple testing across 1,419 phenotypes using Bonferroni and Benjamini-Hochberg FDR methods, these associations no longer remain statistically significant (Bonferroni p=1.0, FDR p=0.922). These findings therefore need to be interpreted with caution. Notably, no significant association with CAD was observed, likely due to the underrepresentation of East Asian populations in this cohort, where rs6903956 has shown stronger effects.

Category	Clinical Diagnosis	pval	p_bonferroni	p_fdr	Effect size (SE)	Number of samples
Circulatory system	Coronary atherosclerosis	0.28	1.0	0.99	0.012 (0.011)	20454 / 374655
Circulatory system	Essential hypertension	0.023	1.0	0.92	0.015 (0.0066)	77465/328796
Circulatory system	Hypertension	0.025	1.0	0.92	0.015 (0.0066)	77714/328796
Circulatory system	Disease of capillaries	0.05	1.0	0.99	0.23 (0.12)	144/3987056
Circulatory system	Nonrheumatic tricuspid valve disorders	0.073	1.0	0.99	0.25 (0.14)	105/400662
Circulatory system	Varicose veins of lower extremity, symptomatic	0.087	1.0	0.99	0.10 (0.058)	623/367909

Table 3. PheWAS UK Biobank clinical diagnosis associated with rs6903956 filtered for circulatory system. Effect sizes are reported with their standard errors (SE), and the number of samples is presented as cases/controls. Both Bonferroni (p_bonferroni) and FDR-adjusted p-values (p_fdr) are included to account for multiple comparisons across 1,419 tested phenotypes.

2.3.3 Rs6903956 'A' risk allele is associated with elevated diastolic blood pressure

Given the associations of rs6903956 with cardiometabolic traits in previous studies, we investigated traditional cardiometabolic risk factors in a cohort of 1,027 individuals from Guilin, Guangxi, China. Our analysis revealed a significant association between the number of 'A' alleles at rs6903956 and elevated diastolic blood pressure (DBP), with a regression coefficient of 2.31 ($p = 0.045$), suggesting potential allele-specific effects on blood pressure phenotypes (Table 4).

Notably, no significant associations were observed for other traditional risk factors, highlighting a potentially specific effect of rs6903956 on DBP. The lack of association with lipid profiles or glucose-related traits suggests that the mechanistic link between rs6903956 and cardiometabolic risk may involve alternative pathways, independent of classical metabolic factors.

Cardiometabolic risk factors from Guilin, Guangxi, China		
Outcome	**Effect size (SE)**	**p-value**
LDLC	-0.007 (0.084)	0.929
TG	0.178 (0.125)	0.153
HDLC	-0.044 (0.041)	0.287
GLU	0.125 (0.167)	0.454
HbA1C	0.082 (0.105)	0.431
SBP	-0.093 (2.839)	0.974
DBP	**2.310 (1.151)**	**0.045**

Table 4. Association of cardiometabolic risk factors with outcomes in a population from Guilin, Guangxi, China. Effect sizes with standard errors (SE) are reported for each outcome, alongside p-values. Outcomes include: systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI), fasting plasma glucose (FPG), triglycerides (TG), total cholesterol (TC), and high-density lipoprotein cholesterol (HDL-C).

2.3.4 Association of rs6903956 with relevant subclinical phenotypes

Variations in blood pressure are often accompanied by vascular adaptations, including structural remodelling, endothelial dysfunction, and heightened vascular reactivity. To further investigate the vascular-specific effects of rs6903956, we performed a meta-analysis using data from the MESA cohort[152], a study of subclinical cardiovascular disease risk factors involving 6,814 asymptomatic individuals across four ethnic groups: 38% White, 28% African American, 22% Hispanic, and 12% Asian.

We specifically analysed vascular function and structural parameters in 566 Chinese American participants within MESA. Associations of rs6903956 with arterial stiffness, carotid intima-media thickness (CIMT), and FMD were examined (Table 5). Arterial stiffness measurements—including aortic distensibility (β=0.113, p=0.118), large artery elasticity index (β=0.039, p=0.422), and small artery elasticity index (β=−0.017, p=0.726)—revealed no significant associations. Similarly, rs6903956 showed no significant association with CIMT (β=0.048, p=0.385), suggesting limited impact on vascular stiffness and structural markers in this cohort.

Notably, the rs6903956 A allele was significantly associated with reduced % FMD compared to the G allele (A allele: 3.959% [95% CI: 3.313, 4.605]; G allele: 4.812% [95% CI: 4.585, 5.038]; p=0.015, dominant model) (Figure 8A). A similar association was observed for mm FMD (A allele: 0.163 mm [95% CI: 0.138, 0.187]; G allele: 0.197 mm [95% CI: 0.188, 0.205]; p=0.010, dominant model) (Figure 8B).

FMD, a standardized measure of endothelial function, assesses the vascular response to endothelial nitric oxide release during reactive hyperaemia. Decreased % FMD in individuals carrying the A allele reflects compromised endothelial responsiveness, highlighting the potential role of rs6903956 in endothelial dysfunction among Chinese Americans.

Interestingly, no significant association between rs6903956 and % FMD was observed in White Caucasians (GG: 4.639% [95% CI: 4.327, 4.910]; GA+AA: 4.807% [95% CI: 4.598, 5.017]; p=0.335) or Hispanics (GG: 4.303% [95% CI: 4.039, 4.567]; GA+AA: 4.437% [95% CI: 4.025, 4.669]; p=0.454). However, in African Americans, the A allele was associated with increased % FMD (GG: 3.286% [95% CI: 2.974, 3.597]; GA+AA: 3.766% [95% CI: 3.506, 4.026]; p=0.021), contrasting with the effects observed in Chinese Americans. These findings suggest population-specific molecular mechanisms underlying rs6903956's effects on endothelial function.

Taken together, our findings implicate rs6903956 as a modulator of endothelial function, particularly in Chinese Americans. The observed reductions in FMD associated with the A allele provided a key direction for our study, leading us to prioritize endothelial cells for regulatory data mining and mechanistic studies aimed at understanding the functional genetics of rs6903956 and its role in CAD.

Physiological Measurement	Model	Beta	t	Sig
%FMD	Additive	-0.092	-2.072	**0.039**
	Dominant	-0.108	-2.445	**0.015**
FMD	Additive	-0.091	-2.099	**0.036**
	Dominant	-0.112	-2.579	**0.010**
Aortic distensibility	Additive	0.113	1.570	0.118
	Dominant	0.113	1.570	0.118
Large artery elasticity indices	Additive	0.052	1.064	0.288
	Dominant	0.039	0.804	0.422
Small artery elasticity indices	Additive	-0.017	-0.342	0.733
	Dominant	-0.017	-0.351	0.726
CIMT	Additive	0.042	0.753	0.452
	Dominant	0.048	0.870	0.385

Table 5. Associations of rs6903956 genotypes with physiological measurements of vascular function in 566 Chinese American participants from the MESA cohort. Association of rs6903956 genotype with arterial stiffness (aortic distensibility, large artery elasticity index, small artery elasticity index), carotid intima-media thickness (CIMT), and FMD (%FMD and mm FMD) under dominant and additive models, using univariate analyses of covariance adjusted for age and gender. Standardized coefficients (Beta), t-values, and p-values are presented, with genotype distributions of rs6903956 individuals with GG = 504, GA = 59, and AA = 3.

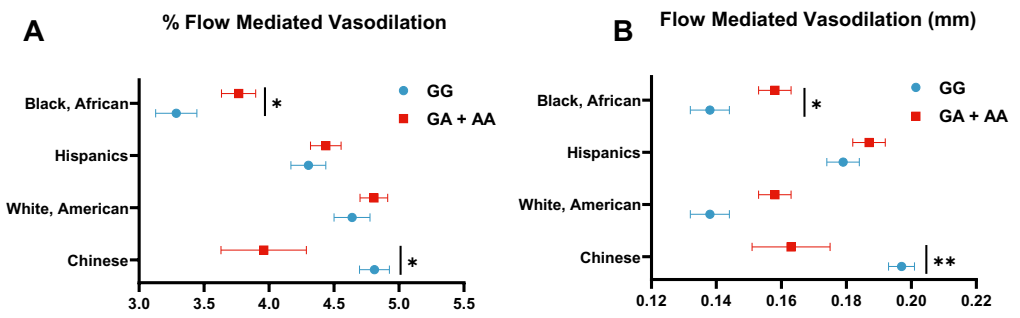


Figure 8. rs6903956 risk allele association with endothelial dysfunction in multi-ethnic study. (A and B) Univariate analysis of covariance for %FMD (A) and mm FMD (B) for rs6903956 genotypes following dominant model, stratified according to population. Line graph showing means with 95% CI, $^{**}p \leq 0.01$, adjusted for age and gender (GG n individuals = 504, GA+AA n individuals = 62, for Chinese population).

2.3.5 Regulatory role of rs6903956

Having identified endothelial cells as the relevant cell type for rs6903956, we sought to investigate its potential regulatory roles using *in silico* analyses. Alternative splicing graphs indicated that rs6903956 does not lie within splice site boundaries, ruling out RNA splicing disruption as a mechanism for disease causation (Figure 9A). Furthermore, rs6903956 was found outside of CTCF sites, suggesting it is unlikely to play a role in chromatin looping or boundary insulation (Figure 9B).

We next examined regulatory chromatin regions surrounding rs6903956 by leveraging ChIP-Seq data from Nakato et al., who profiled nine types of human endothelial cells to map the vascular epigenomic landscape[159]. Analysis of histone modification signals associated with active enhancers (H3K27ac) and promoters (H3K4me3) revealed weak enrichment across upper body endothelial cells (HPAEC, HCCaEC, HaoEC, HCoAEC, and HENDC) within a 1 kb region centred on rs6903956 (Figure 9C). These findings suggest a potential role as a weak enhancer in endothelial cells.

Variant effect predictions categorized rs6903956 primarily as an intronic variant (71%), with minor probabilities of being a regulatory region variant (14%) or a nonsense-mediated mRNA decay (NMD) transcript variant (14%) (Figure 9D). This further supports a potential regulatory role for the variant, rather than one related to coding or splicing disruptions.

Regulatory variants typically impact gene expression by altering TFBS, either by disrupting, creating, or modulating the affinity of these sites[161]. To examine whether rs6903956 plays a regulatory role, we explored its

potential to serve as a binding site for transcription factors and mediate cooperative enhancer-like activity.

To explore whether rs6903956 functions as a TFBS, we analysed DNase I foot printing data from HUVEC cells, part of the ENCODE Project, which identified rs6903956 as a critical nucleotide located within a DNase hypersensitive site (DHS) (Figure 2A) (GEO accession: GSM1014528). Regions lying within DHS suggest potential hallmarks of active transcription factor binding and regulatory activity[162].

Taken together, these findings indicate that rs6903956 resides within a region with weak enhancer-like activity and potential regulatory significance in endothelial cells.

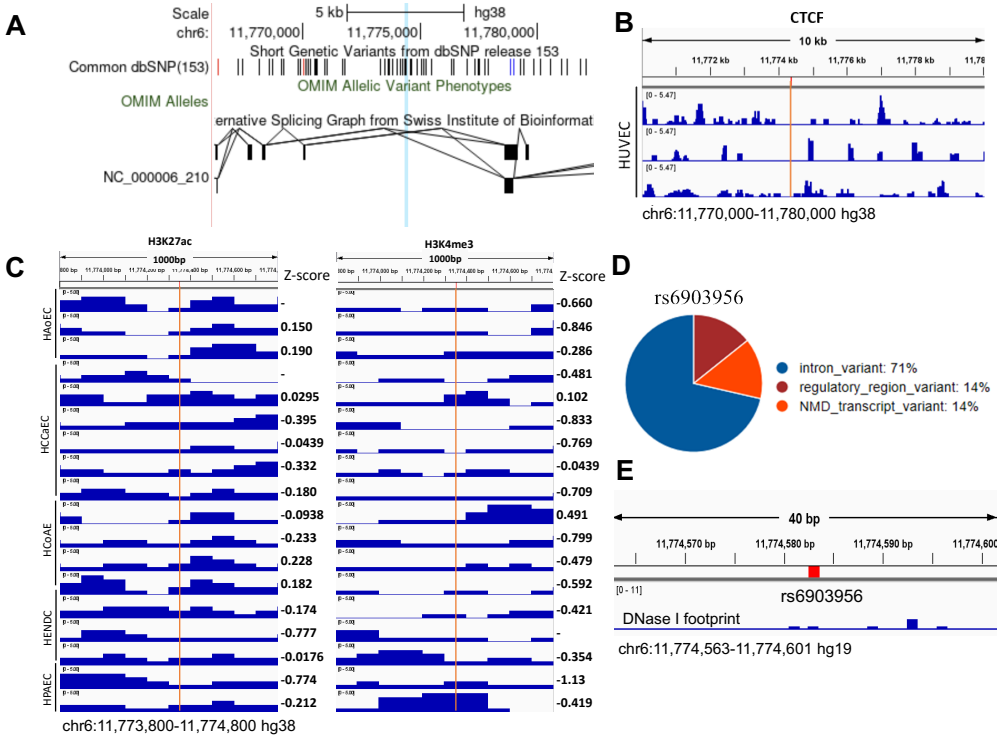


Figure 9. Potential regulatory role of rs6903956. (A) Visualization of UCSC alternative splicing sites located on ADTRP. Location of rs6903956 is annotated at in blue. **(B)** Visualization of CTCF ChIP-Seq on hg38 chr6:11,770,000-11,780,000 of HUVEC. Location of rs6903956 is annotated in red. **(C)** Visualization of H3K4me3 and H3K27Ac ChIP-Seq on hg38 chr6:11,773,800-11,774,800 in upper body endothelial cells from Nakato et al.[159]. Location of rs6903956 is annotated in red. **(D)** Prediction of rs6903956 variant role by Ensembl Variant Effect Predictor. **(E)** DNAse I footprinting of HUVEC on hg19 chr6:11,774,563-11,774,601. Location of rs6903956 is annotated in red.

2.4 Discussion

In Chapter 2, we present a detailed analysis of rs6903956, focusing on its potential regulatory role and cell-type-specific implications in endothelial dysfunction. By integrating fine-mapping, population-based analyses, and functional annotations, we identified rs6903956 as a variant of interest for further investigation.

Fine-mapping demonstrated that rs6903956 is in LD with rs140361069, a tandem repeat variant specific to East Asian populations. This association was not observed in European cohorts, highlighting the importance of population-specific genetic architectures in the interpretation of GWAS data. The strong LD (D' ≈ 1.0) in East Asians indicates a regional mechanism potentially mediating the effects of rs6903956.

Non-coding variants like rs6903956 often affect gene regulation in a cell-type-specific manner. Most prior studies on rs6903956 used leukocytes for eQTL analyses[18], which are accessible but may not fully represent the variant's functional role in endothelial cells. For example, endothelial-specific effects of rs9349379 at the *PHACTR1* locus were identified through targeted studies in human aortic endothelial cells, emphasizing the need for disease-relevant cellular models[7]. Consistently, our findings link rs6903956 with reduced FMD, a functional measure of endothelial activity, without significant effects on structural vascular parameters or traditional cardiometabolic markers. These results, supported by analyses in multi-ethnic cohorts like MESA, establish endothelial cells as the relevant context for rs6903956.

One limitation of the MESA analyses is that our association models were adjusted only for age and gender, without including additional cardiometabolic confounders such as BMI, smoking status, or cholesterol levels. It is therefore important to interpret the observed associations with caution, as residual confounding may have influenced the strength or direction of the effects.

Functional assays by Luo et al.[144] provide evidence for the regulatory role of rs6903956. Their dual luciferase assays in HeLa cells showed that a 519 bp region flanking rs6903956 enhances transcriptional activity of the SV40 promoter in the presence of GATA2. While this suggests a regulatory function, validation in endothelial cells is required to confirm its physiological relevance.

Through data mining of publicly available ChIP-Seq datasets, we observed that rs6903956 resides within regions exhibiting modest

enrichment of histone marks H3K27ac and H3K4me3 in endothelial cells. Notably, the signal intensities at the 1kb visualized locus are subtle when compared to genome-wide distributions, however, consistent enrichment across multiple endothelial-relevant cell types suggests that rs6903956 may function as a weak, cell-type-specific enhancer in vascular endothelium. Further experimental validation in endothelial contexts is warranted to elucidate the regulatory significance of rs6903956.

These findings highlight the challenges of interpreting non-coding variants, where regulatory effects may differ across cell types and populations. Our workflow, integrating fine-mapping, epigenomic data, and functional validation, addresses these complexities and provides a model for prioritizing and studying causal variants. Large-scale datasets like the UK Biobank and ENCODE further enhance this process by enabling comprehensive contextualization of regulatory variants.

Chapter 2 identifies rs6903956 as a cell-type-specific modulator of endothelial function, underscoring the importance of studying regulatory variants within their appropriate biological contexts. This chapter establishes a foundation for understanding the role of rs6903956 in cardiovascular disease and highlights the importance for targeted functional genomics approaches to elucidate GWAS variant function.

Chapter 3: Establishment of CRISPR-edited (Δ63-89bp including rs6903956) iPSC-derived endothelial cell model and whole transcriptomic profiling

Chapter 3: Establishment of CRISPR-edited (Δ63-89bp including rs6903956) iPSC-derived endothelial cell model and whole transcriptomic profiling

3.1 Chapter Background

In Chapter 2, we identified rs6903956, a CAD-associated SNP, as a potential regulatory variant influencing vascular function. Non-coding SNPs like rs6903956 present challenges in functional genomics, as they often exert subtle regulatory effects in a cell-type-specific and context-dependent manner. Our findings from MESA suggested that rs6903956 may play a role in endothelial dysfunction, a hallmark of CAD. However, its precise functional mechanisms remain unclear, necessitating further investigation in a controlled and disease-relevant model.

To explore the role of rs6903956 in endothelial biology, we generated iPSCs derived from individuals carrying either the rs6903956 risk A or non-risk G alleles. To isolate the functional role of rs6903956, we perform CRISPR-Cas9 to precisely delete the SNP in iPSCs, followed by differentiation into endothelial cells. This strategy ensures that the genetic background remains identical across edited and unedited lines, isolating endothelial phenotypic effects solely attributable to rs6903956. Additionally, we targeted rs140361069, a tandem repeat in weak LD with rs6903956, identified in Chapter 1 as potentially modulating the regulatory effects of rs6903956 in a population-specific manner. This chapter establishes a CRISPR-edited iPSC-derived endothelial cell model with a 63–89 bp deletion encompassing rs6903956 and rs140361069.

3.1.1 Generation of human iPSCs from PBMCs

iPSCs are very useful for investigating human-specific regulatory mechanisms, as they can be differentiated into cell types relevant to various diseases while preserving the donor's genetic and epigenetic profiles. These cells are a type of pluripotent stem cell generated by reprogramming adult somatic cells into a pluripotent state.

iPSCs were first introduced by Shinya Yamanaka and colleagues[163] in 2006, who demonstrated that somatic cells could be reprogrammed using a combination of four transcription factors: Oct4, Sox2, Klf4, and c-Myc. This technology provided a significant breakthrough, enabling the generation of pluripotent cells with the same differentiation potential as embryonic stem cells (ESCs), but without the ethical concerns tied to using human embryos.

The reprogramming process typically involves the delivery of these factors into somatic cells. Viral vectors, such as retroviruses and lentiviruses, are commonly used for this purpose due to their high efficiency in gene transfer[164,165]. The use of integrating vectors, however, poses a risk of insertional mutagenesis, which can potentially lead to oncogene activation or tumour formation. As a response to these safety concerns, alternative non-integrating methods have been developed[166,167], including the use of adenoviruses, episomal plasmids, and the Sendai virus, as well as synthetic mRNA and protein transduction approaches. These methods prevent genetic alterations to the host genome, enhancing the clinical applicability of iPSC technology.

Verification of successful iPSC generation involves stringent testing to ensure their pluripotency and genetic integrity. Morphological assessment of iPSC colonies, which should exhibit tight cell clusters with defined borders and a high nucleus-to-cytoplasm ratio, is typically the first step[168]. Molecular characterization includes the expression analysis of pluripotency markers such as SSEA-4, Tra-1-60, Nanog, and Oct4[169]. Additionally, the silencing of exogenous transgenes used during reprogramming must be confirmed to ascertain the iPSCs' self-reliance for maintaining pluripotency[170]. Functional validation involves in vitro differentiation into three germ layers or in vivo teratoma formation assays[171], which demonstrate the cells' ability to differentiate into various cell types corresponding to ectoderm, mesoderm, and endoderm, affirming their pluripotent status.

3.1.2 Validation of differentiated endothelial cells

Human iPSCs have revolutionized our approach to modeling and treating cardiovascular diseases. These cells can differentiate into a range of critical cell types, including endothelial cells[172-174] and smooth muscle cells[175], which are foundational for personalized medicine, including disease modeling, drug screening, and cell replacement therapy. Endothelial cells, in particular, are critical mediators of vascular homeostasis and are directly implicated in the pathogenesis of CAD[176]. Their dysfunction, often characterized by increased oxidative stress, inflammation, and impaired repair capacity, is a key contributor to the progression of atherosclerosis[177].

To explore the role of rs6903956 in CAD, we utilize iPSC-derived endothelial cells as a model system. Our protocol[178] begins with the induction of lateral plate mesoderm, the precursor to vascular lineages, using fibroblast growth factor 2 (FGF2) and bone morphogenetic protein 4 (BMP4) to direct the iPSCs towards a mesodermal lineage essential for forming vascular structures. We also apply a phosphoinositide 3-

kinase inhibitor (LY294002) to modulate the PI3K/AKT pathway, enhancing the efficiency of this induction.

After a five-day period of mesodermal induction, we enhance endothelial specification with a combination of transforming growth factor β (TGF-β) inhibitor SB431542, which helps maintain the specificity of endothelial differentiation by preventing the signaling that might lead iPSCs into non-endothelial lineages like smooth muscle cells. The addition of vascular endothelial growth factor (VEGF) promotes the maturation and survival of endothelial cells.

After another five days, the cells sorted based on CD144+ expression to achieve a high degree of purity. These mature ECs are expanded on collagen I-coated surfaces in endothelial growth media (EGM-2). Validation of these differentiated ECs includes demonstrating their ability to form tube-like structures on Matrigel and respond to inflammatory stimuli—key functional characteristics of native endothelial cells. Endothelial markers such as CD31, CD144, and von Willebrand factor are analyzed to confirm the endothelial identity of the cells, ensuring they retain critical endothelial functions indicative of a successful differentiation protocol.

3.1.3 CRISPR-Cas9 editing

CRISPR-Cas9 technology provides an efficient framework for linking genetic variants identified by GWAS to functional outcomes in disease contexts, such as in CAD. For instance, this technology has been utilized to explore the impact of the regulatory role of rs17114036[8], a CAD-associated variant, in modulating *PLPP3* expression in human endothelial cells.

The CRISPR mechanism employs a gRNA that directs the Cas9 nuclease to specific genomic sites, creating DSBs[179]. The predominant repair mechanisms include NHEJ and HDR. NHEJ, while more common, tends to introduce errors such as insertions or deletions at the repair site, potentially leading to gene disruption. In contrast, HDR provides a more accurate repair process but is less efficient than NHEJ, posing challenges for precise gene editing.

In Chapter 3, we employed a specialized pMIA3 plasmid[180] for dual-gRNA CRISPR editing. This plasmid harbors an enhanced version of Cas9, known as SpCas9, driven by the EF1-A promoter for high-efficiency gene editing. It includes dual gRNA duplex oligos aimed at specifically targeting the rs6903956 SNP and an adjacent tandem repeat rs140361069, excising a segment spanning 63-89bp. The

plasmid also integrates mRuby2, linked via a P2A peptide to SpCas9, which facilitates the identification of successfully transfected cells through fluorescence. Moreover, the inclusion of a dominant negative mutant p53 (mtp53) enhances iPSC survival following double-strand breaks, optimizing the overall efficiency of the editing process. After transfection, selected clones are expanded and analyzed, with Sanger sequencing used to confirm the edits and assess any off-target effects.

Recently, advancements in CRISPR technology[181] have led to the development of more precise editing techniques such as base editing[100] and prime editing[182]. Base editing, facilitated by a deactivated Cas9 (dCas9), allows for the direct alteration of DNA bases without inducing double-strand breaks, thereby minimizing unintended genomic changes. Prime editing further enhances editing precision by employing a reverse transcriptase linked to a Cas9 nickase, enabling the introduction all possible base-to-base conversions without significant DNA cleavage. Future experiments can consider utilizing these systems to correct the risk SNP and interrogate the precise role of rs6903956.

In this chapter, we utilized a dual-gRNA strategy within the pMIA3 plasmid to target both rs6903956 and rs140361069, to induce a large chunk deletion of 63-89bp long. This model enables our functional interpretation of the 6p24.1 risk loci for detailed phenotypic and functional studies, enabling experiments such as chromatin accessibility assays, transcription factor binding analyses, and reporter assays to elucidate the regulatory mechanisms of these loci. By leveraging this robust system, we aim to advance the understanding of how rs6903956 contributes to endothelial dysfunction and CAD pathogenesis.

3.2 Methodology

3.2.1 Patient enrolment and study approval

This study received approval from the Local Ethics Committee of Nanyang Technological University Singapore Institutional Review Board (IRB18/09/02 and IRB-2020-09-011), the National Healthcare Group (DSRB: 2013/00937), and the Agency for Science, Technology and Research (IRB Reference 2020-096) in Singapore. It adheres to the principles of the Helsinki Declaration. Prior to participating, all individuals provided written informed consent after being briefed on the objectives and potential implications of the research.

Our study builds upon data from the 'Singapore Coronary Artery Disease Genetics Study,' which involved genotyping rs6903956 alleles in CAD patients from the National University Hospital angiography centre. CAD patients were identified as non-ST-elevation myocardial infarction (NSTEMI) cases via angiography, while control subjects were selected based on self-reported medical histories.

3.2.2 Peripheral blood mononuclear cell isolation from whole blood

Each subject provided 10 ml of fresh blood drawn into heparin vacutainers. Following centrifugation using Ficoll-PaquePremium (GE Healthcare, catalog no.17-5442-03), the buffy coat layer containing PBMCs was isolated. For cryopreservation, PBMCs were suspended in heat-inactivated FBS (Thermo Fisher Scientific Life Sciences, catalog no. 10082139) with 10% DMSO, and frozen at -80 °C.

3.2.3 Sendai virus reprogramming to generate iPSCs from CAD patient PBMCs

iPSCs were derived from donor PBMCs using the CytoTune-iPS Sendai Reprogramming kit (Thermo Fisher Scientific Life Sciences, catalog no. A16518, following the manufacturer’s protocol. Cryopreserved PBMCs were thawed and seeded at a density of 5×10^5 cells/ml in complete PBMC medium for 4 days, followed by transduction with CytoTune-iPS Sendai reprogramming vectors. iPSC colonies meeting quality criteria were selected, and passaged for further expansion.

3.2.4 Induced pluripotent stem cell maintenance and characterization

iPSCs were cultured on Matrigel-coated plates (Corning, catalog no. 354230) in mTeSR1 medium, with passages carried out every 4 to 5

days using ReLeSR (StemCell Technologies, catalog no. 05872). Our iPSCs underwent comprehensive characterization through immunofluorescence, karyotyping, and teratoma formation assays. Karyotyping analysis was conducted by the cytogenetics lab at Singapore General Hospital, while in vivo teratoma formation assays were performed at the animal facility of the Biological Resource Centre of ASTAR Singapore. Following teratoma extraction, samples were forwarded to the Advanced Molecular Pathology Lab of IMCB (ASTAR) for the preparation of paraffin blocks and sectioning. Subsequent H&E staining enabled the validation of pluripotency based on the presence of three germ layers in our iPSCs.

3.2.5 Endothelial cells derived from lateral plate mesoderm differentiation

We adhered to previously established protocols for the differentiation of endothelial cells derived from lateral plate mesoderm, as outlined by Cheung et al.[183] and Narmada et al.[178].

3.2.6 Immunohistochemistry

Immunostaining was performed on both iPSCs and iPSC-derived endothelial cells for characterization. Cells were fixed with 4% paraformaldehyde and permeabilized in 0.1% Triton X-100 for 10 minutes at room temperature. Following fixation, the cells were incubated with primary antibodies, anti-human PECAM1 (Abcam, catalog no. ab32457), anti-human VWF (Abcam, catalog no. ab9378), anti-human eNOS (Abcam, catalog no. ab5589), anti-human Sox2 (R and D Systems, catalog no. AB2018), anti-human NANOG (R and D Systems, catalog no. AF1997), diluted in 1% bovine serum albumin (BSA) blocking solution for one hour under dark conditions. Secondary antibodies, Alexa Fluor 488 and 568, were applied to visualize the staining, and images were acquired using a ZEISS Celldiscoverer7 microscope.

3.2.7 Endothelial characterization

iPSC-derived endothelial cells were characterized using flow cytometry and immunocytochemistry. Cells were stained with anti-human PECAM1 (CD31; BioLegend, catalog no. 102507) or anti-human ICAM-1 (CD54; BioLegend, catalog no. 353111) in PBS with 2% FBS for 30 minutes. Flow cytometry was performed using a BD LSRFortessa X-20 analyzer, and data were analyzed using FlowJo v10.7.1. Functional assays included tube formation tests, where 25,000 cells/well were seeded on Matrigel in EGM-2 medium. Human coronary artery endothelial cells

(HCAEC, ATCC, catalog no. PCS-100-020) were used as positive controls, and iPSCs served as negative controls. Images were captured with a Nikon Ti-E microscope, and structure quantification was conducted using ImageJ's Angiogenesis Analyzer. Interleukin-8 (IL-8) secretion was measured via ELISA (Abcam, catalog no. ab100575). iPSC-derived endothelial cells were cultured at 100,000 cells/well in 2 mL EGM-2 for 72 hours. Media were collected post-centrifugation at 13,000g for 10 minutes to remove debris. IL-8 levels were normalized to total protein, quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, catalog no. 23227).

3.2.8 CRISPR/Cas9 NHEJ guide selection and cloning

Cas9 gRNAs with predicted cleavage sites were identified using a computational algorithm scoring system available at https://benchling.com/. A NHEJ double sgRNA-guided Cas9 system was employed. Validation of sgRNA cleavage efficiency was performed using co-transfection of HEK293T cells with pMIA3 and pCAG-EGxxFP-target plasmid. A 528bp region flanking rs6903956 and rs140361069 is inserted between the EGFP fragments and this construct is used as a target plasmid. Two gRNA duplex oligos were cloned and ligated with the pMIA3 plasmid (Addgene #109399) using T4 DNA ligase (New England Biolabs, catalog no. M0202S).

3.2.9 CRISPR/Cas9 NHEJ 6p24.1 genome editing in iPSCs

Prior to nucleofection, 1.5 × 106 cells were pretreated with 10 μM ROCK inhibitor (ROCKi) (StemCell Technologies, catalog no. 72302) for one hour. A single-cell suspension was prepared using accutase (StemCell Technologies, catalog no. 07922) and resuspended in 100 μl of nucleofection solution along with 10 μg of pMIA3 plasmid containing the selected gRNA pairs. Nucleofection was carried out using the Amaxa4D nucleofector (Lonza, catalog no. AAF-1002B) and P3 primary kit (Lonza, catalog no. V4XP-3024) according to the manufacturer's instructions. The nucleofected cells were then plated on Matrigel-coated 6-well plates using mTeSR medium supplemented with CloneR. After 48 hours, the cells were treated with ROCK inhibitor and dissociated into single cells. Fluorescence-activated cell sorting (FACS) was utilized to enrich for targeted cells, with RFP-positive cells being sorted and plated onto 6-well plates containing mTeSR with CloneR, penicillin-streptomycin (P/S), and gentamicin. Colony formation was observed 8 days after FACS. Subsequently, 24 single colonies were manually picked into 12-well plates, amplified, and split with RelesR onto 6-well plates. The remaining cells were utilized for genomic DNA extraction and

genotyping. Clones exhibiting successful CRISPR targeting were expanded.

3.2.10 Gel electrophoresis and sanger sequencing

Polymerase Chain Reaction (PCR) was carried out using Q5® High Fidelity DNA Polymerase (New England Biolabs, Ipswich, Massachusetts, United States, https://www.neb.sg/, catalog no. M0491S) to amplify 582 base pair regions flanking rs6903956. PCR amplification was performed under universal cycling conditions. Amplified PCR products were run on a 1% agarose gel and the DNA band observed at 582bp was cut out and extracted using FavorPrep™ Tissue Genomic DNA extraction miniprep kit (Favorgen, Taiwan, http://www.favorgen.com/, catalog no. FATGK 000). Clones with successful deletion were verified via Sanger sequencing (Bio Basic Asia, Singapore, https://www.biobasic-asia.com/).

Primer sequences used in chapter 3 are as follows:
rs6903956 F: 5'-GCCCATTCACGGAATGTCA-3'
rs6903956 R: 5'-CGTCGTCATCACCCTCAACT -3'

3.2.11 Statistical analysis

All statistical analyses were performed using GraphPad Prism version 9. Data normality was evaluated using the Shapiro-Wilk test. For datasets with a normal distribution, unpaired two-tailed Student's t-tests were used to compare two groups, and one-way ANOVA followed by Tukey's post-hoc tests was applied for comparisons involving more than two groups. For non-normally distributed data, appropriate non-parametric tests, such as the Mann–Whitney U test for two-group comparisons or the Kruskal-Wallis test for multiple groups, were utilized. Statistical significance was defined as $p < 0.05$. Results are presented as mean ± standard deviation (SD). Specific statistical tests applied to individual experiments are detailed in the corresponding figure legends.

3.3 Results

3.3.1 Donor-derived hiPSCs express classical pluripotency markers and validation of pluripotency through in vivo teratoma formation assay

We generated donor-derived hiPSCs from individuals carrying the rs6903956 risk variant (AA) associated with CAD. Control individuals, who carried the rs6903956 non-risk variant (GG), were generously provided by Assoc Prof Adrian Teo's laboratory. Gender, age, and ethnicity were carefully matched between CAD and control individuals to minimize potential confounding variables (Table 6). To ensure a comprehensive understanding of the genomic context in 6p24.1, we also genotyped all SNPs found to be in LD with rs6903956 (Table 7).

To validate the pluripotency and quality of the hiPSCs, we performed immunostaining for classical pluripotency markers Sox2 and Nanog, confirming their robust expression in both CAD and control lines (Figure 10A). TRA-1-60 staining confirmed the maintenance of an undifferentiated state in hiPSC colonies (Figure 10B). Relative expression analysis against the well-established H9 cell line revealed comparable levels of core pluripotency markers (Oct3/4, Nanog, Sox2, and Dppa) in CAD and control hiPSCs, further ensuring the consistency and quality of the lines used (Figure 10C).

Additional validation was conducted using the in vivo teratoma formation assay, a gold-standard test for pluripotency. This assay demonstrated that both CAD and control hiPSCs were capable of differentiating into tissues representative of all three germ layers, confirming their multipotency (Figure 10D). Moreover, karyotyping revealed no chromosomal abnormalities in either CAD or control hiPSCs, indicating that the lines were genomically stable and appropriate for use in subsequent analyses (Figure 10E).

Collectively, these results comprehensively validate the pluripotency, genomic stability, and differentiation potential of the hiPSC lines generated in this study.

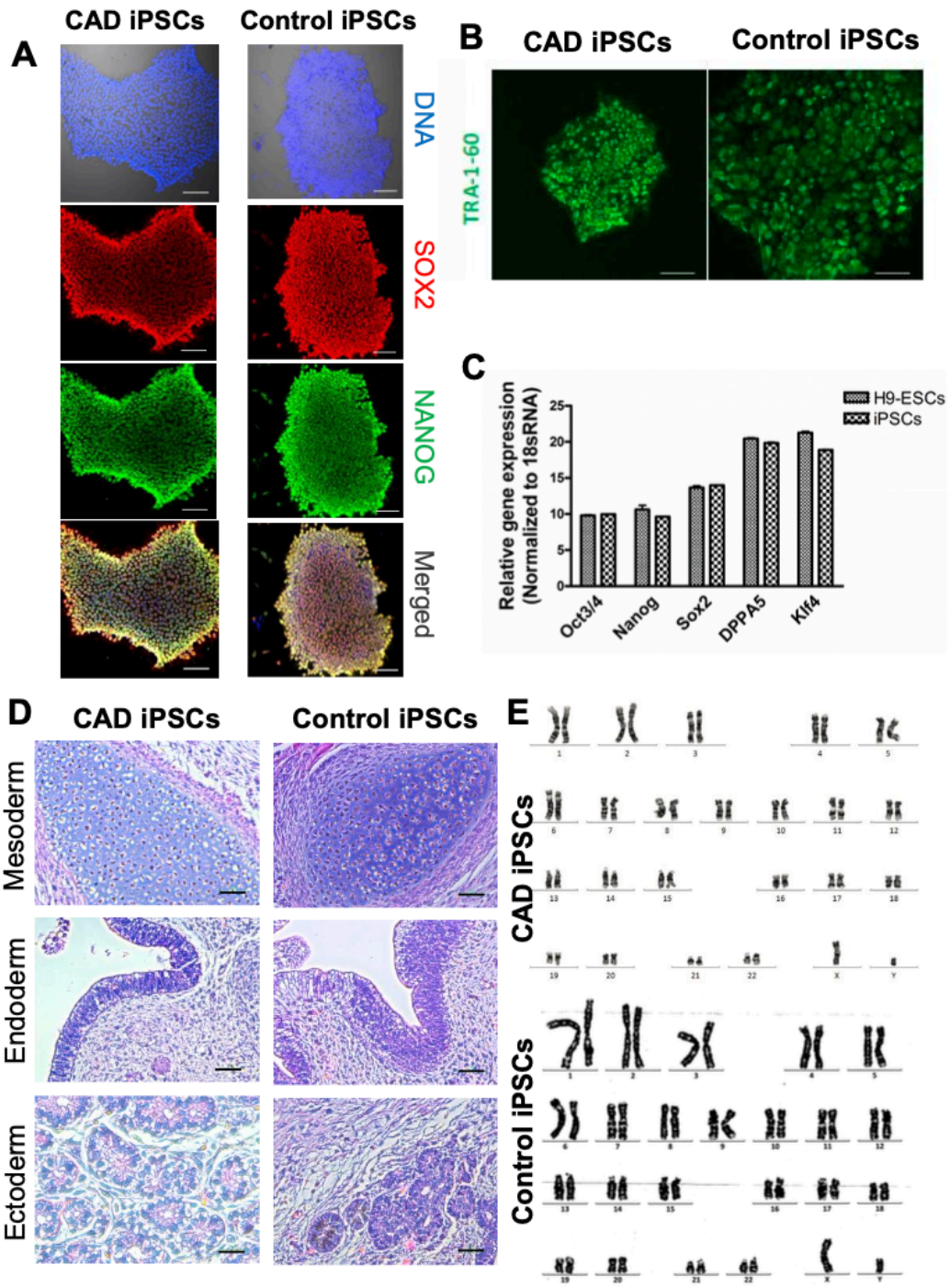


Figure 10. Generation and characterization of donor-derived hiPSCs. (A) Immunostaining of pluripotency markers, SOX2 and NANOG, on iPSC colonies (scale bar, 100 µm). **(B)** Immunostaining of undifferentiated iPSC maintenance markers, TRA-1-60, on iPSC colonies (scale bar, 100 µm). **(C)** Relative gene expression of iPSCs for classical pluripotency markers (Oct3/4, Nanog, Sox2, NPPA5, Klf4) normalized to 18S rRNA with H9-ESCs as a positive control**. (D)** *In vivo* teratoma formation assay of derived iPSCs. H&E staining of teratomas for differentiated tissues of three distinct germ layers (scale bar, 25 µm). **(E)** Karyotyping of CAD iPSC and control iPSC to establish genetic stability of cell lines based on chromosome numbers and structural features.

Staining and qPCR done by Nicole Min Qian Pek and Dr Dalakrishnan Chakrapani Narmada.

Disease status	Donor IDs	Diagnosis	Ethnicity	Gender	Age	rs6903956 allele
CAD	07	NSTEMI	Chinese	M	68	AA
	BBCC11	NSTEMI	Chinese	M	63	AA
Control	173	Normal	Chinese	M	58	GG
	i70b	Normal	Chinese	M	56	GG

Table 6. Information of donors used in this study. Donors are ethnicity, gender and age matched. CAD donors have AA rs6903956 genotype and control donors have GG rs6903956 genotype.

SNP id	CAD 07	CAD BBCC11	Control 173	Control i70b
rs2235388	GG	GG	CC	CC
rs4713894	TT	TT	CC	CC
rs35064760	AA	AA	A	AA
rs4713898	CC	CC	CC	CC
rs140361069	$-(AAAT)_7$	$-(AAAT)_7$	$-(AAAT)_{10}$	$-(AAAT)_{10}$
rs6903956	AA	AA	GG	GG
rs6929400	TT	TT	TT	TT

Table 7. Genotypes of all donors for SNPs in LD with rs6903956.

3.3.2 iPSC endothelial cells characterization

In Chapter 2, we demonstrated that the rs6903956 A risk allele is associated with decreased FMD, suggesting that this variant contributes to endothelial dysfunction. To further investigate its role in a disease-relevant context, we differentiated hiPSCs derived from CAD and control individuals into endothelial cells using our previously established protocol[178]. This approach provides a robust and physiologically relevant model for studying endothelial function.

To mimic the microenvironment of coronary vasculature, the hiPSCs were differentiated through a lateral plate mesoderm population, a critical precursor for cardiac lineages, under chemically defined conditions. After 10 days of differentiation, we achieved a high-purity population of PECAM-1 expressing endothelial cells (>98%), as confirmed by flow cytometry. Magnetic-activated cell sorting (MACS) was employed to further enrich PECAM-1 positive cells, ensuring a homogenous endothelial population for downstream experiments (Figure 11A, B).

To validate the endothelial identity of these cells, we analysed the expression of mature endothelial markers PECAM1, VE-cadherin, VWF, and NOS3. Both CAD and control endothelial cells exhibited marker expression levels comparable to HCAECs, demonstrating successful differentiation into mature endothelial cells (Figure 11C). Functional

validation through tube formation assays further confirmed the angiogenic potential of the iPSC-derived endothelial cells, as all lines formed robust cord-like structures (Figure 11E).

We also evaluated the expression of *ADTRP*, the gene on which rs6903956 resides, in iPSC-derived endothelial cells. As *ADTRP* is a relevant gene for CAD and is implicated in endothelial biology, its expression was assessed to confirm the functional relevance of the model. *ADTRP* was detected in iPSC-derived endothelial cells and positive control HCAECs, but it was absent in undifferentiated iPSCs, confirming that its expression is specific to endothelial lineage differentiation (Figure 11D).

These findings establish that the iPSC-derived endothelial cells faithfully recapitulate key features of the endothelial phenotype, including marker expression, angiogenic potential, and lineage-specific gene expression.

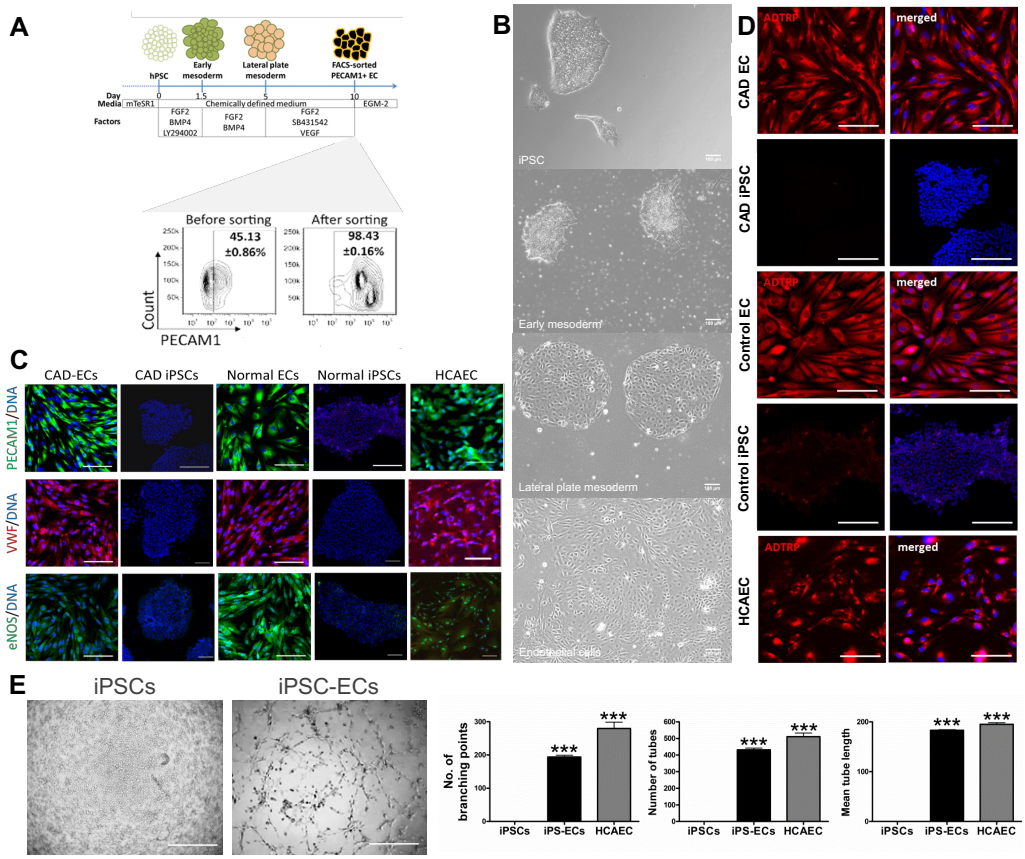


Figure 11. Differentiation donor derived iPSCs into endothelial cells and characterization of endothelial phenotype. (A) Schematic representation of the differentiation protocol used to generate endothelial cells from iPSCs. iPSCs were sequentially differentiated into early mesoderm and lateral plate mesoderm stages using chemically defined media supplemented with FGF2, BMP4, LY294002, and VEGF. Flow cytometry analysis before and after FACS sorting demonstrates PECAM1+ (CD31+) cell populations, showing 45.13 ± 20.86% PECAM1 positivity before sorting and enrichment to 98.43 ± 0.16% PECAM1+ cells after sorting. **(B)** Representative phase-contrast

images showing morphological changes during differentiation from iPSCs to endothelial cells. Progression from iPSCs through early mesoderm, lateral plate mesoderm, and mature endothelial cells is shown (scale bar, 100 µm). **(C)** Immunostaining of differentiated endothelial cells and control cells for endothelial markers PECAM1 (CD31), von Willebrand factor (VWF), and endothelial nitric oxide synthase (eNOS) CAD-ECs and control ECs demonstrate robust expression of endothelial markers, whereas iPSCs and HCAECs (human coronary artery endothelial cells) serve as negative and positive controls, respectively. **(D)** ADTRP expression analysis by immunostaining in CAD-derived ECs, CAD iPSCs, control ECs, and control iPSCs (scale bar, 100 µm). **(E)** Left: Representative images of tube formation after 2 hours by iPSCs and iPSC-derived endothelial cells (scale bar, 200 µm). Right: Image quantification of tube formation attributes of endothelial cord-like structures. Negative control was iPSCs, while positive control was HCAECs. Bar graphs showing means ± s.e.m. (n = 3 biological replicates), ***p ≤0.001, one-way ANOVA with post-hoc Tukey's tests.

Staining and tube formation assay done by Nicole Min Qian Pek and Dr Dalakrishnan Chakrapani Narmada.

3.3.3 CAD patient iPSC endothelial cells recapitulate disease phenotype

To assess the ability of CAD endothelial cells to model disease-relevant phenotypes, we evaluated oxidative stress, proliferative capacity, and proinflammatory markers. These features are key hallmarks of endothelial dysfunction and are directly implicated in the pathogenesis of CAD.

Oxidative stress

ROS production appeared elevated in CAD endothelial cells compared to controls, as visualized by MitoSOX Red staining (Figure 12A). While no statistical analysis was performed, the increased staining intensity may suggest elevated oxidative stress, a known contributor to endothelial dysfunction through mechanisms such as impaired nitric oxide signalling and proinflammatory activation.

Cell proliferation

Proliferative capacity was visualized using Ki-67 staining. CAD endothelial cells showed reduced Ki-67 staining relative to control cells, which may indicate reduced proliferation, a phenotype consistent with cellular senescence and dysfunction often observed in CAD (Figure 12B).

Proinflammatory status

CAD endothelial cells exhibited elevated IL-8 secretion, as quantified by ELISA (Figure 12C). Surface expression of ICAM-1 was also observed

to be higher in CAD endothelial cells in a representative flow cytometry plot (Figure 12D), although no statistical analysis was performed. ICAM-1 and IL-8 are key mediators of vascular inflammation and leukocyte recruitment, and their increased expression is consistent with the proinflammatory state characteristic of atherosclerosis.

These results collectively demonstrate that iPSC-derived endothelial cells from CAD patients recapitulate key hallmarks of endothelial dysfunction and inflammation observed in cardiovascular disease, although some findings are qualitative and should be interpreted with care.

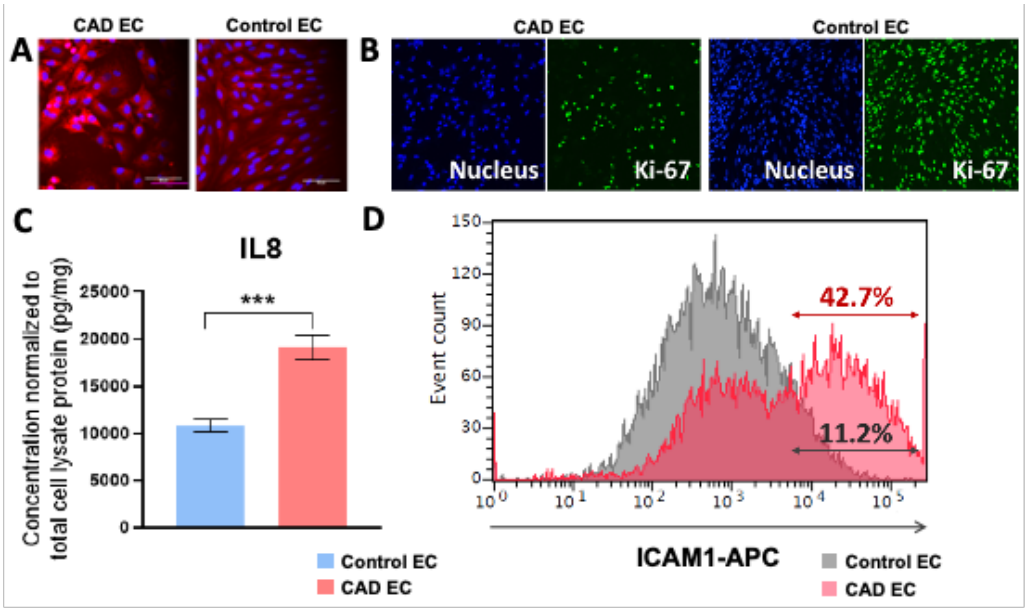


Figure 12. Assessment of proinflammatory status, proliferative capacity, and ICAM-1 expression in CAD and control iPSC-ECs. (A) Assessment of mitochondrial oxidative stress by Mitosox Red assay in CAD and control iPSC-ECs (scale bar, 100 µm). **(B)** Characterisation of cell proliferation by Ki-67 staining in CAD and control iPSC-ECs. **(C)** ELISA-based measurement of secreted IL-8 concentrations in culture media of CAD and control iPSC-ECs, which were conditioned for 72 h. Bar graphs showing means with S.D. (n = 4 from 2 donors with 2 iPSC lines/donor), ***$p \leq 0.001$, two-tailed t-test. **(D)** Flow cytometry characterization of proinflammatory marker ICAM-1 on CAD and control iPSC-ECs.

Mitosox Red assay, Ki-67 staining and IL-8 ELISA done by Nicole Min Qian Pek and Dr Dalakrishnan Chakrapani Narmada. Flow cytometry characterization done by Dr Kanxing Wu.

3.3.4 Generation of pMIA3-double guide RNA plasmid for CRISPR/Cas9 mediated gene editing

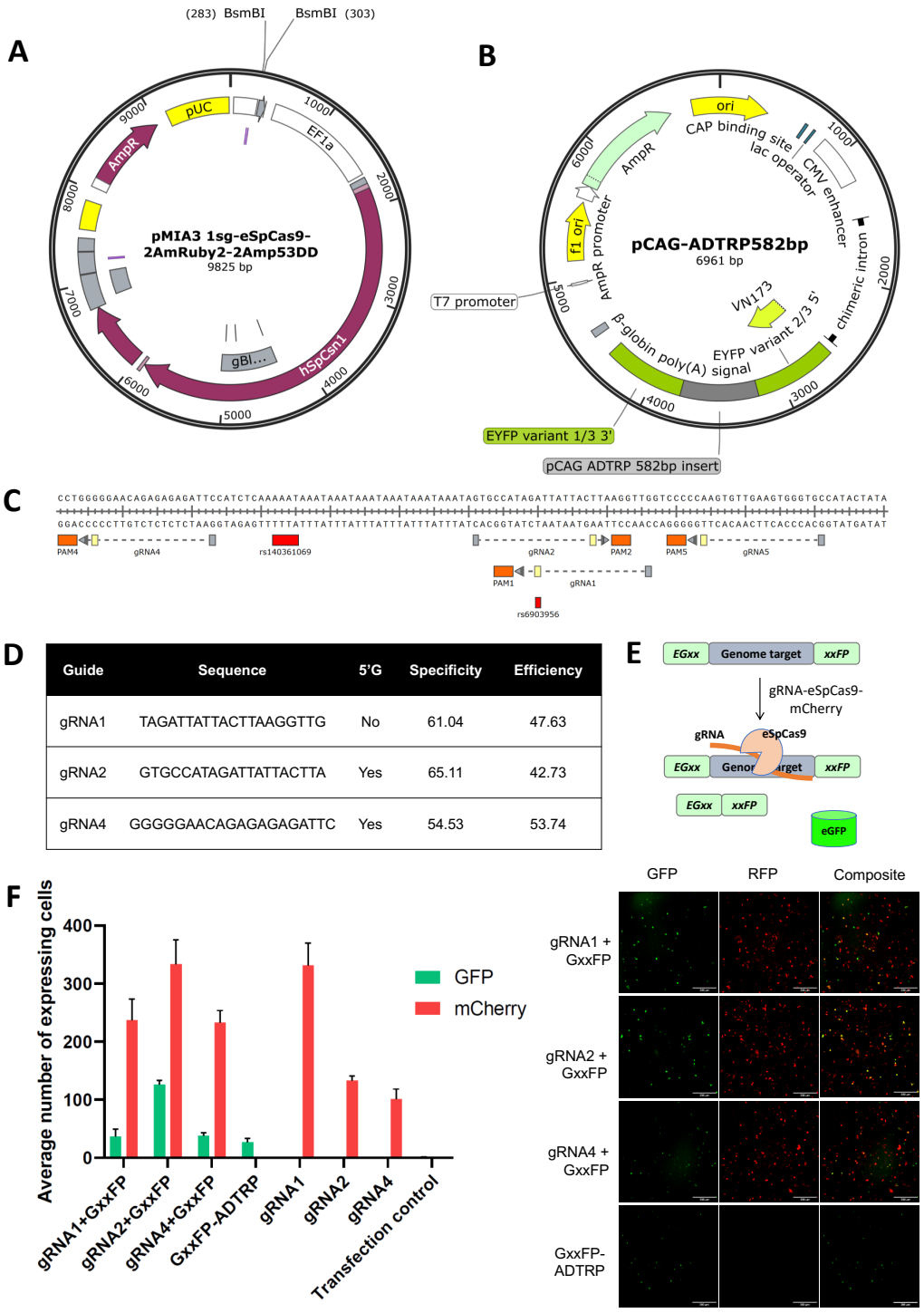
In Chapter 2, fine mapping identified rs140361069, a short tandem repeat variant, in weak LD with rs6903956 (R^2 = 0.21) in East Asian populations. The analysis demonstrated a correlation between the ‘A’

risk allele of rs6903956 and the deletion of (AAAT)3 at rs140361069, a LD pattern absent in European populations. This population-specific LD pattern suggests that rs140361069 may modulate the effects of rs6903956, potentially contributing to the observed differences in CAD susceptibility across ethnicities. To explore the functional relevance of these loci in the context of CAD, we utilized CRISPR/Cas9-mediated gene editing to precisely delete both rs6903956 and rs140361069 in CAD and control iPSCs.

We engineered a pMIA3 plasmid expressing gRNAs designed to target rs6903956 and rs140361069, alongside eSpCas9 and the fluorescent marker mRuby2 for the identification of transfected cells (Figure 13A). PAM sequences flanking the target loci were systematically evaluated to design gRNAs that optimize precise cleavage (Figure 13C, D). To validate the efficiency of these gRNAs, we employed a GFP-reporter assay in HEK293T cells. A 582 bp genomic fragment encompassing rs6903956 and rs140361069 was cloned into the pCAG-EGxxFP plasmid, a construct engineered to disrupt GFP fluorescence unless the target region is successfully cleaved by the gRNA-Cas9 complex (Figure 13B). Restoration of GFP functionality upon cleavage enabled quantification of gRNA efficiency through fluorescence intensity (Figure 13E).

Among the gRNAs tested, gRNA2 demonstrated the highest cleavage efficiency, indicated by the strongest GFP signal and the largest population of GFP-positive cells (Figure 13F). To account for variability in transfection rates, mRuby fluorescence was used as an internal control across experimental conditions, ensuring the reliability of our assay. These results provide evidence of the precision and efficiency of gRNA2 in directing Cas9-mediated cleavage.

Despite the promising results in HEK293T cells, initial attempts to edit iPSCs using single gRNAs were unsuccessful, even after screening over 50 clones. This prompted the adoption of a dual-gRNA strategy, leveraging the capacity of two gRNAs to induce larger deletions through NHEJ. This approach was implemented using the pMIA3 vector to enhance editing efficiency and ensure the precise deletion of both rs6903956 and rs140361069 in iPSCs.



Guide	Sequence	5'G	Specificity	Efficiency
gRNA1	TAGATTATTACTTAAGGTTG	No	61.04	47.63
gRNA2	GTGCCATAGATTATTACTTA	Yes	65.11	42.73
gRNA4	GGGGGAACAGAGAGAGATTC	Yes	54.53	53.74

Figure 13. Design of CRISPR-Cas9 pMIA3 construct for targeting ADTRP locus and evaluating editing efficiency. (A) Schematic map of pMIA3 (Addgene plasmid # 109399), a gift from Roger Foo. **(B)** Schematic map of pCAG-EGxxFP plasmid (Addgene plasmid # 50716), a gift from Masahito Ikawa, disrupted by 582bp ADTRP region. **(C)** Location of gRNA sequences flanking rs6903956 and rs140361069 (chr6:11,714,280-11,714,400). **(D)** Table summarizing individual gRNA sequences with specificity and efficiency scores ranked by algorithm developed by Doench et al.[184]. **(E)** Schematic of the dual-fluorescence reporter system for assessing gRNA activity. Successful targeting by the gRNA-eSpCas9-mCherry complex cleaves the GxxFP cassette, leading to GFP expression, while mCherry indicates transfection efficiency. **(F)** pMIA3 containing a

respective gRNA was co-transfected with pCAG-EGxxFP plasmid[185]. gRNAs 1, 2 and 4 were evaluated for targeting efficiencies by quantifying the average number of expressing cells. pMIA3 plasmids without co-transfection of pCAG-EGxxFP was used as control. Inactive pMIA3 plasmid was used as a negative control. Bar graph showing means with S.D. (n = 3 independent transfections).

3.3.5 Generation of isogenic iPSC lines with 63-89bp deletions on 6p24.1

Using the dual-gRNA pMIA3 plasmid, we successfully targeted rs6903956 and rs140361069 in both CAD and control iPSCs. Nucleofection of the plasmid achieved a transfection efficiency of 1.3% (Figure 14A). Single clones were isolated after two weeks of culture and screened for deletions by PCR amplification of a 582 bp region flanking the target sites. Sanger sequencing confirmed successful deletions ranging from 63 to 89 bp, encompassing both rs6903956 and rs140361069 (Figure 14B).

Editing efficiencies varied depending on the gRNA pairs used (Figure 14C). In control iPSCs, guide pair 1 and 4 yielded a 6% success rate, while guide pair 4 and 5 achieved a higher efficiency of 19%. For CAD iPSCs, guide pair 4 and 5 demonstrated the highest editing efficiency at 46%, while guide pair 4 and 2 resulted in a 2% success rate. Importantly, at least two independently derived edited cell lines were generated for both CAD and control groups using different guide pairs, ensuring the robustness and reproducibility of our experimental model.

To ensure the validity of our model, unedited clones subjected to nucleofection but lacking target deletions, were retained as controls. Immunostaining confirmed that the deletion of rs6903956 and rs140361069 did not compromise pluripotency or growth characteristics, as the edited iPSCs continued to express TRA-1-60, OCT4, and SOX2 at levels comparable to unedited lines (Figure 14D).

The successfully edited (Δ63-89bp) lines were designated as Δ CAD EC and Δ Control EC, while unedited clones were labelled as UNΔ CAD EC and UNΔ Control EC. Parental wild-type lines were designated WT CAD EC and WT Control EC. These isogenic lines serve as a robust platform to dissect the functional roles of rs6903956 and rs140361069 in endothelial biology and CAD pathogenesis.

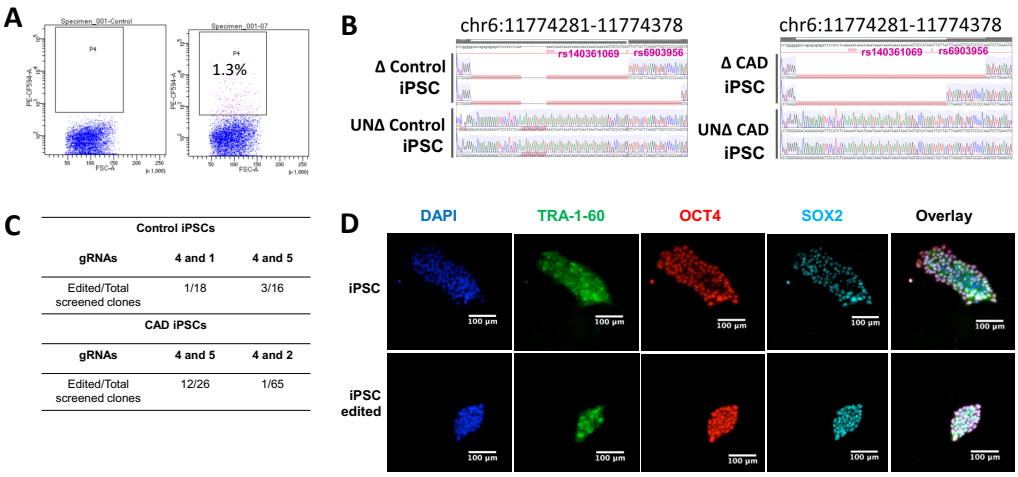


Figure 14. Generation of Δ63-89bp deletions flanking rs6903956 and rs140361069 in CAD and control iPSCs**. (A)** Flow cytometry of successfully nucleofected iPSCs with mCherry-pMIA3. Negative control (left) of iPSC without nucleofection. **(B)** Sequencing chromatograms to validate CRISPR-Cas9 deletion in edited (Δ) *versus* unedited (UNΔ) endothelial cells. Two different sets of gRNAs were rendered on each iPSC-endothelial cell line to control for CRISPR-Cas9 off-targeting effects. **(C)** Table summarizing editing efficiencies for different gRNA pairs in control and CAD-derived iPSCs. **(D)** Immunostaining for pluripotency markers on iPSCs and genetically edited iPSCs (scale bar, 100 µm).

3.4 Discussion

In Chapter 3, we established a CRISPR-edited iPSC-derived endothelial cell model to investigate the role of the 6p24.1 locus, encompassing rs6903956, in endothelial dysfunction and CAD. By using isogenic iPSC lines with and without targeted deletions, we differentiated these cells into endothelial cells and assessed their phenotypic and functional characteristics. The endothelial cells were characterized by marker expression, including PECAM1 (CD31), VE-cadherin, and ICAM-1, and their functionality was confirmed through tube formation assays and IL-8 secretion analysis. Building on Chapter 2, which identified endothelial cells as a relevant model for studying rs6903956, this work provides a powerful and physiologically relevant endothelial *in vitro* model for exploring rs6903956 CAD-associated mechanisms.

Sexual dimorphisms are well-documented in pathways relevant to endothelial dysfunction[186,187]. In the GWAS that identified rs6903956 as a CAD susceptibility locus[18], the OR for the A risk allele was higher in the female population (OR = 1.63) compared to the male population (OR = 1.44). Although this difference was not statistically significant (p = 0.32, Breslow-Day test), it underscores the potential for sex-dependent effects of rs6903956. However, our study only utilized gender-matched samples from male participants. Incorporating both sexes could provide additional insights into the role of rs6903956 in CAD pathogenesis and help identify potential sex-specific mechanisms.

Given that we did not create iPSC lines from the same clinical phenotype harboring both risk and non-risk alleles, the observed phenotypic effects cannot be attributed solely to rs6903956 genotypes. Instead, the results of deleting the 6p24.1 region with ‘AA’ risk alleles in isogenic cell lines (Δ CAD EC vs. UNΔ CAD EC) must be interpreted within the context of potential gene-environment interactions inherent to the CAD background[188,189]. Comparisons between these isogenic lines underscore the importance of considering gene-environment interactions in CAD models. A more accurate CAD comparison could involve the more recent base editing technology[100], which allows precise single nucleotide changes to edit the A risk allele in a CAD model to the G non-risk allele. This approach would also preserve the surrounding genomic elements, unlike the large chunk deletions (63–89 bp) used in this chapter to target the 6p24.1 locus, which ensured the removal of rs6903956 and rs140361069 but lacked specificity and risked disrupting neighbouring regulatory elements like TFBS that may influence the observed phenotypes.

In the differentiation of iPSCs into endothelial cells, we achieved a PECAM1 positivity of 45.13%, which is notably lower compared to newer protocols. For instance, a recent study by Ang et al.[190] published in 2022 reported arterial endothelial cell (aEC) purities as high as 97.2%. While our protocol effectively generated general endothelial cells, aECs are more relevant for CAD studies due to their distinct gene expression profiles and susceptibility to flow-associated endothelial dysfunction, necessitating a more specialized differentiation approach[191-195]. These cells, distinguished by artery-specific markers such as SOX17, CXCR4, and DLL4, can be generated by promoting arterial fate through TGF-β activation while simultaneously inhibiting PI3K to suppress vein formation. Adopting these strategies in future differentiation protocols could enhance the generation of aECs, thereby providing a more accurate and disease-relevant model.

The CRISPR editing strategy initially employed single gRNAs, which resulted in low editing efficiencies. To address this, we adopted a double-guide RNA strategy that improved editing success by inducing larger deletions through NHEJ. However, off-target effects remain a concern. To mitigate this, future experiments using this cell line should include at least two guide pairs per target locus to ensure reproducibility, as different guide pairs may have distinct off-target profiles. Additionally, RNA sequencing comparisons should be expanded to include wild type iPSCs alongside edited and unedited lines. This approach would provide a more comprehensive understanding of the global impact of CRISPR edits on gene expression and cellular function. It would also help identify unintended consequences of the edits and ensure that novel pathways uncovered are influenced by the 6p24.1 locus rather than by off-target effects.

Functional assays performed in this study included tube formation assays, which confirmed the angiogenic potential of the iPSC-derived endothelial cells. These cells formed robust cord-like structures on Matrigel, similar to HCAECs, validating their functionality. Additionally, IL-8 secretion was assessed as a marker of proinflammatory activity, and CAD-derived endothelial cells exhibited elevated IL-8 levels compared to controls, consistent with a proinflammatory state characteristic of CAD. To further refine functional assessments, flow-based assays mimicking arterial shear stress could be employed. These assays would offer a more physiologically relevant evaluation of endothelial cell function under conditions that closely replicate those in coronary vasculature[193,196].

In conclusion, while this study successfully established and characterized a CRISPR-edited iPSC-derived endothelial cell model, there are more opportunities to refine genetic manipulation strategies, improve differentiation protocols, and expand functional assays. These recent advancements can provide more detailed insights into the molecular pathways affected by rs6903956 and its role in endothelial dysfunction and CAD progression.

Chapter 4: Inter-chromosomal interaction of CAD risk loci 6p24.1 and 10q11.21 is associated with endothelial damage in coronary artery disease

Chapter 4: Inter-chromosomal interaction of CAD risk loci 6p24.1 and 10q11.21 is associated with endothelial damage in coronary artery disease

4.1 Introduction

In Chapters 2 and 3, we established the endothelial-specific regulatory role of rs6903956 in CAD and developed CRISPR-edited iPSC-derived endothelial cell models to investigate its functional impact.

While these efforts provided insights into the variant's contribution to endothelial dysfunction, the broader regulatory mechanisms mediated by rs6903956 remain unexplored. In this chapter, we expand upon this foundation by examining the cis- and trans-regulatory effects of rs6903956, aiming to uncover the gene networks and chromatin interactions that drive CAD pathogenesis.

4.1.1 Cis and trans eQTLs

eQTL analysis plays a critical role in linking SNPs identified through GWAS to variations in gene expression[197]. An eQTL consists of a variant and a gene, where the genotype at the SNP locus is significantly associated with the expression levels of the target gene. To identify eQTLs, biological samples are genotyped to map genetic variations, and then measure gene expression in these samples using methods like RNA sequencing. Statistical models are then employed to associate gene expression data with genetic variants across the genome.

Transcriptional regulatory factors are categorized into cis and trans-regulatory factors. Cis-regulatory elements include promoters, enhancers, and silencers. Conversely, trans-regulatory factors, such as transcription factors, noncoding RNAs, and signaling molecules, influence the expression of distally located genes. The intricate interplay between cis-regulatory sequences and trans-acting factors typically governs gene expression.

The interactions between cis and trans eQTLs can be complex[198]. For example, a SNP within a gene's promoter region that encodes a transcription factor might alter its expression (a cis-eQTL effect), subsequently affecting the expression of other genes targeted by this transcription factor across the genome (a trans-eQTL effect). This scenario illustrates how changes in cis-regulatory elements can lead to broad trans-regulatory impacts.

Mechanistic studies from GWAS-derived eQTLs have largely focused on cis-eQTLs, where genetic variants influence the expression of genes in

their proximity. Notable examples include the lead SNP rs9349379 on 6p24.1 within *PHACTR1*, affecting the expression of endothelin-1 *(EDN1)* located 600kb upstream[7], and cardiovascular disease-associated SNPs in the 9p21 locus impacting *ANRIL* levels[130]. Additionally, the SNP rs7539120 within an enhancer of *NOS1AP* on chromosome 1q has cardiac-specific eQTL effects on *NOS1AP* expression[69].

Although most eQTL studies primarily concentrate on cis-eQTLs, these account for only a limited fraction of the heritability associated with complex traits and diseases. On the other hand, trans-eQTLs, which are significantly more difficult to detect, could potentially explain a more considerable portion of trait variance[199,200]. The challenge in identifying trans-eQTLs stems from the need to evaluate a vast number of SNP-gene pairings for trans associations[201]. This process demands larger sample sizes and more substantial effects to achieve statistical significance.

The study of trans-eQTLs is crucial as they have the potential to reveal complex gene regulatory networks underlying diseases by showing how distant genetic variants influence gene expression across different chromosomes[202]. The GTEx project[86] has identified both cis- and trans-eQTLs in multiple human tissues, although data on endothelial cell types are not available in GTEx.

To investigate the potential cis and trans effects of rs6903956, we performed RNA sequencing to identify genes and pathways impacted by the deletion of rs6903956 in iPSC-derived endothelial cells. Transcriptomic analysis revealed a network of genes associated with cardiovascular disease pathways driven by the presence of the AA risk genotype. Interestingly, deletion of the AA risk genotype in CAD endothelial cells minimally impacted the expression of nearby cis genes, such as *ADTRP*. Instead, we observed that *CXCL12*, a distal gene located at 10q11.21, displayed reduced expression in edited CAD endothelial cells, reaching levels comparable to those in normal (non-risk GG) endothelial cells. This finding suggests that rs6903956 mediates trans-regulation of *CXCL12*, a well-established player in vascular biology and CAD pathogenesis.

4.1.2 Chromatin conformation capture

Chromatin conformation capture techniques provide insights into 3D organization of the genome, which is essential for regulating gene expression[203]. These methods identify interactions between distant chromatin regions, enabling regulatory elements to modulate

transcription by interacting with gene promoters. Such interactions are cell-type specific, reflecting the unique epigenetic controls that define cellular functions and identity[204].

Super-enhancers, characterized by their dense clusters of enhancers and high concentration of transcriptional activators[205], are known to engage in more chromatin interactions compared to typical enhancers. These interactions often span across multiple genes essential for cell identity[206]. Super-enhancers, along with TADs, help maintain the integrity of gene regulation in complex diseases[207]. TADs segregate genomic segments to prevent inappropriate enhancer-promoter interactions and misexpression of genes. Boundaries of TADs are marked by the presence of structural proteins, with CTCF facilitating the formation of these boundaries through orientation-dependent binding, and cohesin aiding in loop formation through its extrusion mechanism[208].

The Chromosome Conformation Capture (3C) technique and its derivatives, such as 4C, 5C, ChIA-PET, HiChIP, and Hi-C, are crucial for detecting DNA sequences that are physically proximate within the cell nucleus, often providing insights into transcriptional regulation[209]. Hi-C, in particular, was developed by Lieberman-Aiden et al.[84] as a method to map genome-wide chromatin interactions by deep sequencing. This approach marked a significant advancement over previous methods that required predefined regions of interest. To enhance the usability of Hi-C data, the same team introduced an automated pipeline called Juicer[210]. This tool processes Hi-C data from raw sequence information to structured outputs. It transforms raw data into Hi-C contacts, pairs of genomic positions adjacent to each other in 3D space, and compiles these into contact matrices. Juicer then annotates these matrices to identify structural genomic features such as loops, loop anchor motifs, and contact domains. However, despite these advancements, the resolution of Hi-C is still limited by restriction enzymes, which fragment DNA into segments approximately 4kb in length. These segments can be too large to capture finer details of regulatory interactions, thus posing a challenge to fully resolving complex genomic structures and their functions.

In this chapter, Hi-C analysis of our iPSC lateral mesoderm derived endothelial cells and HUVECs from Rao and Huntley et al.[211] revealed a significant chromatin interaction between approximately 2.2kb region at 6p24.1, where the SNP rs6903956 is located, and a 15kb region upstream of *CXCL12* at 10q11.21. This interaction, potentially mediated by a super-enhancer located on chromosome 10, was stronger in cells carrying the AA risk genotype compared to those with the GG non-risk

genotype. Notably, 10q11.21 is a known CAD risk locus, suggesting that the spatial arrangement of these regions within endothelial cells could be crucial for the coordinated regulation of gene expression.

Our findings propose that CAD risk loci at 6p24.1 and 10q11.21 are organized within endothelial chromatin to co-regulate gene expression, potentially playing a role in CAD pathogenesis[212]. By integrating chromatin conformation and transcriptomic data, this chapter aims to elucidate the combined effects of cis and trans-regulatory mechanisms on endothelial biology and CAD susceptibility, providing a clearer view of the complex regulation by non-coding variants.

4.1.3 CXCL12

GWAS have linked the 10q11 genomic locus, containing the *CXCL12* gene, to an increased risk of CAD[213]. Variants associated with CAD in this region correlate with plasma levels of CXCL12, although the precise mechanisms underlying this relationship remain unclear[214]. CXCL12, also known as SDF-1α, is secreted by stromal cells and plays a critical role in angiogenesis[215], inflammation[216], wound healing[217], and embryonic development[218].

CXCL12 interacts with its receptors, CXCR4 and ACKR3, and binds to glycosaminoglycans (GAGs) on the endothelium, enhancing leukocyte presentation[219]. This interaction promotes angiogenesis independent of VEGF, facilitating migration and tubulogenesis in HUVECs[220].

Elevated serum levels of CXCL12 have been observed in patients with cardiovascular diseases[214,221,222], where endothelial-derived CXCL12 plays a significant role in atherosclerosis progression and contributes to circulating CXCL12 levels[223]. CXCL12 has previously been associated with a salutary effect on atherosclerotic plaque stability[224,225]. In contrast, other studies have shown that inflammation mediated by CXCL12 and its receptors have long been linked to vascular injury, involving processes of monocyte differentiation and macrophage infiltration, aggravating pro-inflammatory responses on vascular endothelium[226]. CXCL12 is also involved in other pro-atherogenic processes, including hyperlipidemia and insulin resistance that may in turn affect endothelial integrity[227].

In Apoe$^{-/-}$ murine models, augmented expression of *CXCL12* has been observed to accelerate the progression of atherosclerosis[228]. Furthermore, Hostler et al.[229] showed that conditional knockout of CXCL12 in endothelial cells was shown to significantly affect ischemic tissue viability, alter reparative processes, and modify tumorigenesis,

while not influencing embryogenesis or morphogenesis. CXCL12 deficiency in endothelial cells disrupts essential interactions with fibroblasts, crucial for stromal development and vascularization. These observations illuminate the complex role of CXCL12 in vascular biology and underscore its potential implications in the pathogenesis of CAD.

4.2 Methodology

4.2.1 RNA Sequencing and analysis

Total RNA from iPSC-derived endothelial cells was extracted using the RNeasy Micro Kit (Qiagen, catalog no. 74004). PolyA library preparation and 150 bp paired-end sequencing were performed on an Illumina HiSeq platform (Novogene, Singapore) with an average depth of 60 million reads. Reads were aligned to the human genome (hg38) using STAR aligner v2.4.1a, and gene-level read counts were extracted. Data were normalized using the edgeR package (v3.34.0)[230] with TMM, and differentially expressed genes were analyzed with the limma package (v3.46.0)[231]. Log2-counts-per-million (logCPM) values were calculated using the voom transformation to model the mean-variance relationship. Heatmaps were created with the pheatmap R package (v1.16.0). Principal component analysis (PCA) plots were generated with scatterplot3d R package v0.3-41. Gene ontology annotations[232] from Molecular Signatures Dataset (MSigDB)[233] was used to perform Fast Gene Set Enrichment Analysis (fgsea) R package v1.18.0. The functional disease networks were generated through the use of IPA (QIAGEN Inc.)[234].

4.2.2 RNA extraction and quantitative RT-PCR

Total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, catalog no. 74134) following the manufacturer's instructions and converted to cDNA using the LunaScript RT SuperMix Kit (NEB, catalog no. E3010S). Real-time PCR was conducted with SYBR Green assays (NEB, catalog no. M3003S) on a QuantStudio 6 system (Applied Biosystems). Gene expression was normalized to the housekeeping gene GAPDH.

Primer sequences used for qRT-PCR in chapter 4 are as follows:
ADTRP F: 5’-GCCCATTCACGGAATGTCA-3’
ADTRP R: 5’-CGTCGTCATCACCCTCAACT -3’
EDN1 F: 5’-GCCCATTCACGGAATGTCA-3’
EDN1 R: 5’-CGTCGTCATCACCCTCAACT -3’
PHACTR1 F: 5’-GCCCATTCACGGAATGTCA-3’
PHACTR1 R: 5’-CGTCGTCATCACCCTCAACT -3’

4.2.3 Hi-C

Hi-C was conducted using the Arima-HiC kit (Arima Genomics, San Diego) following the manufacturer's protocol up to DNA purification. Library preparation and sequencing were completed by Integrated

Genome Analytics Platform. Hi-C reads were aligned to the human reference genome (hg38) using BWA-MEM (v0.7.15), with forward and reverse reads independently mapped to detect ligation events between genomic loci. The Juicer pipeline was used for read processing, filtering out unmapped, duplicate, and low-quality reads (MAPQ < 30), as well as self-ligated fragments (<2 kb apart)[210]. Filtered read pairs were used to generate chromatin contact maps.

For identification of toplogically associating domains (TADs), the resulting .hic matrix files (MAPQ > 30) were used as input for Arrowhead to identify TADs. Arrowhead ran automatically as part of Juicer pipeline post-processing. The following default parameters were set for Arrowhead: Resolution (r) = 5kb; Normalization (k) = Knight-Ruiz balancing (KR); Size of sliding window (m) = 2000.

For visualization of Hi-C output, Hi-C contact matrix are visualized using 3D genome browser and Juicebox software. Circular visualization of HUVEC (CC-2517) trans chromosomal interactions were visualized using Rondo[235].

4.2.4 3C-droplet digital PCR

3C libraries were prepared as previously described[236]. Briefly, 1 x 10^7 cells were crosslinked with 2% formaldehyde, quenched with 0.125 mM glycine, and lysed with cold lysis buffer (10 mM Tris-HCl, pH 7.5; 10 mM NaCl; 5 mM MgCl2; 0.1 mM EGTA; 1x protease inhibitor, Roche, catalog no. 11836145001). Nuclei were digested overnight at 37°C with 400 U FastDigest HindIII (Thermo Fisher Scientific, catalog no. FD0504) and ligated at 16°C for 4 h in 1.15x T4 ligation buffer (New England Biolabs, catalog no. B0202S) with 100 U T4 ligase. Samples were de-crosslinked with Proteinase K (Thermo Fisher Scientific, catalog no. EO0492), treated with RNase (Thermo Fisher Scientific, catalog no. EN0531), and purified using phenol-chloroform extraction. DNA concentrations were determined via qPCR against reference genomic DNA using internal primers amplifying GAPDH.

ddPCR was performed on the QX200 Droplet Digital PCR System (Bio-Rad Laboratories) following the manufacturer's protocol and a 3C-digital PCR workflow. Reactions (24 µL total) included 12 µL 2x ddPCR Supermix for Probes (No dUTP; Bio-Rad, catalog no. 1863024), 900 nM primers, 250 nM probes, and 400 ng 3C template DNA. Negative controls (ddH2O) and control primers upstream of the constant fragment were used to ensure accuracy. Cycling conditions were 95°C for 10 min, followed by 45 cycles at 94°C for 30 s and 60°C for 2 min, and final denaturation at 98°C for 10 min. Signals were quantified using the

QX200 Droplet Reader and analyzed with Quantasoft Pro v1.7. Interaction frequency was calculated as target copy number per 100 ng of DNA: 24 × copies/µL divided by 100 × DNA input.

Primer sequences used for 3C-ddPCR in chapter 4 are as follows:
3C *GAPDH* internal primer F: 5'-TCTGACTTCAACAGCGACAC-3'
3C *GAPDH* internal primer R: 5'-GCTGTAGCCAAATTCGTTGTC -3'
3C constant fragment probe: 5'-AGGTGGGGCCACATAACTGGGATC TCCAGG-3'
3C constant fragment primer: 5'-AGGGGCCCACTCAGAATCTC-3'
3C T4: 5'-CACCTGCTCCACCTTCACTT-3'
3C T5: 5'-TTGCCAGCTGGTTGACTGAAA-3'
3C T6: 5'-TGTGACAAGGTGACACGAGG-3'

4.2.5 ENCODE and ChIP-Seq analysis

Analysis of H3K27ac and H3K4me3 histone marks was conducted using coordinates chr10:44,297,249-44,312,248 on hg38 for visualization of weak promoter lying ~2kb from fragment 4, and chr10:41,830,816-41,866,615 in hg38 for visualization of super-enhancer region. Individual bigWig (BW) files were visualized on IGV. ChIP-Seq human vascular endothelial cells was analyzed from GEO dataset (GSE131953).

4.2.6 Circulating endothelial cells

A total of 100 µL of 1 million PBMCs were stained with antibodies in the dark for 10 minutes at room temperature, followed by 20 minutes at 4°C on an analog tube rotator. After staining, cells were washed and resuspended in 200 µL of PBS with 1% BSA for analysis. CECs were identified based on the immunophenotypic profile CD45-/CD31+/CD133-/DNA+, and their frequency was expressed as cells per million PBMCs. Antibodies used were Hoechst 33342 (Life Technologies, catalog no. R37165), CD45-FITC (Thermo Fisher Scientific, catalog no. 11-9459-42), CD133-APC (BioLegend, catalog no. 372806), CD31-PECy7 (BioLegend, catalog no. 303118). Flow cytometry was performed using a BD LSRFortessa X-20 with FACSDiva software (BD Biosciences), and data were analyzed with FlowJo v10.7.1 (Becton Dickinson). Each analysis included a minimum of 30,000 events.

For CXCL12 quantification in CECs, total RNA was isolated from sorted CECs and probe-based ddPCR was performed following the method described above. Absolute quantity of DNA per sample (copies/µL) was processed using QuantaSoft (v.1.0.596) and converted to copies/CEC according to amount of input sample.

4.2.7 Statistical analysis

All statistical analyses were performed using GraphPad Prism version 9. Data normality was evaluated using the Shapiro-Wilk test. For datasets with a normal distribution, unpaired two-tailed Student's t-tests were used to compare two groups, and one-way ANOVA followed by Tukey’s post-hoc tests was applied for comparisons involving more than two groups. For non-normally distributed data, appropriate non-parametric tests, such as the Mann–Whitney U test for two-group comparisons or the Kruskal-Wallis test for multiple groups, were utilized. Statistical significance was defined as $p < 0.05$. Results are presented as mean ± SD. Specific statistical tests applied to individual experiments are detailed in the corresponding figure legends.

4.3 Results

4.3.1 Cis-environment of 6p24.1 (Hi-C)

The regulation of non-coding SNPs such as rs6903956 is closely tied to chromatin interactions, as chromatin architecture dictates the spatial and functional relationships between enhancers, promoters, and their target genes. SNPs within regulatory regions can modulate chromatin accessibility, looping dynamics, or the recruitment of transcriptional machinery, impacting the expression of both nearby (cis) and distal (trans) genes. To investigate whether rs6903956 may regulate cis-genes through such mechanisms, we generated Hi-C genomic contact maps of CAD and control endothelial cells at a 10-kb resolution (Figure 15A, B).

Using *in situ* Hi-C, we identified an average of 101 million long-range (≥ 20 kb) intra-chromosomal contacts per sample, after aligning and filtering 301 million Hi-C read pairs. Chromatin interactions are typically confined within TADs—spatial compartments demarcated by boundaries that restrict enhancer-promoter interactions to defined genomic regions. TADs are known to be highly conserved across species and cell types, providing a stable structural framework for gene regulation[237]. Comparing Hi-C data from CAD EC, control EC, and a well-characterized HUVEC line[211] revealed highly similar TAD characteristics across all three datasets, indicating that the chromatin organization of our iPSC-derived endothelial models faithfully recapitulates a canonical endothelial system (Figure 15C).

Our Hi-C contact maps showed that rs6903956 lies within a TAD that includes neighbouring genes *ADTRP*, *EDN1*, and *PHACTR1* (Figure 15D). This TAD structure raises the possibility that rs6903956 could exert cis-regulatory effects by modulating enhancer-promoter interactions within the domain. To test this hypothesis, we analysed the expression of genes within the TAD in edited (Δ63–89bp) and unedited endothelial cell lines using qRT-PCR. Despite previous studies reporting decreased *ADTRP* expression in leukocytes carrying the A risk allele at rs6903956, our results showed no significant changes in *ADTRP* or *EDN1* expression in Δ vs. UNΔ ECs.

Interestingly, *PHACTR1* expression was upregulated in Δ Control EC compared to all CAD EC lines, but there were negligible differences in PHACTR1 expression between edited and unedited lines within CAD or control groups. These results suggest that the Δ63–89bp region encompassing rs6903956 does not substantially alter the expression of cis-genes within this TAD. Our data suggest that the regulatory influence

of the Δ63–89bp region flanking rs6903956 in iPSC-derived endothelial cells may operate through mechanisms that extend beyond cis-regulatory interactions.

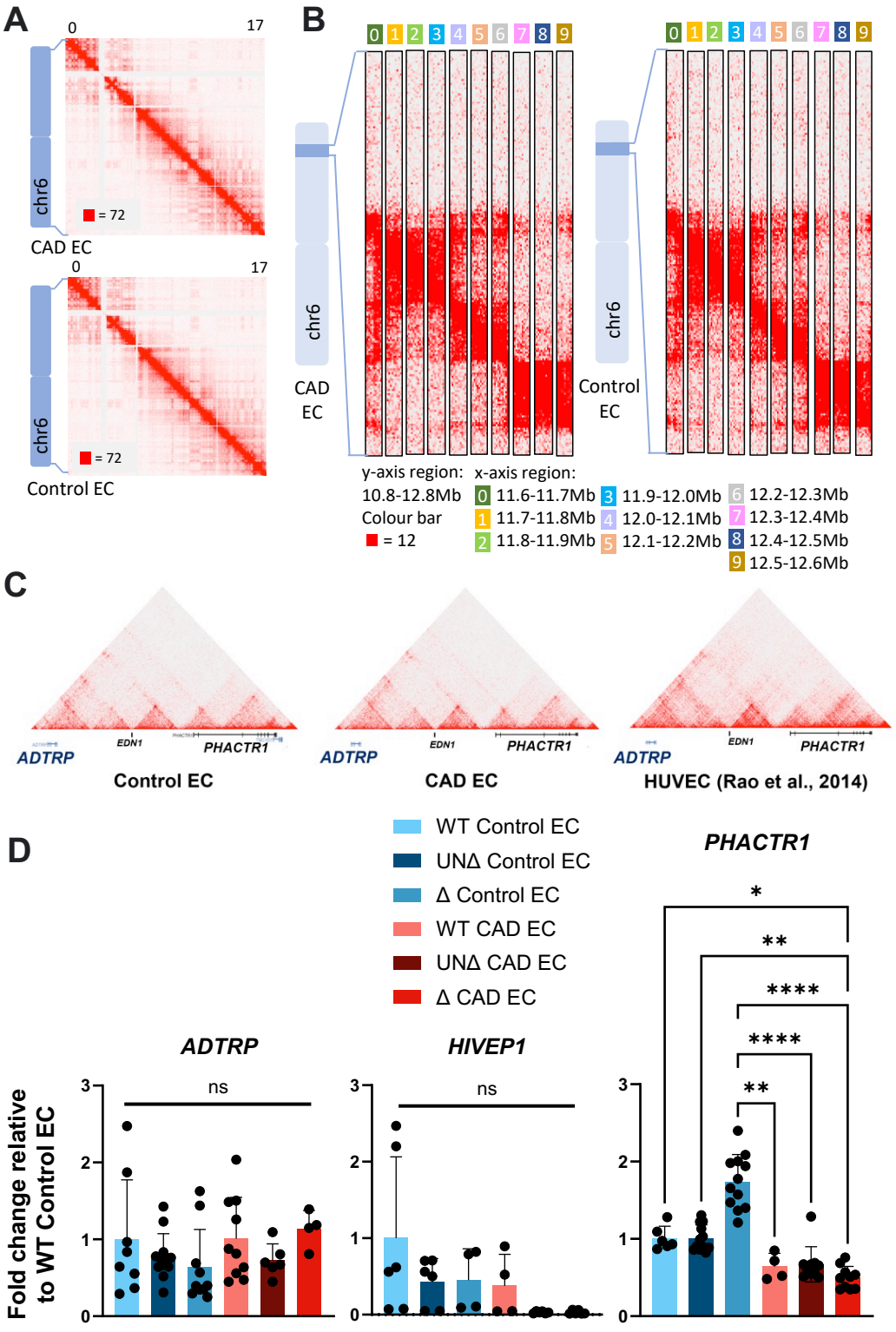


Figure 15. Cis environment of 6p24.1 in CAD and control endothelial cells. (A) Hi-C contact maps for chromosome 6 in CAD and control endothelial cells. The contact

intensity is represented by the heatmap, with darker red indicating stronger interactions. **(B)** Zoomed-in view of the 11.6–12.6 Mb region on chromosome 6. The x-axis regions are subdivided into 100 kb bins (color-coded from 0 to 9). **(C)** Hi-C contact map visualization of 6p24.1 locus in Control EC, CAD EC and HUVEC to reflect TADs involving rs6903956 in *ADTRP* and neighbouring cis genes. **(D)** Quantitative RT-PCR of *ADTRP* and cis genes in WT, isogenic Δ and UNΔ CAD and Control EC. Bar graphs showing means with S.D. (n = 2–6, from 2 donor cell lines with 2 biological replicates per cell line generated by 2 different guide RNA pairs), ***$p \leq 0.001$, Kruskal-Wallis test.

4.3.2 Genetically edited endothelial cells reveal downregulation of molecular pathways relating to endothelial instability

To uncover potential trans effects mediated by the rs6903956 locus, RNA sequencing was conducted on iPSC-derived endothelial cells with risk genotype (AA) and non-risk genotype (GG). This included WT CAD EC and WT Control EC (n = 3 biological replicates from 3 independent differentiation batches), as well as their Δ and UNΔ counterparts (n = 6 from 2 independently generated cell lines using distinct gRNA pairs, 3 independent differentiation batches per cell line).

Principal component analysis (PCA) demonstrated minimal variance between biological replicates, highlighting the robustness of the RNA sequencing datasets (Figure 16A). Additionally, minimal differences between cell lines generated using different gRNA pairs suggested minimal off-target effects of the CRISPR editing strategy.

Differential expression analysis revealed 183 upregulated and 270 downregulated genes in Δ CAD EC compared to UNΔ CAD EC (Figure 16B, left). In contrast, 1682 upregulated and 2001 downregulated genes were identified in Δ Control EC compared to UNΔ Control EC (Figure 16B, right). To validate that observed gene expression changes were not merely artifacts of CRISPR editing, we compared Δ CAD EC and Δ Control EC against their respective WT datasets. Differentially expressed genes identified between WT CAD EC and Δ CAD EC, and between WT Control EC and Δ Control EC, were largely consistent with the comparisons of Δ EC to their UNΔ counterparts (Figure 16C).

Gene set enrichment analysis (GSEA)[233,238] was performed to identify biological processes impacted by the differentially expressed genes. Comparing Δ CAD EC to UNΔ CAD EC, we observed significant downregulation of pathways related to vascular development, abnormal vascular physiology, endothelial cell migration, and sprouting angiogenesis (Figure 16D). These findings suggest that the deletion of the Δ63–89bp region at rs6903956 leads to a less activated/proliferative endothelial phenotype, potentially reversing pathogenic endothelial activation associated with CAD.

In contrast, Δ Control EC compared to UNΔ Control EC displayed substantially more differentially expressed genes; however, GSEA did not reveal any pathways directly linked to endothelial function or vascular biology (Figure 16E). Instead, the deleted region on 6p24.1 containing the GG non-risk genotype in Δ Control EC appeared to broadly impact fundamental biological processes unrelated to endothelial activation.

These findings suggest that the G-to-A allele substitution at rs6903956 may represent a gain-of-function mutation. Deletion of the AA risk genotype in Δ CAD EC, but not deletion of the GG non-risk genotype in Δ Control EC, was associated with significant downregulation of endothelial activation pathways. This supports the hypothesis that the presence of the A allele may confer a pathogenic regulatory role, leading to an activated endothelial phenotype associated with CAD.

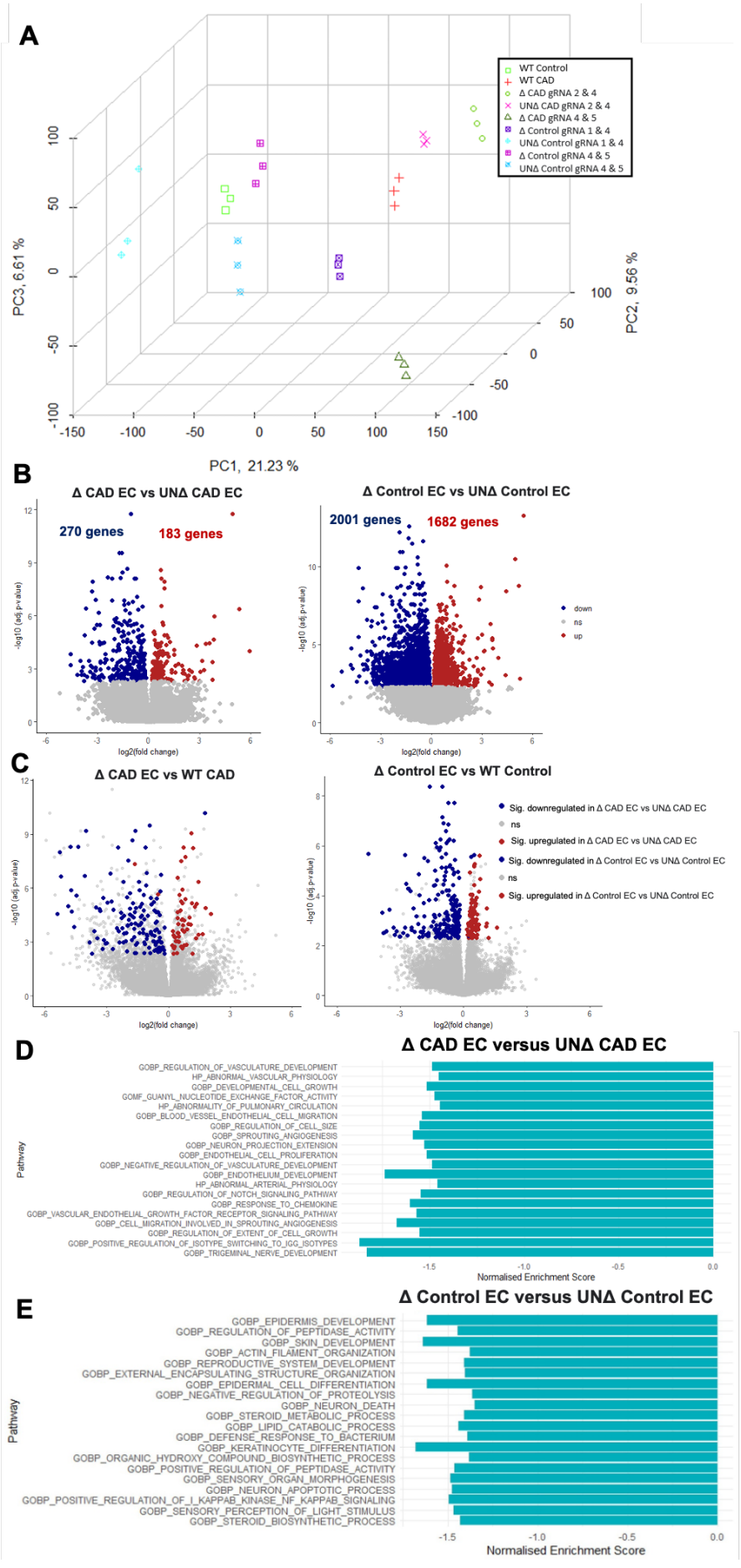

A
WT Control
WT CAD
Δ CAD gRNA 2 & 4
UNΔ CAD gRNA 2 & 4
Δ CAD gRNA 4 & 5
Δ Control gRNA 1 & 4
UNΔ Control gRNA 1 & 4
Δ Control gRNA 4 & 5
UNΔ Control gRNA 4 & 5
PC3, 6.61 %
PC2, 9.56 %
PC1, 21.23 %
B
Δ CAD EC vs UNΔ CAD EC
270 genes
183 genes
Δ Control EC vs UNΔ Control EC
2001 genes
1682 genes
-log10 (adj p-value)
log2(fold change)
down
ns
up
C
Δ CAD EC vs WT CAD
Δ Control EC vs WT Control
Sig. downregulated in Δ CAD EC vs UNΔ CAD EC
ns
Sig. upregulated in Δ CAD EC vs UNΔ CAD EC
Sig. downregulated in Δ Control EC vs UNΔ Control EC
ns
Sig. upregulated in Δ Control EC vs UNΔ Control EC
D
Δ CAD EC versus UNΔ CAD EC
GOBP_REGULATION_OF_VASCULATURE_DEVELOPMENT
HP_ABNORMAL_VASCULAR_PHYSIOLOGY
GOBP_DEVELOPMENTAL_CELL_GROWTH
GOMF_GUANYL_NUCLEOTIDE_EXCHANGE_FACTOR_ACTIVITY
HP_ABNORMALITY_OF_PULMONARY_CIRCULATION
GOBP_BLOOD_VESSEL_ENDOTHELIAL_CELL_MIGRATION
GOBP_REGULATION_OF_CELL_SIZE
GOBP_SPROUTING_ANGIOGENESIS
GOBP_NEURON_PROJECTION_EXTENSION
GOBP_ENDOTHELIAL_CELL_PROLIFERATION
GOBP_NEGATIVE_REGULATION_OF_VASCULATURE_DEVELOPMENT
GOBP_ENDOTHELIUM_DEVELOPMENT
HP_ABNORMAL_ARTERIAL_PHYSIOLOGY
GOBP_REGULATION_OF_NOTCH_SIGNALING_PATHWAY
GOBP_RESPONSE_TO_CHEMOKINE
GOBP_VASCULAR_ENDOTHELIAL_GROWTH_FACTOR_RECEPTOR_SIGNALING_PATHWAY
GOBP_CELL_MIGRATION_INVOLVED_IN_SPROUTING_ANGIOGENESIS
GOBP_REGULATION_OF_EXTENT_OF_CELL_GROWTH
GOBP_POSITIVE_REGULATION_OF_ISOTYPE_SWITCHING_TO_IGG_ISOTYPES
GOBP_TRIGEMINAL_NERVE_DEVELOPMENT
Pathway
Normalised Enrichment Score
E
Δ Control EC versus UNΔ Control EC
GOBP_EPIDERMIS_DEVELOPMENT
GOBP_REGULATION_OF_PEPTIDASE_ACTIVITY
GOBP_SKIN_DEVELOPMENT
GOBP_ACTIN_FILAMENT_ORGANIZATION
GOBP_REPRODUCTIVE_SYSTEM_DEVELOPMENT
GOBP_EXTERNAL_ENCAPSULATING_STRUCTURE_ORGANIZATION
GOBP_EPIDERMAL_CELL_DIFFERENTIATION
GOBP_NEGATIVE_REGULATION_OF_PROTEOLYSIS
GOBP_NEURON_DEATH
GOBP_STEROID_METABOLIC_PROCESS
GOBP_LIPID_CATABOLIC_PROCESS
GOBP_DEFENSE_RESPONSE_TO_BACTERIUM
GOBP_KERATINOCYTE_DIFFERENTIATION
GOBP_ORGANIC_HYDROXY_COMPOUND_BIOSYNTHETIC_PROCESS
GOBP_POSITIVE_REGULATION_OF_PEPTIDASE_ACTIVITY
GOBP_SENSORY_ORGAN_MORPHOGENESIS
GOBP_NEURON_APOPTOTIC_PROCESS
GOBP_POSITIVE_REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_SIGNALING
GOBP_SENSORY_PERCEPTION_OF_LIGHT_STIMULUS
GOBP_STEROID_BIOSYNTHETIC_PROCESS
Pathway
Normalised Enrichment Score

Figure 16. Transcriptomic analysis of WT, isogenic Δ63–89bp CAD and control endothelial cells. (A) 3D-PCA plot of wild-type iPSC-derived endothelial cells (n = 3 biological replicates from 3 independent differentiation batches), as well as edited and unedited endothelial cells (n = 6 from 2 cell lines generated by 2 different guide RNA pairs, 3 independent differentiation batches/ cell line). **(B)** Volcano plots showing the adjusted p-values and log2 fold change values of genes in Δ CAD EC *versus* UNΔ CAD EC (left), as well as in Δ Control EC *versus* UNΔ Control EC (right). Differentially expressed genes are indicated by red (upregulated) and blue (downregulated) dots. **(C)** Volcano plots showing the adjusted p-values and log2 fold change values of genes in Δ CAD EC versus WT CAD EC (left), as well as in Δ Control EC versus WT Control EC (right). The blue (downregulated in Δ versus UNΔ) and red (upregulated in Δ versus UNΔ) dots on the plot represent the significant genes which are similarly differentially expressed in their respective Δ versus UNΔ counterparts. **(D)** Top 20 ontology gene sets normalized enrichment scores from Gene Set Enrichment Analysis, ranked according to *p*-value. Signed fold change of genes differentially expressed in Δ CAD EC *versus* UNΔ CAD EC was used as input. **(E)** Signed fold change of genes differentially expressed in Δ Control EC versus UNΔ Control EC were used as input.

4.3.3 Transcriptomic analysis reveals CXCL12 as a potential downstream trans target

To identify candidate effector genes potentially regulated by the deleted region on 6p24.1, we employed a double-pronged approach combining network analysis and expression dynamics.

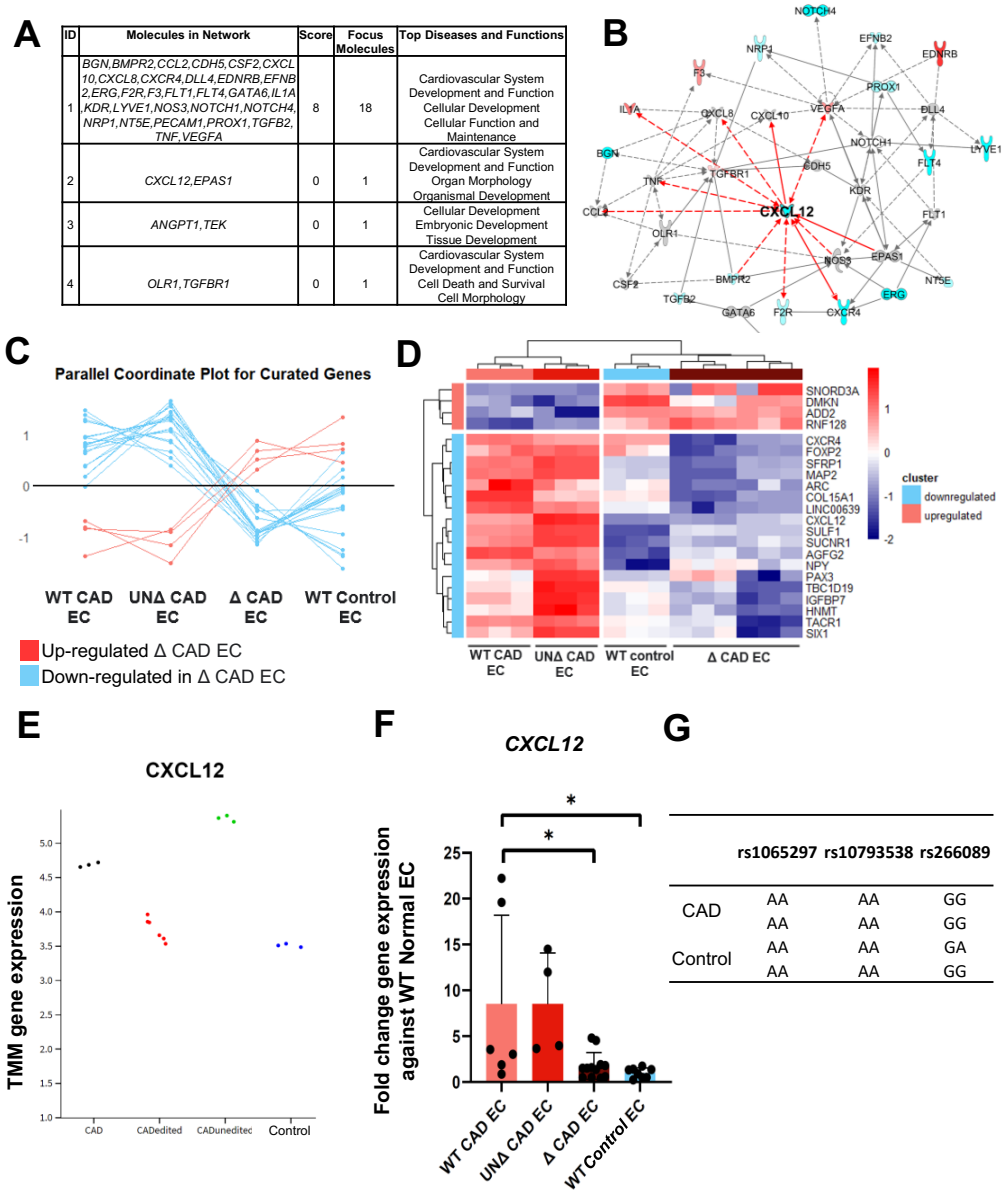
First, we conducted a 'diseases and functions' network analysis using Ingenuity Pathway Analysis (IPA) on differentially expressed genes between Δ CAD EC and UNΔ CAD EC. This analysis revealed that the network associated with "cardiovascular system development and function" was among the most dysregulated pathways in Δ CAD EC (Figure 17A). Further exploration of the network interactome indicated that most genes within this network were downregulated, with *CXCL12* emerging as a central node in this interactome (Figure 17B).

To complement this, we visualized gene expression dynamics using parallel coordinate plots to identify candidate genes whose expression levels were restored in Δ CAD EC to resemble those of WT normal EC harbouring the non-risk GG genotype (Figure 17C). Heatmap visualization helped refine the candidate list, highlighting genes normalized to WT normal EC levels upon deletion of the AA risk genotype in Δ CAD EC (Figure 17D). Consistently, *CXCL12* surfaced as a strong candidate effector gene (Figure 17E).

qRT-PCR analysis further validated the expression pattern of *CXCL12*, showing significant downregulation in genetically edited Δ CAD EC compared to its parental WT CAD EC (Figure 17F). Importantly, this reduction in *CXCL12* expression upon deletion of the AA risk genotype suggests a trans-regulatory relationship mediated by the 6p24.1 locus.

Coincidentally, *CXCL12* resides on chromosome 10q11.21, a well-established CAD susceptibility locus[141,239]. Three *CXCL12* variants (rs266089, rs1065297, rs10793538) have previously been associated with CAD in the Chinese Han population[240]. To evaluate whether these variants contributed to the observed *CXCL12* expression differences, we genotyped our iPSC lines. All lines harboured common alleles at these *CXCL12* SNPs, except for one normal iPSC line with a heterozygous 'GA' genotype at rs266089, which has been postulated to confer CAD risk (Figure 17G). However, no significant difference in *CXCL12* expression was observed between our WT normal EC lines, suggesting that rs266089 has negligible impact on *CXCL12* expression in our iPSC-derived endothelial cell models.

These findings motivated us to explore the potential for trans-chromosomal interactions linking the 6p24.1 locus, where rs6903956 resides, to the 10q11.21 locus harbouring *CXCL12*.



ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	*BGN,BMPR2,CCL2,CDH5,CSF2,CXCL10,CXCL8,CXCR4,DLL4,EDNRB,EFNB2,ERG,F2R,F3,FLT1,FLT4,GATA6,IL1A,KDR,LYVE1,NOS3,NOTCH1,NOTCH4,NRP1,NT5E,PECAM1,PROX1,TGFB2,TNF,VEGFA*	8	18	Cardiovascular System Development and Function Cellular Development Cellular Function and Maintenance
2	*CXCL12,EPAS1*	0	1	Cardiovascular System Development and Function Organ Morphology Organismal Development
3	*ANGPT1,TEK*	0	1	Cellular Development Embryonic Development Tissue Development
4	*OLR1,TGFBR1*	0	1	Cardiovascular System Development and Function Cell Death and Survival Cell Morphology

Figure 17. CXCL12 *trans*-regulated gene identified from list of curated genes. (A) 'Diseases and functions' network analysis by Ingenuity Pathway Analysis for differentially expressed genes in Δ CAD EC versus UNΔ CAD EC. **(B)** An interactome of differentially expressed genes in Δ CAD EC *versus* UNΔ CAD EC, arising from the 'cardiovascular system development and function' network based on Ingenuity Pathway Analysis. Genes indicated in red and blue were upregulated and downregulated in Δ CAD EC respectively. **(C)** Parallel coordinate plot for curated genes that showed gene expressions in Δ CAD EC 'normalized' to the levels of that in WT Control EC. **(D)** Heatmap visualization of curated genes from (C). **(E)** CXCL12 TMM genes expression from RNA Sequencing across WT CAD, Δ CAD, UNΔ CAD and Control EC. **(F)** Quantitative RT-PCR of *CXCL12*, in WT, isogenic Δ and UNΔ CAD and Control EC. Bar graphs showing means with S.D. (n = 2–6, from 2 donor cell lines with 2 biological replicates per cell line generated by 2 different guide RNA pairs), ***$p \leq 0.001$, Kruskal-Wallis test. **(G)** CAD and control donor genotype of CXCL12 variants identified by GWAS[141,239].

4.3.4 Contact between 6p24.1 and 10q11.21 by three-dimensional chromatin structure

To investigate whether *rs6903956* at 6p24.1 influences *CXCL12* expression through inter chromosomal interactions, we analysed the chromatin architecture between 6p24 and 110q11.21 and its potential regulatory association with *CXCL12*.

Hi-C data from Rao et al.[211] revealed that the 6p24.1 and 10q11.21 loci are closely intertwined in 3D genomic space in HUVEC. Notably, the 10q11.21 locus harbours a ~30kb super enhancer region (Figure 18A). Super enhancers are critical for orchestrating gene expression programs that define cell identity[241].

To validate the endothelial specificity of this super enhancer, we analysed ChIP-seq data from various human vascular endothelial cell types[159]. Consistently, increased H3K27Ac enhancer marks were observed at the 10q11.21 super enhancer in all vascular endothelial cell types examined, including HCCaEC, HCoAEC, HaoEC, HPAEC, and HENDC, but not in the non-endothelial IMR90 lung fibroblast cell line (Figure 18C).

To directly test whether the AA risk genotype at *rs6903956* facilitates interactions with regulatory elements activating *CXCL12*, we performed chromatin conformation capture (3C) on WT CAD EC (risk AA) and WT normal EC (non-risk GG). Using a constant HindIII fragment encompassing *rs6903956* as a fixed interaction hotspot, we examined its ligation frequency with 27 HindIII fragments spanning the *CXCL12* genomic region.

Given the challenges inherent in 3C, including low ligation frequencies for inter-chromosomal interactions and biases toward cis interactions[242,243], we employed droplet digital PCR (ddPCR) for precise quantification. ddPCR bypasses the need for normalization controls required by conventional TaqMan qPCR[244] and delivers absolute quantification of ligation products. Pooled samples from five independently differentiated iPSC-derived endothelial batches ensured sufficient cell numbers (10 million) and minimized batch-to-batch variability.

Our findings revealed frequent chromatin interactions between the control fragment flanking *rs6903956* and the *CXCL12* genomic region (Figure 18B). Specifically, fragment 4, located within the 5’ UTR of *CXCL12*, demonstrated significantly stronger interaction signals with the

control fragment flanking AA risk genotype compared to the GG non-risk genotype.

To explore whether regions around fragment 4 could be functionally regulating *CXCL12* expression, we reanalysed ChIP-seq data from vascular endothelial cells[159]. A weak but consistent H3K4me3 histone peak, potentially indicative of promoter activity, was identified in fragment 5, located ~2kb downstream of fragment 4 (Figure 18D). These findings suggest that the CAD background coupled with AA risk genotype at rs6903956 seemed to confer greater interaction frequency near a weak promoter on *CXCL12*.

This highlights a possible trans-regulatory mechanism wherein the AA risk genotype at *rs6903956* promotes physical interactions with a putative weak *CXCL12* promoter at 10q11.21. The involvement of a super enhancer at 10q11.21 underscores the functional importance of inter-chromosomal interactions in regulating endothelial gene expression and CAD pathogenesis.

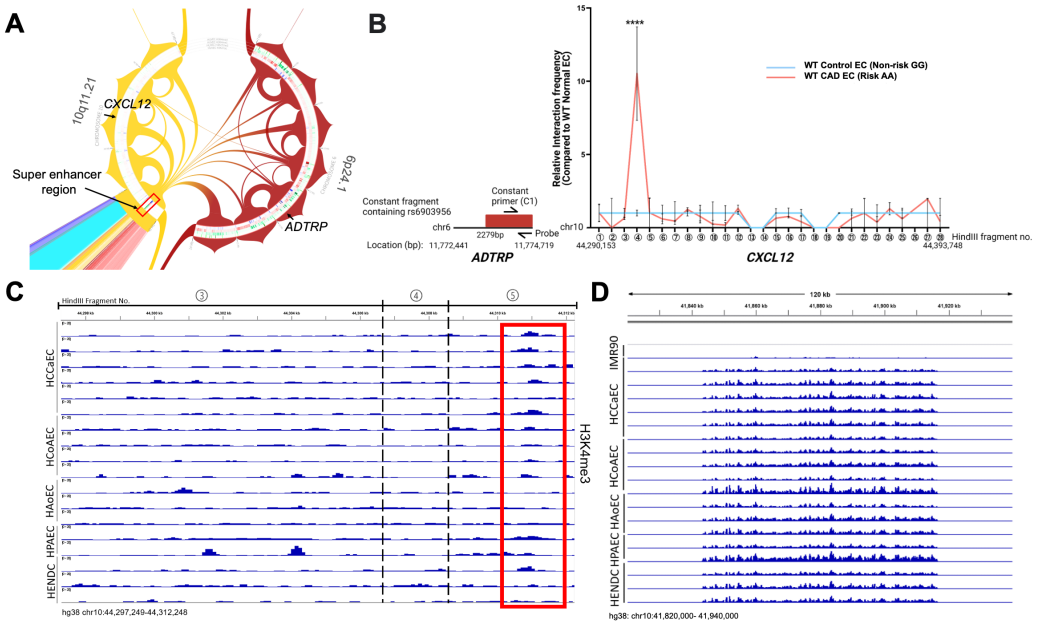


Figure 18. Frequent chromatin contacts between 6p24.1 and 10q11.21. (A) Chord diagram visualizing HUVEC dataset from Rao et al.[211]. Hi-C inter-chromosomal loops between 6p24.1 and 10q11.21 on hg19. Only significant interactions (q-value <0.01) are shown. Presence of a super enhancer region on hg38: chr10:41,830,816- 41,866,615 (boxed in red) where most of the chromatin contacts on 6p24.1 and 10q11.2 are anchored on, including *CXCL12* and *ADTRP*. **(B)** Chromatin conformation capture (3C)-digital PCR of WT Control EC and WT CAD EC. Anchoring on HindIII fragment harboring rs6903956 (constant fragment; 2279bp resolution), we probed for HindIII fragments for the entire CXCL12 gene locus. Data represent fold changes against WT Control EC. Line graph showing means with S.E.M. (n = 4 from 2 donor cell lines with 2 technical replicates/ cell line), ****p ≤0.0001, two-way ANOVA comparing WT CAD EC with WT

Control EC for each individual fragment. **(C)** ChIP-Seq of H3K4me3 histone marks in vascular endothelial cells[159] on hg38 chr10:44,297,249–44,312,248: Common Carotid Artery ECs (HCoAEC, *n* = 4), Coronary Artery ECs (HCCaEC, *n* = 6), Human Aortic ECs (HAoEC, *n* = 3), Human Pulmonary Artery ECs (HPAEC, *n* = 2), Human Endocardiac Cells (HENDC, *n* = 3). Presence of a weak promoter ~2 kb downstream of fragment 4 was boxed in red. **(D)** ChIP-Seq of H3K27Ac histone marks in vascular endothelial cells (3). Presence of an endothelial specific super enhancer region on hg38: chr10:41,830,816- 41,866,615. Normal lung fibroblast cell line (IMR90) was used as a non-vascular control.

4.3.5 Rs6903956 risk alleles intensified vascular injury in patients with CAD

To test the effect of risk allele A at rs6903956, we measured CECs as a biomarker for vascular injury[245], in the blood of CAD patients stratified by genotypes. CECs are damaged endothelial cells found in the peripheral blood, which have been detached from the blood vessel lining because of vascular injury[126,245,246].

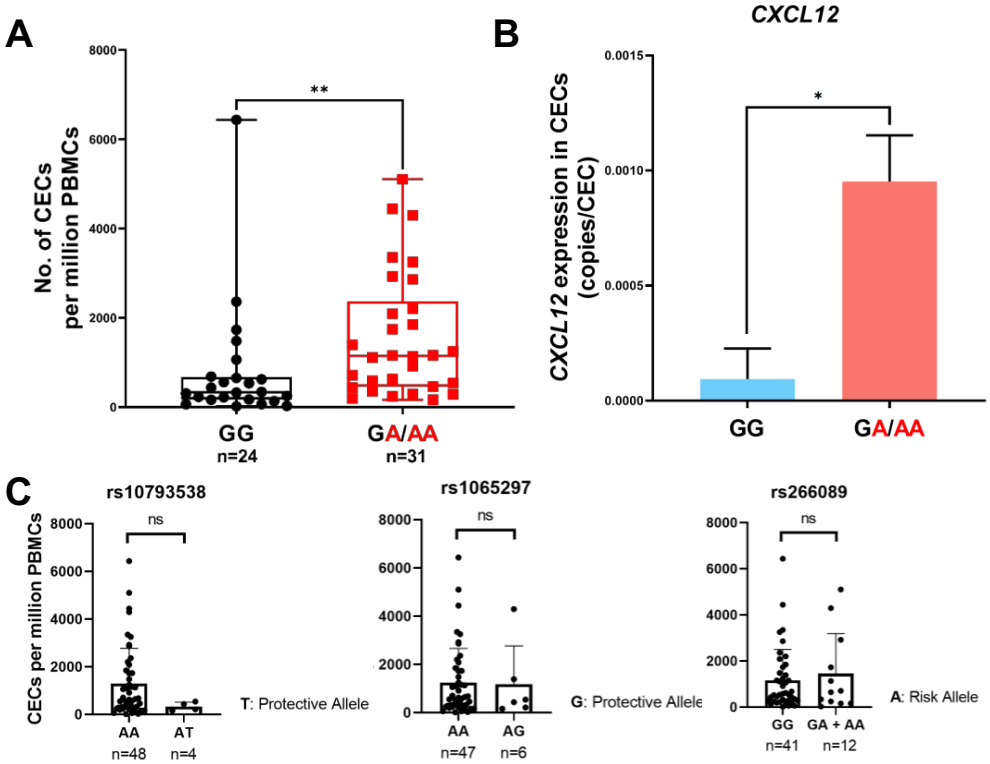
With all CAD patients age- and gender-matched, we achieved appropriate sample sizes for the non-risk GG (n = 24) and risk AA/AG (n = 31) genotypes, under the assumption of a dominant model (Table 8). While GG (n = 24) and AG (n = 23) genotype groups met sample size requirements, only a small fraction of the cohort carried the homozygous AA genotype (0.7% of CAD patients, n = 7). As a result, the dominant model was adopted due to the low MAF of rs6903956.

Remarkably, patients with risk genotypes AA/AG displayed significantly higher numbers of CECs compared with those carrying the non-risk GG genotype (Figure 19A). These findings associate the A risk allele of rs6903956 with heightened vascular injury.

To assess whether rs6903956 influences endothelial expression of *CXCL12*, we isolated CECs by FACS and grouped them based on the presence or absence of risk alleles (GA/AA, n = 27; GG, n = 20). Given the low numbers of isolated CECs, we performed ddPCR for precise quantification of *CXCL12* transcripts. Patients with risk genotypes (AA/AG) exhibited approximately 10-fold higher *CXCL12* copy numbers per CEC compared with the non-risk GG genotype (Figure 19B). These results further support the association between the rs6903956 *A* allele and elevated *CXCL12* expression.

Coincidentally, three *CXCL12* variants (rs1065297, rs266089, rs10793538) have also been reported to associate with CAD susceptibility in the Chinese Han population. Of these, rs1065297 and rs10793538 are protective variants, while rs266089 increases CAD risk. To attribute the observed CEC counts solely to rs6903956, we

genotyped these *CXCL12* SNPs in our patient population (Figure 19C). No significant correlation was observed between CEC counts and any of these *CXCL12* variants, potentially due to their low minor allele frequencies, as reported in both our cohort and previous studies by Zhang et al.[240]. Our findings underscore that the heightened vascular injury observed in CAD patients is likely linked to the rs6903956 A risk allele.



SNP	Base/Change	MAF (%) (Case)	MAF (%) (Control)
rs1065297	G/A	4.6	7
rs266089	A/G	19	16
rs10793538	T/A	4.6	6.8

Figure 19. Circulating endothelial cell profiling reveals association of rs6903956 risk allele with vascular injury. (A) Association between number of CECs per million PBMCs in patient samples and their genotypes at rs6903956. Bar graphs showing means with S.D. (n = 31 'GA/AA' and n = 24 'GG'), $^{**}p \leq 0.01$, Mann–Whitney U test. **(B)** Droplet digital PCR of *CXCL12* transcript copies from isolated CECs pooled from 'GG' (n = 20) and 'GA/AA' (n = 27) patient samples. Bar graph showing means with S.D., $^{*}p \leq 0.05$, Mann–Whitney U test. **(C)** Association of CECs with rs1065297, rs266089 and rs10793538 genotypes. Bar graphs showing means with S.D. (n = 31 'GA/AA' and n = 24 'GG'), ns ≥ 0.05, Mann–Whitney U test.

Genotypes		Analysis of CECs Non-risk GG (*n* = 24)	Risk AA/ AG (*n* = 31)
Age		54 (52.75 – 56.25)	53 (52 – 57)
Male		24 (100)	31 (100)
Race			
1)	Chinese	24 (100)	26 (83.9)
2)	Malay-Indian	0 (0)	5 (16.1)

Table 8: Demographics of CAD patients from whom samples were used for analysis of patient CECs. Data were presented as median (interquartile range) for continuous variables or n (%) for categorical variables.

4.4 Discussion

Our genetic experimentation using patient-derived iPSC endothelial cells provides new insights into the molecular basis of susceptibility SNP rs6903956 and its role in CAD pathogenesis. In Chapter 4, we explored the potential mechanistic contributions of rs6903956, focusing on its trans-chromosomal regulatory impact on *CXCL12*, an atherosclerosis-implicated gene. Our findings suggest that this trans-regulatory interaction may represent one of the key pathways through which rs6903956 exerts its deleterious effects, as evidenced by elevated vascular injury levels in patients harbouring the A risk allele. Facilitated by CRISPR-Cas9 deletions (Δ63-89bp) on 6p24.1 including rs6903956 in our iPSC-derived endothelial cells, we demonstrated that AA risk genotype at rs6903956 dysregulates vascular physiology transcriptional networks. On the other hand, removal of the non-risk GG genotype at rs6903956 had minimal effect on endothelial-specific pathways.

Our transcriptomic and chromatin conformation analyses uncovered a key trans-regulatory mechanism where upregulated *CXCL12* expression was associated with increased inter-chromosomal interaction frequency between 6p24.1 (with the AA risk genotype) and a putative promoter region upstream of *CXCL12* on 10q11.21. These findings propose a model where long-range chromatin interactions drive CAD pathogenesis, implicating *CXCL12* as a downstream effector of rs6903956. Furthermore, our observations are supported by the detection of significantly higher levels of damaged CECs in patients harbouring the A risk allele at rs6903956, reinforcing its role in endothelial injury.

SNP rs6903956 lies within the first intron of *ADTRP* that has been found to enhance the anticoagulant properties of endothelial cells in the presence of androgen[145]. Risk allele A at rs6903956 has been associated with decreased *ADTRP* mRNA expression in leukocytes[18]. More recently, the locus containing rs6903956 was postulated to act as an enhancer of *ADTRP*. Luo et al. transfected a 519bp region spanning rs6903956 and a 1513bp promoter/regulatory region of *ADTRP* into HeLa cells and found higher *ADTRP* promoter activity with preferential binding of transcription factor GATA2 to rs6903956 with common allele G[144]. They further showed in HUVECs that *GATA2* siRNA knock down resulted in lower ADTRP protein levels, but a direct enhancer-promoter interaction remained to be proven. Here, we were surprised that deletions of 6p24.1 containing rs6903956 (Δ63-89bp) in both WT CAD EC and WT Control EC did not seem to affect *ADTRP* gene expression in iPSC-derived endothelial cells. As CAD pathogenesis also involves other cells such as smooth muscle cells and macrophages[4], cell-type

specific effects of rs6903956 on *ADTRP* expression warrant further studies.

Causal non-coding SNPs can interfere with normal gene regulation by being regulatory loss- or gain-of-function mutations. On certain occasions, non-coding variants also have the potential to interfere with chromatin topological domain architecture by repositioning regulatory elements between domains or disrupting TAD boundaries, inducing ectopic gene activation and causing misexpression[247]. TADs form an intermediate level of chromatin organization by constraining interactions of cis-regulatory sequences with their cognate genes. Our lab found previously that many cardiovascular disease-associated SNPs contribute to variation of gene expressions within the same TAD as the variants concerned[248]. Hi-C contact maps of our iPSC-derived endothelial cells revealed that *ADTRP* lies within the same TADs as neighboring genes *EDN1* and *PHACTR1*. However, there was minimal cis-regulatory effect of rs6903956, other than alteration of *EDN1* expression associated with non-risk homozygous G alleles. The minimal cis-regulatory effect of rs6903956 may be due to the large segment deletion affecting other nearby SNPs or the differentiation of iPSCs into a general lateral mesoderm derived endothelial cell type, which may not reveal specific SNP impacts.

Instead, we observed using iPSC-endothelial cells that the G-to-A allele substitution at rs6903956 may confer CAD risk through long-range chromatin interactions with *CXCL12* residing on 10q11.21. We demonstrated the feasibility of obtaining high-resolution trans interaction by coupling 3C with droplet digital PCR. Higher inter-chromosomal contact frequency was found between 6p24.1 containing AA risk genotype at rs6903956 and a 1,792bp region lying within 5' untranslated region of *CXCL12*, which appeared to be a weak promoter. We note that rs6903956 is in LD with rs140361069, a short tandem repeat which was missing in our CAD EC harboring homozygous A alleles at rs6903956. Sun et al. had shown that some disease-associated tandem repeats may be located with chromatin domain boundaries, affecting insulation of genomic neighborhoods[249].

It remains to be elucidated how G-to-A substitution at rs6903956 and/or missing tandem repeats at rs140361069 alter chromatin dynamics, leading to elevated interaction frequency with a distal promoter site on *CXCL12*. It is also unknown if rs6903956 region can activate expression of a putative promoter of *CXCL12* via a direct enhancer-promoter interaction. We have not proven the specific enhancer role of rs6903956 region, and no strong distinct H3K27Ac enhancer histone marker peak

has been detected at rs6903956 region in endothelial ChIP-Seq. It is necessary to substantiate the enhancer role of rs6903956 region with further experimental validation using reporter assays in an endothelial cell system.

Our Hi-C data mining suggested there was an endothelial-specific super-enhancer on 10q11.21 that might be responsible for the regular contacts between 6p24.1 and 10q11.21. Super-enhancers have been thought to facilitate cell-identity specific regulatory response via the formation of phase separation bodies[250]. We find it fascinating that 6p24.1 and 10q11.21, both known loci harboring genetic polymorphisms associated with CAD, are closely interacting in 3D space. It will be worth investigating the potential global regulation of super-enhancers and how CAD risk loci on different chromosomes can mutually regulate gene expressions in chromatin space, opening the door for discovery of coregulated gene clusters via super-enhancer-mediated inter-chromosomal interaction.

In this study, rs6903956 was associated with increased *CXCL12* expression in endothelial cells. To validate our findings, we measured CECs as a biomarker to examine the functional consequence of rs6903956 on vascular injury in CAD patients, as CECs have been shown to be an indicator of vascular injury as they are shed into the circulation following vascular damage[123,251]. CECs are valuable markers of vascular dysfunction in a variety of vascular disorders including myocardial infarction, acute ischemic stroke[252], atherosclerosis, vasculitis, CAD[253] etc. We detected a significantly higher number of damaged CECs in patients harboring risk alleles A at rs6903956 than those with homozygous non-risk alleles G, supporting the role of rs6903956 in endothelial dysfunction.

There are limitations to this chapter that should be acknowledged. Firstly, creating a precise single nucleotide switch from G to A at rs6903956 has proven challenging. As a result, our findings, based on CRISPR-edited iPSC-derived endothelial cells with small deletions (Δ63–89bp) at 6p24.1, may be confounded by uncharacterized regulatory elements within the deleted region. This limitation underscores the importance of further refinement in gene editing techniques to isolate the specific contribution of rs6903956.

Secondly, we have not yet validated whether the 6p24.1 locus, where rs6903956 resides, functions as an enhancer or interacts with transcription factors, which could influence the observed differences in long-range chromatin interaction dynamics. Such validation would

provide deeper insights into the role of rs6903956 in altering chromatin topology and its potential contribution to CAD.

Finally, while we demonstrated an association between the A risk allele at rs6903956 and increased *CXCL12* expression alongside greater endothelial damage, the precise mechanism remains elusive. Trans-regulatory effects are typically mediated by diffusible elements, such as transcription factors or RNA molecules, that influence enhancer-promoter interactions. However, the exact molecular players linking rs6903956 to *CXCL12* dysregulation remain to be identified. Future studies should aim to elucidate whether the interaction between 6p24.1 and 10q11.21 is facilitated by intermediate regulatory elements, or indirect causal genes, or both.

Overall, chapter 4 demonstrates the utility of patient-derived iPSC models for dissecting the functional consequences of GWAS loci, shedding light on how non-coding variants like rs6903956 contribute to CAD risk through trans-regulatory chromatin interactions. By uncovering the connection between rs6903956 and *CXCL12* dysregulation, our findings provide critical insights into the complex interplay of chromatin architecture and gene regulation in CAD pathogenesis. This work enhances our understanding of the trans-chromosomal regulatory landscape underlying GWAS loci, offering new perspectives on the molecular mechanisms driving cardiovascular disease.

Chapter 5: Deciphering the cis-regulatory role of rs6903956 using base edited isogenic arterial endothelial cell model

Chapter 5: Deciphering the *cis*-regulatory role of rs6903956 using single base edited isogenic arterial endothelial cell model

5.1 Chapter background

Building on the foundation of our earlier chapters, this final chapter addresses previous limitations in chapter 3 and 4 with a new disease model – single base edit of rs6903956 in 2 CAD iPSC lines and differentiation into arterial-specific endothelial cells, to unravel the specific regulatory transcription factor mechanisms of rs6903956.

5.1.1 Single base edited iPSC-arterial endothelial cell model

In Chapter 3, we described the generation of isogenic endothelial cells via CRISPR-Cas9-mediated NHEJ deletion of a 63–89bp segment encompassing rs6903956 and an adjacent tandem repeat, rs140361069[212]. These cells were derived from CAD patient derived iPSCs and differentiated into general endothelial cells of lateral plate mesoderm origin[178], provided a cellular model for functional studies.

Chapter 5 introduces an improved disease model to specifically isolate the effect of rs6903956. First, we employed a precise adenosine base editing system to create isogenic lines from CAD patient-derived iPSCs, converting the risk AA genotype to the non-risk GG genotype. This strategy avoids the potential confounding effects of the large 63–89bp deletion used previously and allows for a focused investigation of rs6903956 within the specific CAD disease background. Specific base edits are crucial for understanding regulatory mechanism of SNPs, as they allow for the precise dissection of effects such as transcription factor binding, which cannot be elucidated from large segment deletions. This can be supported in the study by Gao et al.[96], which investigated rs11672691, a regulatory SNP associated with aggressive prostate cancer (PCa). Gao et al. converted the A non-risk to G risk allele in the PCa cell line 22Rv1 using base editing and demonstrated altered expression of *PCAT19* and *CEACAM21*. They further identified that rs11672691 acts as an enhancer variant by introducing a TFBS for HOXA9, which regulates its eQTL genes. This study highlights the power of precise genetic editing in dissecting the functional roles of TFBSs and their downstream effects.

Furthermore, instead of using general endothelial cells, we differentiated the base-edited iPSCs into aECs using a recently published protocol by Ang and colleagues[190]. Arterial specification is particularly relevant in the context of CAD pathophysiology, as arteries and veins exhibit distinct molecular and functional profiles. aECs express unique markers such as

EphrinB2, which delineates the arterial phenotype, whereas venous endothelial cells express EphrinB4, a key determinant of venous identity[195]. Arterial fate is promoted by increased activation of and Notch signalling pathways, while venous specification requires reduced VEGF activity and increased expression of COUP-TFII (NR2F2)[254-256]. Additionally, arterial and venous endothelial cells differ in their responses to flow-induced shear stress. aECs become more polarized and exhibit more prominent actin stress fibers aligned in the direction of flow compared to venous cells[191,192]. Arteries are also subjected to higher pressure and are more susceptible to disturbed flow at vascular branch points, which are regions associated with endothelial dysfunction and atherosclerosis[193,194]. Given these distinctions, generating aECs provided a more physiologically and pathologically relevant cellular model for dissecting the cis-regulatory mechanisms of rs6903956 in CAD.

5.1.2 Transcription factors in regulatory elements

Many disease-associated variants are situated within key regulatory domains, including enhancers, which are dense with TFBSs[204]. TFBSs are crucial for coordinating the complex networks of gene expression that respond to environmental stimuli and internal cellular conditions[257]. Within these regulatory landscapes, the dynamic interactions among transcription factors, coactivators, and chromatin remodelling complexes play a vital role in determining tissue-specific transcriptional responses and cellular behaviour[258]. Variants that alter transcription factor binding within these regions can disrupt these finely tuned processes, leading to variations in phenotype and increased susceptibility to diseases.

Transcription factors interact with a spectrum of cis-regulatory elements from promoters at transcription start sites to distant enhancers, silencers, and insulators. These interactions are essential for the accurate placement and function of RNA polymerase II, which orchestrates the assembly of the transcription initiation complex[259]. The stability and efficacy of these regulatory complexes are often ensured by the evolutionary conservation of their binding sites.

Building on findings from Chapter 4, RNA sequencing of Δ63–89bp isogenic endothelial cells indicated significant upregulation of *CXCL12* in unedited CAD endothelial cells, hinting at a potential trans-regulatory effect mediated by rs6903956. Chromatin conformation studies further suggested interactions between the rs6903956 region on 6p24.1 and a weak promoter region of *CXCL12* on 10q11.21. While these findings supported the possibility of long-range chromatin interactions influencing *CXCL12* expression, the nature of most trans-regulatory effects—

commonly mediated by diffusible elements like transcription factors or RNAs—suggested that direct chromatin interactions may not fully account for the observed expression changes[260].

This emphasized the necessity to investigate transcription factor targets and the cis effects of rs6903956. In Chapter 2, rs6903956 was identified within a DNase hypersensitive site, underscoring its potential as a TFBS essential for gene regulation. In this chapter, utilizing the new A-to-G edited iPSC aEC model, we demonstrated that the rs6903956 A allele promotes binding of transcription factors HOXA4 and MEIS1, which are critical for cardiovascular development and endothelial function[261,262]. ChIP-qPCR assays confirmed allele-specific binding at rs6903956 A in aECs. Additionally, dual-luciferase assays connected this binding activity directly to increased *PHACTR1* expression, corroborating eQTL data from the HELIOS study that linked the A allele with elevated levels of *EDN1* and *PHACTR1* in Chinese individuals.

Functional assays further revealed that aECs carrying the A risk allele exhibited enhanced ICAM-1 levels and monocyte adhesion under disturbed flow conditions, implicating the rs6903956 A allele in promoting a pro-inflammatory endothelial phenotype. By integrating precise base editing techniques with a physiologically relevant aEC model, this chapter provides comprehensive mechanistic insights into the cis-regulatory role of rs6903956 and elucidates how modulation of *PHACTR1* expression by this SNP contributes to endothelial inflammation.

5.2 Methodology

5.2.1 Cloning sgRNA into pMIA20 ABE plasmid

A adenosine base editing system was used. Two gRNA duplex oligos with AaRI overhangs were annealed and subcloned into AaRI site and ligated with pMIA20 plasmid using T4 DNA ligase. pMIA20 plasmid was designed to contain EGFP fluorescence protein, a human codon-optimized Cas9(D10A) nickase and expression cassette for guide RNA. pMIA20 gRNA Cas9(D10A) plasmid was transformed into chemically competent E. coli (RapidTrans 96-Tube Chemically Competent E.coli; Active Motif, catalog no. 11096) and its plasmid DNA was extracted and purified using (Promega PureYield™ Plasmid Miniprep System, catalog no. A1223).

5.2.2 Single base editing on CAD iPSCs

A candidate gRNA was designed to match the 5'-to-3' DNA sequence containing rs6903956, located directly before an NGG PAM site: 5'-CATAGATTATTACTTAAGGT-3'. This gRNA duplex oligo was subcloned and ligated into the pMIA20 plasmid (provided by Assoc Prof Roger Foo, unpublished data) using T4 DNA ligase. $1.5x10^6$ cells were pre-treated with 10 μM ROCKi (StemCell Technologies, catalog no. 72302) for an hour. Single cell suspension was prepared using accutase (StemCell Technologies, catalog no. 07922) and resuspended in 100ul of nucleofection solution and 10ug of pMIA3 plasmid containing the selected gRNA pairs. Cells were nucleofected using Amaxa4D nucleofector (Lonza, catalog no. AAF-1002B) and P3 primary kit (Lonza, catalog no. V4XP-3024) as per manufacturer's instructions and plated on matrigel coated 6-well plate using mTeSR with CloneR supplement. After 48 hours, cells were again pre-treated with ROCKi and a single cell suspension was prepared. FACS was performed to enrich targeted cells. EGFP+ cells were sorted and plated onto 6-well plates containing mTeSR with CloneR and P/S. Colony formation was apparent from the individually sorted iPSCs 8 days after FACS. 24 single colonies then were manually picked into 12-well plates. Colonies were amplified and split with RelesR onto 6-well plates. Remaining cells were used for genomic DNA extraction and genotyping. Clones with successful CRISPR targeting were expanded.

5.2.3 TaqMan SNP Genotyping for rs6903956 base editing

Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen, catalog no. 69506). SNP genotyping was conducted with the TaqMan SNP Genotyping Assay, utilizing predesigned TaqMan primers

(Thermo Fisher, catalog no. 4351379) on a QuantStudio 6 instrument (Applied Biosystems). Each 20 µL reaction included 20 ng/µL of purified DNA and TaqMan Genotyping Master Mix (Thermo Fisher, catalog no. 4371353), prepared in a 96-well plate format. The rs6903956 genotypes were analyzed using the QuantStudio Design and Analysis Software integrated with the qPCR system.

5.2.4 Arterial endothelial cell differentiation

aEC differentiation from CAD patient-derived iPSCs was performed following the protocol by Ang Lay Teng et al.[180].

hiPSCs were dissociated using Accutase (Thermo Fisher, catalog no. 00-4555-56) and seeded at 100,000 cells/cm² on Matrigel-coated six-well plates (Corning, catalog no. 354277). CD144+ cells were isolated on Day 10 using MACS technology (Miltenyi Biotec, catalog no. 130-097-857), and flow cytometry was performed before and after enrichment. Enriched endothelial cells were further expanded in Endothelial Cell Growth Medium 2 (EGM2, Lonza, catalog no. CC-3162).

5.2.5 qPCR analysis

Real-time PCR was performed using SYBR green gene expression assays (New England Biolabs, catalog no. M3003S) on a QuantStudio 6 instrument (Applied Biosystems). Gene expressions were normalized to endogenous GAPDH housekeeping gene.

Primer sequences used in chapter 5 are as follows:
HOXA4 F: 5’-TGTCAGCGCCGTTAACCC-3’
HOXA4 R: 5’-GCCGGGTCAGGTATCGATTG-3’
MEIS1 F: 5’-CCAGTCCAACCGAGCAGTAA-3’
MEIS1 R: 5’-CACTCATAGGTCCTGGTGCT-3’
PHACTR1 F: 5’-CGAAGACGACGACAGCTCAT-3’
PHACTR1 R: 5’-TTCTTCCAGCTCTCGCTTGG-3’
EDN1 F: 5’-AAGGCAACAGACCGTGAAAAT-3’
EDN1 R: 5’-CGACCTGGTTTGTCTTAGGTG-3’
ADTRP F: 5’-GAAGACAGGACTCACCTTGCTG-3’
ADTRP R: 5’-GGCAAACACAGGATACACCCAG-3’
GAPDH F: 5’-CCGTCAAGGCTGAGAACGG-3’
GAPDH R: 5’-CTCAGCGCCAGCATCGC-3’

5.2.6 Transcription factor binding prediction

DNA motif of ~60bp flanking rs6903956 were used as input for Tomtom meme suite to identify potential TFs binding to rs6903956 region. Only the TF associations p <0.001 with E-value <10 were considered as statistically significant.

5.2.7 Alphafold 3

To analyze the structural context of rs6903956, a 101 bp sequence centred on the SNP (TTCCATCTCAAAAATAAATAAATAAATAAATAAATAAATAGTGCCATAG(*A/G)*TTATTACTTAAGGTTGGTCCCCCAAGTGTTGAAGTGGGTGCCATACTATAA) was input into the AlphaFold 3 model[263] via the AlphaFold Server with the seed set to 1. The model generated five predictions, pre-ranked based on overall complex confidence using the ranking_score metric.

SASA was used to define the interaction residues on the protein complex interacting with the 101 bp DNA molecule centered on rs6903956. Residues were classified as interface residues if at least 15 Å² of their surface area was buried upon interaction with the DNA molecule. The buried SASA for the entire interface was quantified, and chain-chain interactions were considered significant if they resulted in a buried SASA of at least 300 Å². This analysis helped identify and characterize the residues involved in DNA-protein interactions.

Residue interactions were analyzed based on spatial proximity. Residues were classified as potential contact residues if the distance between them was within 8 Å. The distance between chains and individual residues was investigated to identify and characterize interaction interfaces.

5.2.8 ChIP-qPCR

ChIP-qPCR was performed using High-Sensitivity ChIP Kit (Abcam, catalog no. ab185913), following manufacturer's instructions. Briefly, WT (AA), UNΔ (AA) and Δ (GG) aECs were harvested at a density of 5 x 10^5 cells per reaction. Cells were cross-linked with 1% formaldehyde and quenched with 1.25 M glycine. Cross-linked cells were then lysed, and chromatin was extracted. Chromatin was sheared using a probe sonicator set to 25% power output, applying 3-4 pulses of 10-15 seconds each, with 30-40 seconds rest on ice between pulses. The efficiency of shearing was confirmed via gel electrophoresis, ensuring

that the chromatin fragments were between 100-700 bp with a peak size of 300 bp.

ChIP was performed using 2 µg of sheared chromatin per well. Antibodies against HOXA4 (Santa Cruz Biotechnology, catalog no. sc-515418) and MEIS1/2/3 (Santa Cruz Biotechnology, catalog no. sc-101850) were used for the ChIP assays. The positive control used was an antibody against RNA polymerase II. The negative control was non-immune IgG, which serves as a baseline to measure non-specific binding. The antibodies were bound to strip wells coated with a protein A/G mix that has a high affinity for IgG antibodies. Chromatin samples were then added to these wells along with ChIP buffer, enrichment enhancer, and blocker solution, and incubated to allow antibody binding to the target chromatin fragments.

After incubation, the reaction wells were washed four times with 200 µL of Wash Buffer to remove non-specifically bound chromatin. This was followed by a single wash with DNA Release Buffer. Cross-links between DNA and proteins were reversed by incubating with RNase A solution at 42°C for 30 minutes, followed by Proteinase K at 60°C for 45 minutes. The DNA was then purified using a spin column and eluted with DNA Elution Buffer.

For qPCR analysis, input DNA (i.e., a fraction of the chromatin before immunoprecipitation) was used as a control to normalize the data. The purified DNA was quantified using Sybr Green qPCR primers for Allele-Specific qPCR[144] as well as TaqMan qPCR primers amplifying a 147bp region flanking rs6903956 A and G.

Allele-Specific qPCR primers: F (5′-GGGGACCAACCTTAAGTAATAACCTATG-3′), R (5′-GAGCCAAGATTGTGCCACTGC-3′).

TaqMan qPCR primers: F (5’-GCCTGTGAACAGTGAGAAGA-3’), Probe (5’-CCAAGTGTTGAAGTGGGTGCCATACT-3’), R (5’-GGGAACAGAGAGAGATTCCATC-3’).

qPCR reactions were run with an initial activation step at 95°C for 7 minutes, followed by 40 cycles of 95°C for 10 seconds, 55°C for 10 seconds, and 72°C for 8 seconds. The fold enrichment of target sequences was calculated using the amplification efficiency of the ChIP sample relative to the input DNA.

5.2.9 siRNA knockdown of HOXA4 and MEIS1

UNΔ (AA) and Δ (GG) aECs were seeded at a density of 1.3 x 10^5 cells per well in 24-well plates and incubated for 24 hours. SMARTPool siRNAs targeting HOXA4 and MEIS1 were used (Dharmacon, catalog no. L-011693-00-0005, L-011726-00-0005). siRNAs were resuspended by preparing a 20 µM stock solution. 5 nmol of siRNA was dissolved in 250 µL of 1x siRNA buffer (Dharmacon, catalog no. B-002000-UB-100). The solution was gently mixed and incubated on an orbital shaker for 30 minutes at room temperature. The concentration of siRNA was confirmed using spectrophotometry at 260 nm. Resuspended siRNAs were aliquoted and stored at -20°C.

For transfection, to obtain a final concentration of 5nM siRNA, 0.5 uL/well of 5uM stock siRNA was diluted in 49.5uL of Opti-MEM™ reduced serum medium (Thermo Fisher Scientific, catalog no. 31985062). 1uL/well of DharmaFECT 1 transfection reagent (Dharmacon, catalog no. T-2001-03) was used. The siRNA and transfection reagent mixtures were incubated separately for 5 minutes before combining and incubating for an additional 20 minutes at room temperature to form siRNA-lipid complexes. Transfection medium was prepared by mixing the siRNA-lipid complexes with EGM2 containing 10% HI FBS, resulting in a final siRNA concentration of 5 nM. The existing culture medium was removed from the cells, and the transfection medium was added. Cells were incubated for 72 hours. siRNA knockdown was confirmed and *PHACTR1* transcript levels quantified using qPCR.

5.2.10 Dual luciferase assay

The 519 bp ADTRP regulatory region flanking rs6903956 (-789 bp to +724 bp) (chr6:11,774,037–11,774,555) from Luo et al. was cloned into the pGL3 Control Vector (Addgene, Plasmid #212937) using KpnI and XhoI, upstream of the SV40 promoter and firefly luciferase coding region. Additionally, the same 519 bp region was cloned into the pGL3 Basic Vector (Addgene, Plasmid #212936) using XhoI and HindIII. Subsequently, a 1781 bp regulatory region of PHACTR1 (-766 bp to +1015 bp) (chr6:12,716,001–12,717,781) was subcloned into the pGL3-Basic vector. Cloning was performed by GenScript Biotech (Piscataway, NJ, USA).

The expression plasmids pcDNA3.1(-)-HOXA4 and pcDNA3.1(-)-MEIS1 were also constructed by GenScript Biotech.

HUVECs were seeded at a density of 5000 cells per well in a 96-well plate. 24 hours after seeding, cells were transfected with 50 ng of the respective reporter construct and 0.5 ng of the pNL1.1 PGK[Nluc/PGK] vector control (Promega, Madison, WI, USA, catalog no. N1441) using 0.5 µL/well of DharmaFECT 1 transfection reagent (Dharmacon, catalog no. T-2001-03). Salmon sperm DNA was used as carrier DNA to normalize plasmid DNA concentrations. Cells were harvested 72 hours post-transfection, and firefly and NanoLuc® luciferase levels were measured using the Nano-Glo® Dual-Luciferase® Reporter Assay System (Promega, catalog no. N1531). Firefly/NanoLuc® activity was determined for each sample using a Synergy H1 microplate reader (BioTek) and measured in triplicates.

5.2.11 Monocyte adhesion endothelial disturbed flow assay using Ibidi Flow System

Inverted µ-Slide γ-shaped slides (ibidi GmbH, catalog no. 80126) were coated with collagen I and seeded with iPSC-derived aECs (1.5×10^6 cells/mL) in Endothelial Cell Growth Medium (Lonza, catalog no. CC-3162) supplemented with ROCK inhibitor. After 24 hours of culture at 37°C and 5% CO_2, THP-1 monocytes (1.5×10^6 cells/mL) labeled with LuminiCell Tracker 670 (Millipore, catalog no. SCT011) were introduced into the assay. Shear stress of 12 dyn/cm² was applied for 24 hours using the ibidi Pump System (ibidi GmbH, catalog no. 10962). Post-assay, cells were fixed, permeabilized, and stained with primary and secondary antibodies, followed by DAPI staining. Imaging was performed using a Nikon Eclipse TiE microscope with Metamorph software.

Antibodies Used:

CD144: Rabbit polyclonal (Abcam, catalog no. ab33168), 0.1-1 µg/mL; secondary: Goat anti-rabbit IgG, Alexa Fluor™ 488 (Invitrogen, catalog no. A-11011), 2 µg/mL.

ICAM-1: Mouse monoclonal (Invitrogen, catalog no. MA5407), 1:250 dilution; secondary: Goat anti-Mouse IgG Alexa Fluor™ 488 (Invitrogen, catalog no. A-11029), 1:400 dilution.

5.2.12 Statistical analysis

All statistical analyses were performed using GraphPad Prism version 9. Data normality was evaluated using the Shapiro-Wilk test. For datasets with a normal distribution, unpaired two-tailed Student's t-tests were used to compare two groups, and one-way ANOVA followed by Tukey’s

post-hoc tests was applied for comparisons involving more than two groups. For non-normally distributed data, appropriate non-parametric tests, such as the Mann–Whitney U test for two-group comparisons or the Kruskal-Wallis test for multiple groups, were utilized. Statistical significance was defined as $p < 0.05$. Results are presented as mean ± SD. Specific statistical tests applied to individual experiments are detailed in the corresponding figure legends.

5.3 Results

5.3.1 Precise base editing of rs6903956 in arterial endothelial cells

To explore the regulatory role of rs6903956, we recognized the need to isolate its specific effects, addressing limitations in the broader NHEJ deletion approach described in Chapters 3 and 4. The previously generated Δ63–89bp deletions included rs6903956 and rs140361069 but could not pinpoint the precise impact of rs6903956 alone. Therefore, we employed adenosine base editing, a recent genome editing technology, to target rs6903956 directly[100]. This enabled precise conversion of the risk genotype (AA) to the non-risk genotype (GG) in CAD patient-derived iPSCs.

To achieve this, we employed single-base editing on CAD patient-derived iPSCs to convert the AA risk genotype to the GG non-risk genotype (Figure 20A). To enable precise single-base editing at the rs6903956 locus, we developed a system integrating D10A Cas9, a gRNA cassette, and an adenosine base editor (ABEmax) into a single construct (Figure 20B). This plasmid construct (pMIA13) offered advantages over traditional base-editing methods, which typically require co-transfection of multiple plasmids for Cas9, gRNA, and the base editor[100]. By integrating all components into a single plasmid, we improved transfection efficiency and reduced cellular stress.

Using this optimized system, we successfully base-edited iPSCs derived from two CAD patients (CAD donor 1 and CAD donor 2) (Figure 20C, D, E), converting the rs6903956 risk genotype (AA) to the non-risk genotype (GG). To control for potential off-target effects, we also generated unedited CAD iPSCs by subjecting them to the same CRISPR-Cas9 nucleofection protocol without achieving successful base editing.

We conducted initial screening using the TaqMan SNP genotyping assay, followed by Sanger sequencing to confirm both the accuracy of the base substitution and the absence of indels (Figure 20D). Editing success rates ranged between 30–35% for both donors. For further expansion, two successfully edited clones and one unedited clone per donor were selected based on the cleanliness of sanger sequencing results and maintenance of healthy iPSC morphology without signs of spontaneous differentiation.

This strategy enabled the generation of an isogenic CAD iPSC model, designated as WT AA for parental iPSCs, Δ GG for successfully edited iPSCs, and UNΔ AA for unedited controls.

Recognizing the limitations of the endothelial model generated in Chapter 3, particularly the need to study aECs in CAD, we differentiated these base-edited iPSCs into aECs using our recently established protocol developed in collaboration with Ang and colleagues[190] (Figure 20F). This protocol efficiently generates pure populations of human aECs from pluripotent stem cells within 3–4 days. AECs, unlike their venous counterparts, exhibit distinct functional and gene expression profiles that are critically relevant to CAD pathology[254-256]. These differentiated aECs consistently expressed >80% endothelial CD144 and artery marker DLL4, confirmed via flow cytometry, and were CD144+ MACS-sorted before downstream applications to ensure quality control (Figure 20G). The resulting cells were referred to as WT AA aECs, UNΔ AA aECs, and Δ GG aECs. We present a powerful base-edited aEC isogenic model to study the genetic basis of rs6903956 in endothelial dysfunction.

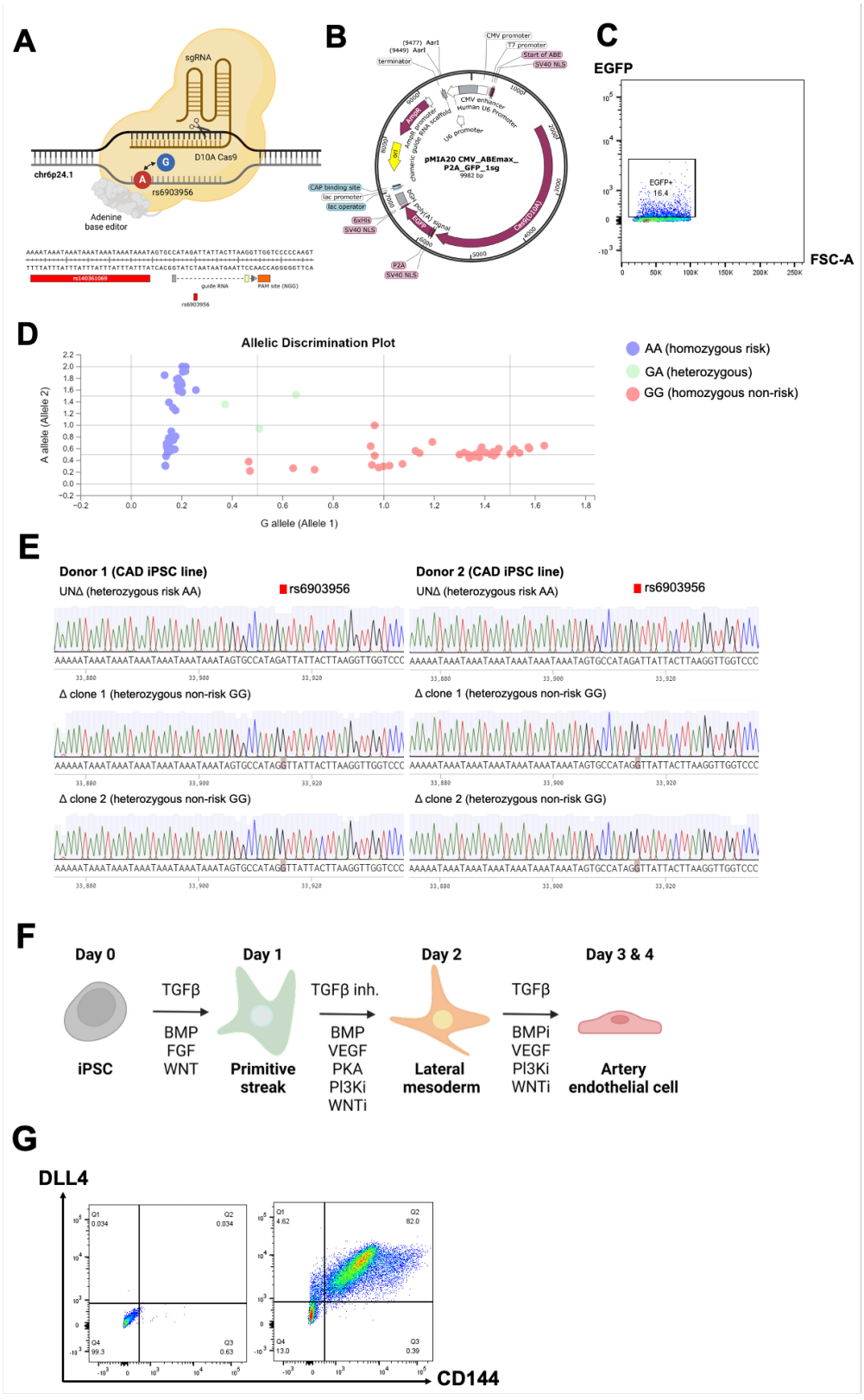


Figure 20. Generation and validation of iPSC-derived isogenic cell lines with rs6903956 single base edits from n=2 CAD patient. (A) Top: Schematic of single base editing targeting rs6903956 using pMIA13 with D10A Cas9, sgRNA targeting rs6903956 and adenosine base editor. Bottom: Guide RNA sequence designed for rs6903956 base editing, featuring NGG PAM site and rs6903956 in red. Tandem repeat rs140361069 is

also shown. **(B)** pMIA13 plasmid design. Empty backbone which expresses sgRNA cassette (AarI cloning site) and human U6 promoter driven D10A Cas9 linked with EGFP. **(C)** Flow cytometry analysis of iPSCs post-nucleofection with pMIA13 plasmid targeting rs6903956. Expression of EGFP indicates successful nucleofection of the pMIA13 plasmid containing the sgRNA targeting rs6903956. 16.4% of iPSCs (right) showed EGFP expression, compared to the control group (left), without nucleofection. **(D)** Allelic discrimination plot output from rs69039056 TaqMan genotyping assay, each dot representing a cell line expanded from a single cell isolated from FACS. Samples containing A allele (homozygote) are labelled blue, samples containing G allele (homozygote) are labelled red and samples which are heterozygous (GA) are labelled green. **(E)** Sequencing chromatograms validating single base edits at rs6903956 on hg38 (chr6:11,714,312–11,714,371). The rs6903956 variant is highlighted in red. Editing was performed on iPSCs derived from two CAD donors (n=2), resulting in two homozygous non-risk GG genotype clones (Δ clone 1 and Δ clone 2). An unedited control (UNΔ), carrying the heterozygous risk AA genotype, was subjected to pMIA13 nucleofection but did not achieve a successful edit. **(F)** Schematic of the 4-day differentiation protocol from iPSCs to aECS, adapted from Ang et al., 2022[190]. **(G)** Flow cytometry validation of aEC identity. Differentiated cells were analysed for the expression of arterial marker DLL4 and endothelial surface marker CD144 to confirm arterial endothelial identity. Control included undifferentiated iPSCs as negative control.

5.3.2 Base editing reveals rs6903956-mediated cis-regulation of PHACTR1 and EDN1 in arterial endothelial cells

In Chapter 3 and 4, we generated NHEJ lines (Δ63-89bp) that deleted both rs6903956 and rs140361069, a short tandem repeat in LD with rs6903956 in East Asian populations[212]. This study implicated *CXCL12*, a chemokine receptor, as a downstream target of rs6903956 through trans-effects. However, trans-effects are often indirect and mediated via diffusible elements acting on nearby cis-regulatory genes[264]. To address this, we employed our new base-editing system to specifically investigate the cis-effects of rs6903956.

To explore these cis effects, we examined the 3D chromatin architecture surrounding rs6903956. Hi-C contact maps from Chapter 4 revealed that rs6903956 resides within a TAD encompassing neighbouring genes such as *EDN1* and *PHACTR1*[212]. Given that SNPs frequently regulate genes within the same TAD, we hypothesized that the A-to-G substitution at rs6903956 could influence the expression of these neighbouring genes. Using our base-edited aEC isogenic model, we aimed to test this hypothesis.

We performed quantitative RT-PCR on WT AA, UNΔ AA, and Δ GG aECs to assess gene expression of the cis-genes (Figure 21A). Although *ADTRP*, located within the first intron of rs6903956, was previously linked to reduced expression of the A risk allele in leukocytes[18], we found no significant differences in ADTRP expression across the three cell lines. These results indicate that rs6903956 may

not influence *ADTRP* expression in endothelial cells, suggesting cell-type specificity in its regulatory effects.

In contrast, significant downregulation of *PHACTR1* and *EDN1* expression was observed in Δ GG compared to WT AA and UNΔ AA aECs. Since *PHACTR1* overexpression has been shown to upregulate *EDN1*[265], it is possible that one is the primary causal gene while the other might be a downstream effect. *PHACTR1* locus has been identified as a 6p24.1 risk locus for five vascular diseases, with the lead SNP rs9349379 linked to a putative enhancer for *EDN1*[7], whereas *EDN1* has not been implicated in large-scale GWAS studies for CAD. Multiple studies have associated *PHACTR1* polymorphisms with blood pressure traits, including hypertension, systolic blood pressure, and pulse pressure[266-269]. Consistent with this, our Guilin cohort identified a significant association between the number of 'A' alleles at rs6903956 and diastolic blood pressure, supporting the role of *PHACTR1* in blood pressure regulation (Table 4). Importantly, genotyping confirmed that none of the cell lines used carried *PHACTR1* hypertension risk alleles (Table 9), demonstrating that the observed effects are specific to rs6903956 and not linked to known *PHACTR1* risk variants.

We hypothesize that individuals carrying A risk allele would have increased *EDN1* and *PHACTR1* expression. As such, we performed an eQTL analysis using data from the HELIOS cohort, our longitudinal study comprising 10,004 individuals of Chinese, Indian, and Malay descent. In a subset of 850 Chinese participants, we examined the PBMC gene expression. After adjusting for age and gender, we observed an additive effect of the rs6903956 risk allele A on *EDN1* and *PHACTR1* expression (Figure 21B). Taken together, individuals carrying A allele exhibited higher normalized expression of *EDN1* and *PHACTR1*, mirroring the trend observed in our base-edited aEC models.

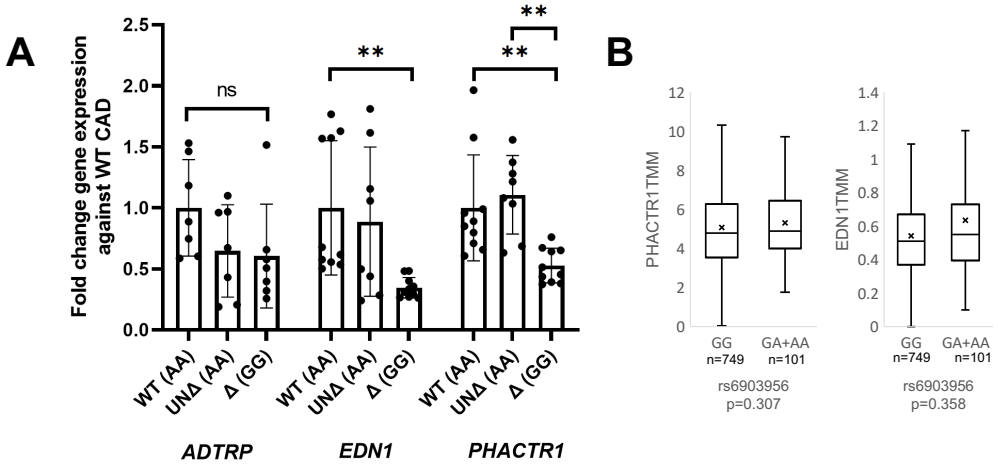


Figure 21. Isogenic base-edited aECs to study the *cis*-regulation of rs6903956. (A) Quantitative RT-PCR of rs6903956 cis genes in WT, UNΔ (AA) and Δ CAD (GG) aECs. Bar graphs showing means with S.D. (*n* = 2 from 2 donor cell lines, n = 2-3 differentiation batch, n=2 technical replicate), $^{**}p \leq 0.01$, Kruskal-Wallis test. **(b)** eQTL analysis of *EDN1*, *PHACTR1*, and *ADTRP* in the HELIOS Chinese cohort, adjusted for gender and ethnicity, included genotypes GG (n=749), GA (n=99), and AA (n=2). Data were normalized using Trimmed Mean of M-values (TMM) and analysed using Mann-Whitney U test.

eQTL RNA Sequencing from the HELIOS cohort was performed by Dr Darwin Tay Jia Xian and Asst Prof Marie Loh.

	Donor 1	Donor 2	Major allele	Minor allele	Disease link	Citation
rs9369640	A	A	A	C	Minor allele C is protective for hypertension	PMC4251179
rs1223397	G	G	G	C	rs1223397 risk allele is significantly associated with pulse pressure.	PMC7276519
rs9349379	A	A	A	G	In European: Minor allele G at rs9349379 links to increased vascular disease risk due to higher ET-1 levels.	PMC5785707

Table 9. Genotype of CAD patient donor 1 and donor 2 for PHACTR1 risk variants implicated in blood pressure phenotypes.

5.3.3 HOXA4 and MEIS1 bind specifically to rs6903956 A allele in AlphaFold 3 analysis

To dissect how the rs6903956 AA risk genotype, located ~500 kb from *EDN1* and ~1,000 kb from *PHACTR1*, could affect long-range gene expression, we explored its potential role in influencing TFBS. Regulatory variants such as rs6903956 often impact gene expression by altering TFBSs—either by disrupting, creating, or modulating the affinity of these sites. These changes can drive gene expression changes through cis or trans mechanisms[161].

In Chapter 2, we identified rs6903956 as a critical nucleotide within a DHS using DNase I footprinting data from HUVEC cells, part of the ENCODE Project (GEO accession: GSM1014528) (Figure 22A). DHS regions, which mark active chromatin sites, frequently overlap with TFBSs, suggesting that rs6903956 may function as a regulatory hotspot. This observation raises the possibility that rs6903956 contributes to transcriptional regulation by altering TFBSs and mediating cooperative enhancer-like activity to influence gene expression in both cis and trans.

Using the TOMTOM algorithm for *in silico* TFBS mapping[270], we sought to identify specific transcription factors that exhibit differential binding to the A or G allele of rs6903956. Transcription factors HOXA4 and MEIS2

were predicted to preferentially bind the risk 'A' allele (Figure 22B). HOX and MEIS proteins are well-known transcription factors that regulate key developmental processes, including cardiovascular diseases[262]. HOX proteins typically exhibit low-affinity DNA binding on their own but achieve stable and high-affinity interactions when paired with TALE homeodomain proteins, such as MEIS, which enhance cooperative binding and regulatory function[271].

To determine whether transcription factors that specifically bind rs6903956 A might be responsible for influencing *PHACTR1* or *EDN1* expression, we analysed TFBS in their promoter regions (Figure 22C). Using the GeneCards Enhancer Database and QIAGEN TFBS prediction tools, we identified binding sites for HOXA9 and MEIS1 in the *PHACTR1* promoter. However, no similar TFBS were identified in the *EDN1* promoter, indicating that rs6903956 does not regulate *EDN1* through HOX-MEIS interactions. As a result, *EDN1* was excluded from further analysis. These findings propose a model where rs6903956 modulates *PHACTR1* regulatory activity through cooperative HOX-MEIS binding[272].

The structural similarities among members of the HOX and MEIS transcription factor families present challenges for *in silico* predictions, as TFBS motifs often cannot reliably differentiate between closely related family members (Figure 22D). To overcome this ambiguity, we utilized AlphaFold 3, an advanced deep learning algorithm for 3D molecular interaction modelling, to elucidate the most plausible transcription factor-DNA interactions at the rs6903956 locus[263].

We focused on a 101-bp DNA sequence surrounding rs6903956, centred at position 50, to model its interactions with specific HOX proteins (HOXA4 and HOXA9) and MEIS proteins (MEIS1 and MEIS2) (Figure 22E). Simulations were performed for both the G and A alleles at rs6903956. The output generated five plausible structural models for each condition, sampling diverse conformational spaces. A critical metric used in these simulations was the minimum distance, defined as the shortest spatial distance between rs6903956 and the nearest amino acid residue of the interacting protein. A lower minimum distance suggests a higher likelihood of potential physical or functional interaction between the DNA and protein. We define two residues to be “in contact” if their distance is less than 8 Å[263].

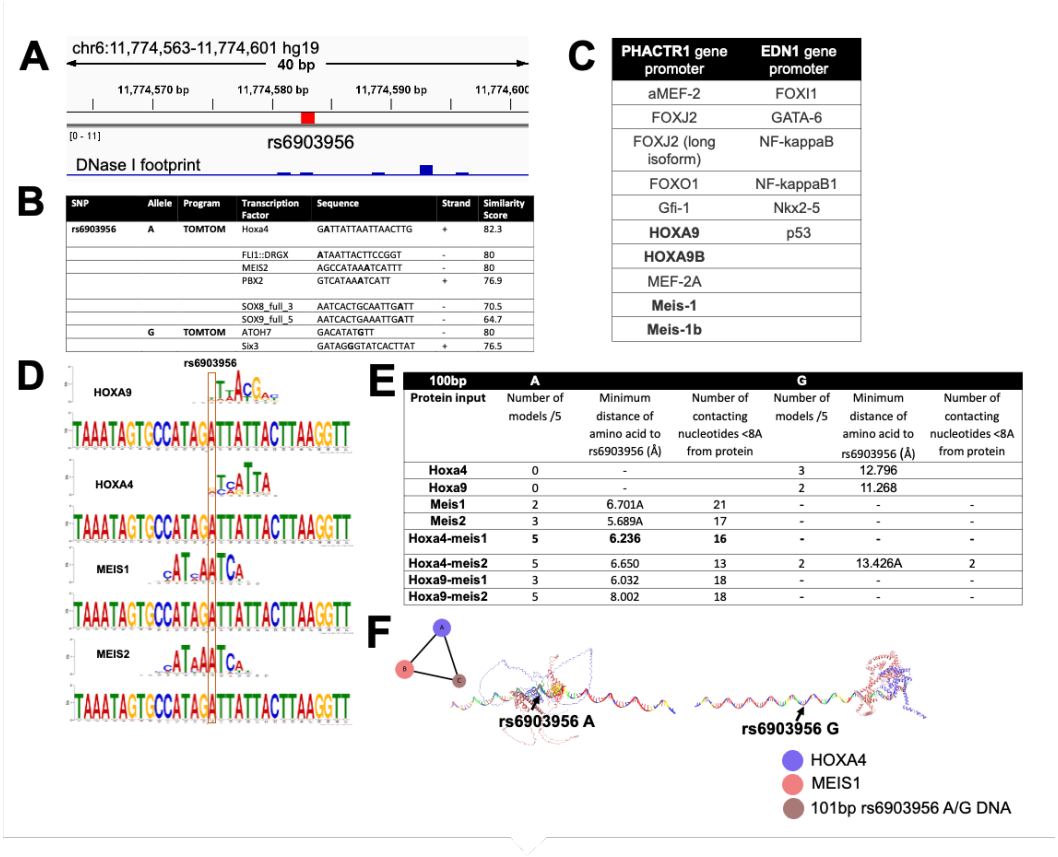
AlphaFold 3 simulations revealed distinct binding preferences for transcription factors at rs6903956 based on the allele (G/A) present (Figure 22F). For individual transcription factors, HOXA4 and HOXA9 demonstrated low-affinity binding to the 'G' allele, with minimum

distances of 12.796Å and 11.268Å respectively. In contrast, MEIS1 and MEIS2 showed exclusive binding to the A allele, with minimum distances of 6.701 Å and 5.689 Å respectively.

When modelled as combinations, HOX and MEIS formed a cooperative tricomplex that exhibited strong binding specificity for rs6903956 A. Notably, HOXA4 and HOXA9, which initially bound rs6903956 G with low affinity and did not bind A, transitioned to high-affinity binding to rs6903956 A in the presence of MEIS1 or MEIS2. This interaction highlights a cooperative mechanism, where MEIS proteins facilitate the recruitment and stabilization of HOX proteins at their specific binding site.

To prioritize the most reliable predictions, we focused on complexes that consistently bound rs6903956 A across all five simulation models. This narrowed the candidates to two key complexes: HOXA4-MEIS1-101bp rs6903956 A DNA and HOXA9-MEIS2-101bp rs6903956 A DNA. Among these, HOXA4-MEIS1-101bp rs6903956 A DNA complex exhibited the lowest minimum distance (6.236Å) between rs6903956 and the nearest interacting amino acid, indicating it as the strongest and most specific binding configuration.

Focusing on HOXA4 and MEIS1, the HOXA4-MEIS1-101bp rs6903956 A DNA complex demonstrated ternary complex formation with direct interactions between all three components—HOXA4, MEIS1, and DNA (Figure 22F, left). In contrast, with 101bp rs6903956 G DNA, HOXA4 and MEIS1 failed to bind at the rs6903956 site (Figure 22F, right). These results from AlphaFold 3 structural predictions highlight the critical role of the rs6903956 A allele in facilitating the binding of HOXA4 and MEIS1.



SNP	Allele	Program	Transcription Factor	Sequence	Strand	Similarity Score
rs6903956	A	TOMTOM	Hoxa4	GATTATTAATTAACTTG	+	82.3
			FLI1::DRGX	ATAATTACTTCCGGT	-	80
			MEIS2	AGCCATAAATCATTT	-	80
			PBX2	GTCATAAATCATT	+	76.9
			SOX8_full_3	AATCACTGCAATTGATT	-	70.5
			SOX9_full_5	AATCACTGAAATTGATT	-	64.7
	G	TOMTOM	ATOH7	GACATATGTT	-	80
			Six3	GATAGGGTATCACTTAT	+	76.5

PHACTR1 gene promoter	EDN1 gene promoter
aMEF-2	FOXI1
FOXJ2	GATA-6
FOXJ2 (long isoform)	NF-kappaB
FOXO1	NF-kappaB1
Gfi-1	Nkx2-5
HOXA9	p53
HOXA9B	
MEF-2A	
Meis-1	
Meis-1b	

100bp	A			G		
Protein input	Number of models /5	Minimum distance of amino acid to rs6903956 (Å)	Number of contacting nucleotides <8A from protein	Number of models /5	Minimum distance of amino acid to rs6903956 (Å)	Number of contacting nucleotides <8A from protein
Hoxa4	0	-		3	12.796	
Hoxa9	0	-		2	11.268	
Meis1	2	6.701A	21	-	-	-
Meis2	3	5.689A	17	-	-	-
Hoxa4-meis1	**5**	**6.236**	**16**	**-**	**-**	**-**
Hoxa4-meis2	5	6.650	13	2	13.426A	2
Hoxa9-meis1	3	6.032	18	-	-	-
Hoxa9-meis2	5	8.002	18	-	-	-

Figure 22. Transcription factor motif and structural prediction using TOMTOM and AlphaFold 3 for rs6903956 with HOX and MEIS proteins. (A) DNAse I footprinting of HUVEC on hg19 chr6:11,774,470-11,774,700. **(B)** Predicted transcription factor motifs based on a 31bp sequence surrounding rs6903956, as analyzed by TOMTOM (https://meme-suite.org/meme/tools/tomtom). The similarity score reflects the alignment percentage between the transcription factor motif and the provided sequence. **(C)** Predicted binding of transcription factors to *PHACTR1* and *EDN1* gene promoter using GeneCards Enhancer Database and QIAGEN TFBS prediction tools. **(D)** Visualization of transcription factor binding for HOXA4, HOXA9, MEIS1, and MEIS2 on the rs6903956 A allele. Nucleotide size variations represent the position frequency matrix (PFM). Rs6903956 is boxed in red. **(E)** Structural prediction of rs6903956 A/G and flanking sequences with transcription factors HOXA4, HOXA9, MEIS1, MEIS2, and their complexes. AlphaFold 3^{263} was used to predict interactions between 101 bp flanking rs6903956 (with the variant as the 50th nucleotide) and individual transcription factors (HOXA4, HOXA9, MEIS1, MEIS2) or their complexes. The output includes predictions from five AlphaFold models, with the number of models supporting significant binding (<8 Å minimum distance from rs6903956) shown. Nucleotides within <8 Å of the protein are defined as the TFBS. **(F)** Model predicting HOXA4 and MEIS1 binding to 101bp flanking rs6903956 A (left), and G (right). HOXA4 is shown in purple, MEIS1 in pink, 101bp DNA strand in rainbow / brown. The triangle structure shows how the three components interact as a ternary complex.

Alphafold 3 analysis was performed by Nguyen Le Uyen Nhi.

5.3.4 HOXA4 and MEIS1 bind rs6903956 A allele to enhance PHACTR1 transcription regulation in aECs

Earlier *in silico* TFBS and protein structural predictions suggested that the rs6903956 A allele creates novel TFBS for HOXA4 and MEIS1 transcription factors, which are key regulators of cardiovascular development and endothelial cell function. To experimentally validate the *in vivo* binding of rs6903956 A with HOXA4 and MEIS1, ChIP assays were performed using antibodies against HOXA4 and MEIS1 in WT AA, UNΔ AA, and Δ GG aECs (Figure 23A). The immunoprecipitated DNA was amplified using TaqMan qPCR to quantify the amount of chromatin bound to each transcription factor (Figure 23B, C).

ChIP-TaqMan qPCR results demonstrated significantly stronger binding of HOXA4 and MEIS1 to the genomic fragment harbouring the rs6903956 A allele in WT AA and UNΔ AA aECs compared to rs6903956 G allele in Δ GG aECs (Figure 23D). These findings support the hypothesis that the rs6903956 A allele establishes TFBS for HOXA4 and MEIS1, facilitating their allele-specific binding.

To investigate our hypothesis that rs6903956 A leads to the upregulation of *PHACTR1* through HOXA4 and MEIS1 binding, we performed siRNA-mediated knockdowns of HOXA4 and MEIS1 in UNΔ AA and Δ GG aECs and measured *PHACTR1* expression using qRT-PCR (Figure 23E). In UNΔ AA aECs, HOXA4 knockdown resulted in a significant reduction in *PHACTR1* expression compared to scrambled control (Figure 23F). In Δ GG aECs, however, no reduction in *PHACTR1* expression was observed, confirming that HOXA4-mediated upregulation of *PHACTR1* is specific to the rs6903956 A allele.

Despite ChIP-qPCR confirming the binding of MEIS1 to the rs6903956 A allele, knockdown of MEIS1 did not significantly decrease *PHACTR1* expression. This suggests that MEIS1 might play a secondary or supportive role in transcriptional regulation of *PHACTR1*. MEIS proteins are often known to function as cofactors, stabilizing HOX-DNA interactions rather than directly driving transcription[273]. Additionally, functional redundancy within the MEIS family, such as compensatory effects from MEIS2, may explain this observation[274,275]. Interestingly, a significant increase in *PHACTR1* expression was observed upon MEIS1 knockdown. MEIS1 depletion has been reported to promote a shift toward oxidative phosphorylation, resulting in increased production of ROS[276]. ROS is a known trigger of endothelial inflammation, and *PHACTR1* has been shown to be upregulated under inflammatory conditions such as exposure to TNF-α, IL-1β, and ox-LDL[277]. This

suggests that MEIS1 knockdown may indirectly promote *PHACTR1* expression through ROS-mediated inflammatory signalling.

Collectively, these results support a model where the G-to-A substitution at rs6903956 generates TFBS for HOXA4, facilitating an potential enhancer-promoter interaction that drives *PHACTR1* expression. In Δ GG aECs, the absence of these TFBS prevents engagement of this regulatory pathway.

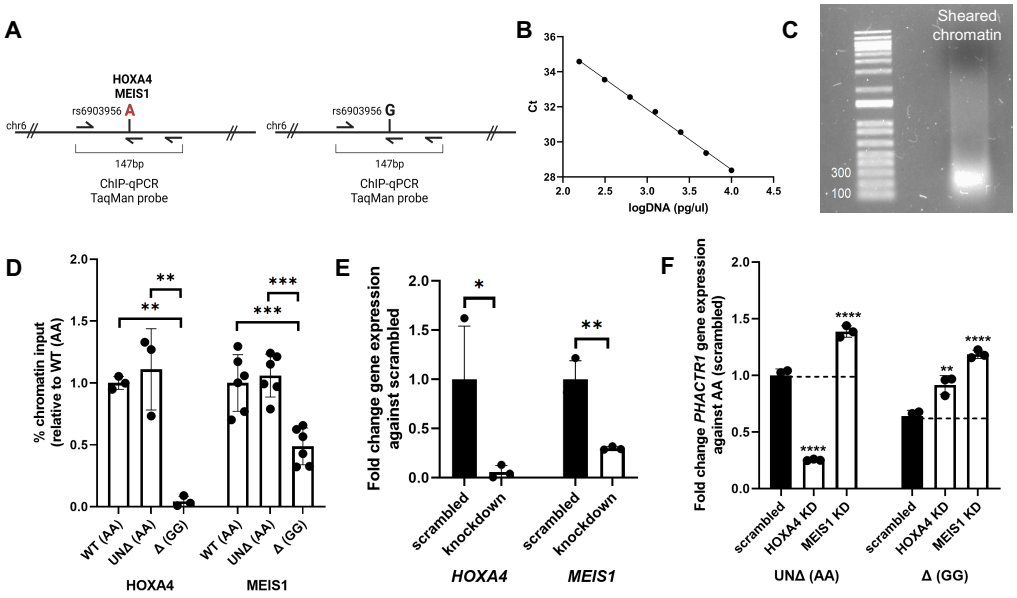


Figure 23. Experimental validation of rs6903956 as a TFBS for HOXA4 and MEIS1 in aECs. (A) Schematic representation of the ChIP-qPCR assay design targeting the rs6903956 enhancer region. Assay setup for wild-type WT AA, UNΔ AA, and Δ GG aECs. Anti-HOXA9 and MEIS1/2 antibodies were used for chromatin immunoprecipitation, followed by TaqMan qPCR amplification targeting a 147 bp region flanking rs6903956. **(B)** Primer dilution curve for ChIP-TaqMan qPCR primer to validate specificity of primers for TaqMan qPCR. DNA used is iPSC genomic DNA averaged from donor 1 and 2. **(C)** Gel electrophoresis estimate of size of 100-300bp sheared chromatin DNA followed by formaldehyde fixing and probe sonication for ChIP-TaqMan qPCR. **(D)** Bar graphs showing the percentage of chromatin input for HOXA4 and MEIS1 for WT AA, UNΔ AA, and Δ GG aECs relative to WT AA aEC. Data represent mean ± S.D. (n=1-2 two donor cell lines, n=3 three well replicate), **p ≤ 0.01, one-way ANOVA with post-hoc Tukey's tests. **(E)** Fold-change gene expression levels of *HOXA4* and *MEIS1* were measured relative to scrambled control siRNA using qRT-PCR, 72 hours post-siRNA knockdown. Expression levels are normalized to the scrambled control. Bar graphs represent means ± S.D. (n = 3 biological well replicates, average of n = 3 technical qPCR replicates), *p ≤ 0.05, **p ≤ 0.01, two-tailed t-tests compared to scrambled control. **(F)** Quantitative RT-PCR analysis of *PHACTR1* expression following siRNA-mediated knockdown of HOXA4 and MEIS1 in UNΔ AA and Δ GG aECs. Expression levels are normalized to scrambled control. Bar graphs represent means ± S.D. (n = 3 biological well replicate, average of n=3 technical qPCR replicate), **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001, two-tailed t-tests compared to scrambled.

5.3.5 *Rs6903956 A allele is located in an enhancer region which drives PHACTR1 expression via HOXA4-MEIS1 complex formation*

To investigate the enhancer activity of rs6903956, we performed dual-luciferase assay with a 519 bp fragment flanking rs6903956. The fragment was cloned into a pGL3-Control vector upstream of an SV40 promoter and transfected into HUVECs (Figure 24A, left). The rs6903956 A allele fragment significantly increased luciferase activity compared to the control plasmid lacking the enhancer region, confirming that the rs6903956 A allele fragment confers enhancer activity to the SV40 promoter (Figure 24B).

To determine whether the enhancer region flanking rs6903956 specifically regulates the *PHACTR1* promoter in an allele-specific manner and to investigate the cooperative roles of HOXA4 and MEIS1, we cloned the same 519 bp fragment (A/G alleles) upstream of the *PHACTR1* promoter in the pGL3-Basic vector (Figure 23A, right). Co-transfections using HUVEC with HOXA4 and MEIS1 expression plasmids were performed individually and in combination.

Without transcription factors, there was no significant difference in luciferase activity between the enhancer region flanking either rs6903956 A or G alleles, indicating that the enhancer activity is transcription factor dependent. Similarly, the expression of HOXA4 alone did not result in significant transcriptional activity, consistent with prior AlphaFold predictions that HOXA4 requires MEIS proteins for high-affinity binding to rs6903956 A allele. When MEIS1 was expressed alone, a modest but statistically insignificant increase in transcriptional activity was observed for both the enhancer region flanking A and G alleles. However, co-expression of HOXA4 and MEIS1 led to significantly higher transcriptional activity for *PHACTR1* promoter observed for the enhancer region flanking rs6903956 A allele compared to the G allele (Figure 23C), indicating that establishment of the HOXA4-MEIS1 complex is critical for *PHACTR1* transcriptional activity with the enhancer region flanking rs6903956 A allele.

Notably, the transcriptional activity of the enhancer region flanking rs6903956 A allele with the *PHACTR1* promoter was significantly higher than with the SV40 promoter (Figure 23D), suggesting that rs6903956 functions as a locus-specific enhancer to regulate *PHACTR1* expression in endothelial cells.

Overall, these findings establish that the region flanking rs6903956 A allele functions as an enhancer that facilitates the recruitment of HOXA4 and MEIS1, driving *PHACTR1* expression in aECs.

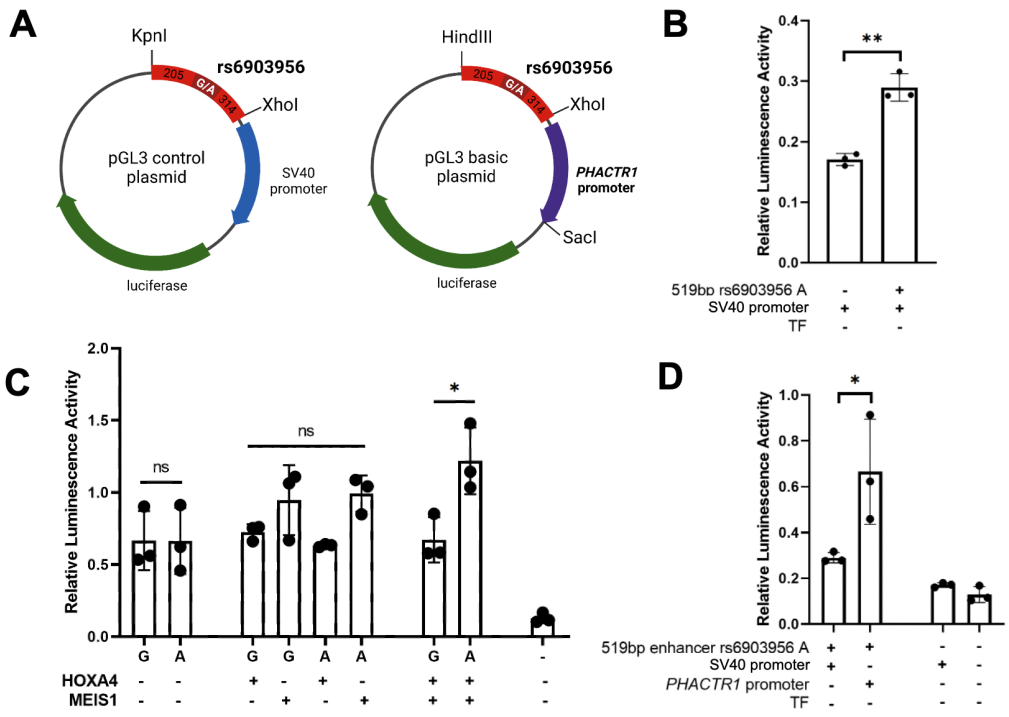


Figure 24. Dual luciferase assay to prove enhancer role of rs6903956. (A) Schematic summary of pGL3 plasmids used in the dual-luciferase assay. The pGL3 control plasmid was engineered to include a 519 bp region flanking rs6903956, cloned using KpnI and XhoI restriction sites under the control of the SV40 promoter. For the pGL3 basic plasmid, the 519 bp flanking region of rs6903956 was cloned using HindIII and XhoI, and the *PHACTR1* promoter was inserted using XhoI and SacI. These constructs were utilized for dual-luciferase assays to assess enhancer activity. **(B)** Relative luminescence activity of pGL3 control plasmids with and without the 519 bp region flanking rs6903956 A transfected into HUVECs. Effect of the 519 bp region flanking rs6903956 A on firefly luciferase activity under SV40 promoter control in the pGL3 control plasmid. Firefly luciferase activity was normalized to Renilla luciferase activity using the pNL1.1 PGK3 plasmid as an internal control. Bar graphs represent means ± S.D. (n = 3 biological well replicates, average of n = 3 technical luciferase reads), **p ≤ 0.01, two-tailed t-tests. **(C)** Relative luciferase activity of pGL3 basic plasmids containing the 519 bp region flanking rs6903956 (alleles A and G) and the *PHACTR1* promoter. The plasmids were transfected with different combinations of HOXA4 and MEIS1 (single and co-transfections) into HUVECs. The control on the far right represents the pGL3 basic plasmid without an enhancer or promoter cloned. Firefly luciferase activity was normalized to Renilla luciferase activity using the pNL1.1 PGK3 plasmid as an internal control. Bar graphs represent means ± S.D. (n = 3 biological well replicates, average of n = 3 technical luciferase reads), *p ≤ 0.05, ns = not significant, two-tailed t-tests. **(D)** Relative luciferase activity of pGL3 plasmids with 519 bp region flanking rs6903956 A under SV40 and *PHACTR1* promoter control. Relative luciferase activity of the pGL3 control plasmid with the 519 bp region flanking rs6903956 A cloned under

SV40 promoter control and the pGL3 basic plasmid with the same enhancer region cloned under the *PHACTR1* promoter transfected into HUVECs. Controls on the right represent the pGL3 control plasmid without the enhancer and the pGL3 basic plasmid without both enhancer and promoter cloned. Firefly luciferase activity was normalized to Renilla luciferase activity using the pNL1.1 PGK3 plasmid as an internal control. Bar graphs represent means ± S.D. (n = 3 biological well replicates, average of n = 3 technical luciferase reads), *p ≤ 0.05, two-tailed t-tests.

5.3.6 UNΔ AA aECs express increased endothelial activation and THP-1 monocyte adhesion compared to Δ GG aECs under disturbed flow

Endothelial *PHACTR1* is known to mediate inflammation by activating NF-κB-dependent ICAM-1 expression under disturbed flow, with si*PHACTR1* reducing ICAM-1 levels in HUVECs[265]. Overexpression of *PHACTR1* also promotes inflammation and monocyte adhesion[277]. To investigate the endothelial phenotypic effects of rs6903956 and the causal gene *PHACTR1*, we hypothesized that UNΔ AA aECs, which express higher *PHACTR1* levels compared to Δ GG aECs, would exhibit enhanced ICAM-1 protein expression and increased monocyte adhesion under disturbed flow.

UNΔ AA and Δ GG aECs were exposed to oscillatory disturbed flow (±6 dynes/cm², 1 Hz) in the presence of THP-1 monocytes, followed by immunofluorescence staining for ICAM-1 and quantification of adhered monocytes (Figure 25A). Compared to Δ GG aECs, UNΔ AA aECs exhibited a 4.18-fold increase in ICAM-1 expression and a 4.25-fold increase in adhered THP-1 monocytes (Figure 25B, C), demonstrating an allele-specific endothelial phenotypic response to disturbed flow. The results suggest that UNΔ AA aECs with elevated *PHACTR1* levels have increased inflammation and monocyte adhesion under disturbed flow conditions compared to Δ GG aECs.

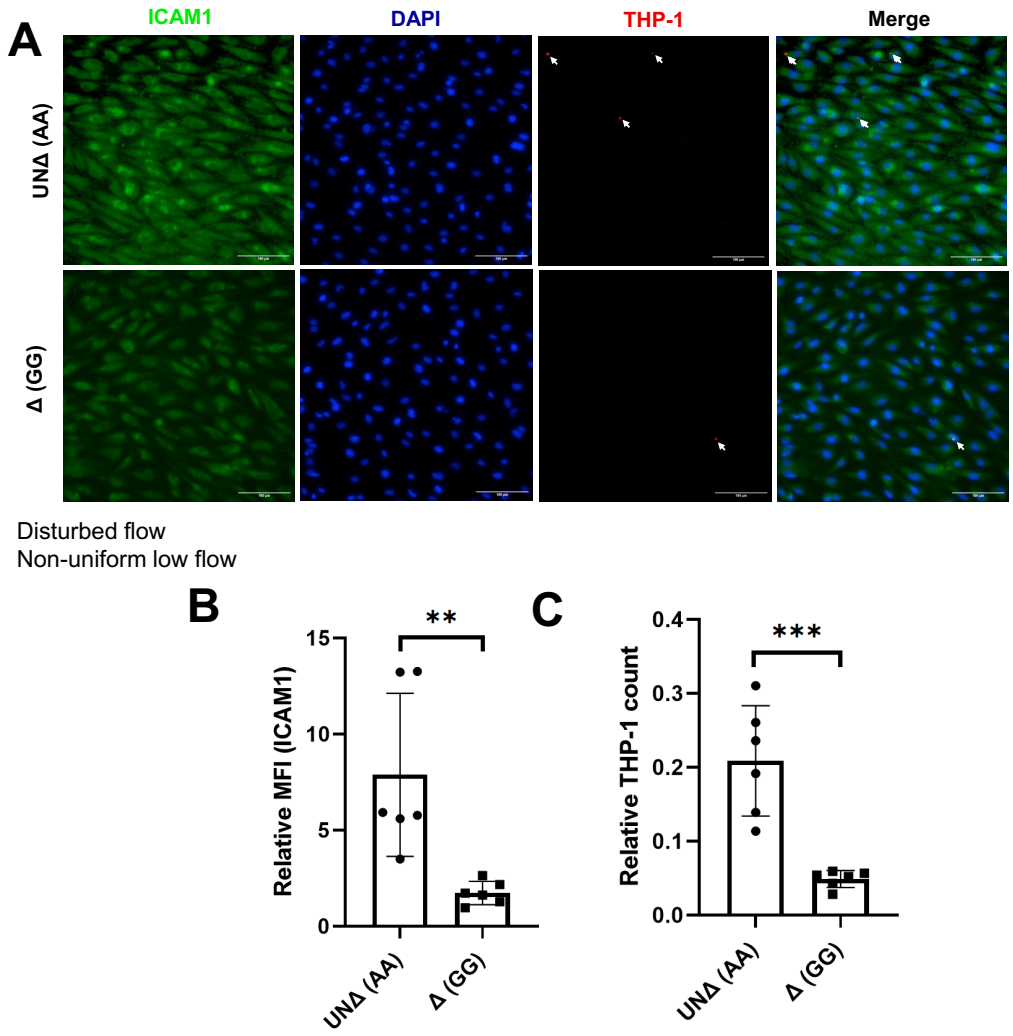


Figure 25. ICAM-1 expression and monocyte adhesion in Δ (GG) compared to UNΔ (AA) aECs under disturbed flow conditions. (A) Representative immunofluorescence images of ICAM-1 (green), DAPI (blue), and THP-1 monocyte adhesion (red) in UNΔ AA and Δ GG aECs. Scale bar: 100 μm. (B) Quantification of relative ICAM-1 mean fluorescence intensity (MFI), normalised to DAPI MFI, in UNΔ (AA) and Δ (GG) cells. Bar graphs represent means ± S.D. (n = 3 regions of interest (ROIs) per biological replicate from 2 donors). **p ≤ 0.01, Mann-Whitney U test. **(C)** Quantification of relative THP-1 monocyte adhesion, normalised to the number of endothelial cells, in UNΔ (AA) and Δ (GG) cells. Bar graphs represent means ± S.D. (n = 3 ROIs per biological replicate from 2 donors). ***p ≤ 0.001, two-tailed t-tests.

Monocyte perfusion under disturbed endothelial flow conditions and ICAM-1 staining was performed by Hannah Wee Su-Ann.

5.4 Discussion

Understanding the regulatory mechanisms of non-coding variants identified by GWAS remains a critical challenge, particularly in the context of cardiovascular diseases like CAD. While over 90% of GWAS-identified variants reside in non-coding regions, their functional interpretation often requires experimental models that integrate genomic, transcriptomic, and cellular insights. In this chapter, we aimed to elucidate the specific mechanism of rs6903956 in regulating CAD-associated cis genes through comprehensive analyses combining base editing, eQTL data, and functional assays.

Previous work by Luo et al.[144] suggested that rs6903956 plays an enhancer regulatory role by impacting *ADTRP* expression through disruption of a GATA2 binding site, as evidenced by dual-luciferase assays in HeLa cells. However, their findings lacked confirmation in endothelial cells, the primary cell type implicated in CAD pathogenesis, and did not establish rs6903956 as an enhancer independent of the promoter. In contrast, our study provides direct evidence for the enhancer activity of rs6903956 in endothelial cells. By cloning a 519 bp fragment flanking rs6903956 into a pGL3-basic vector and conducting dual-luciferase assays in HUVECs, we demonstrated that the rs6903956 A allele significantly enhances transcriptional activity compared to a control lacking the enhancer region.

Using eQTL data from the HELIOS cohort, the largest study to date on rs6903956 in the Chinese population, we established that the rs6903956 A risk allele upregulates *PHACTR1* expression. This was further supported by findings in the Guilin, Guangxi, China population study, where the rs6903956 A allele was associated with increased diastolic blood pressure. Functional validation revealed that rs6903956 enhances *PHACTR1* expression by recruiting HOXA4 and MEIS1 transcription factors, as confirmed by ChIP-qPCR, siRNA knockdown experiments, and dual-luciferase assays. These results directly link the rs6903956 A allele to the activation of a HOXA4-MEIS1-mediated enhancer-promoter interaction that drives *PHACTR1* transcription.

Our findings underscore the complexity of interpreting GWAS-identified variants across cell types. While Wang et al. linked rs6903956 A to decreased *ADTRP* expression in peripheral blood leukocytes[18], neither our NHEJ-edited nor base-edited endothelial cell models revealed significant changes in *ADTRP* expression. Moreover, our eQTL analysis in the HELIOS cohort showed that the GG genotype correlated with lower *ADTRP* expression, suggesting that rs6903956 exhibits cell-type-specific regulatory effects. CAD involves various cell types, including

endothelial cells, smooth muscle cells, and macrophages[278], which may explain the divergent regulatory outcomes observed in different contexts.

Our study provides compelling evidence that rs6903956 influences endothelial cell function through *PHACTR1*, a gene located approximately 942 kb downstream within the same TAD. *PHACTR1* is an actin- and protein phosphatase 1 (PP1)-binding protein involved in F-actin remodelling and PP1 regulation, with the *PHACTR1* locus at 6p24.1 implicated in susceptibility to several vascular diseases, including CAD[47,133,269,279-282]. First characterized in endothelial cells using Suppression Subtractive Hybridisation (SSH), *PHACTR1* has been extensively studied in the context of CAD. Studies have shown that *PHACTR1* knockdown in HUVECs significantly impairs tube formation and induces apoptosis[283] while knockdown in HAECs diminishes endothelial adhesion molecules, such as ICAM-1 and VCAM-1, underscoring its role in oxidative stress and inflammatory responses[284]. PHACTR1 also contributes to disturbed flow-mediated atherosclerosis by repressing PPARγ activity[265].

The complex role of PHACTR1 in CAD remains multifaceted, with some studies supporting its proatherogenic role, while others point to potential atheroprotective effects[285]. The lead SNP on *PHACTR1*, rs9349379, has been found to be associated with higher EDN1 protein levels[7]. *EDN1* encodes for endothelin-1, a potent vasoconstrictor peptide produced by endothelial cells. Overexpression of *PHACTR1* has been shown to upregulate *EDN1* and downregulate eNOS expression in HUVECs, leading to reduced NO production[277]. Though the precise mechanism by which *PHACTR1* overexpression elevates *EDN1* remains unclear, we observed similar downregulation of *EDN1* in Δ GG aECs, suggesting that the rs6903956 A allele, which drives *PHACTR1* upregulation, may contribute to proinflammatory and vasoconstrictive pathways in CAD by elevating *EDN1* levels.

In silico analysis indicated that HOXA4, MEIS2, and PBX2 are likely to bind to the enhancer region of rs6903956, while AlphaFold 3 predictions suggested specific binding of MEIS1/2 and PBX1/2 to the rs6903956 A risk allele, and HOXA9 and MEIS1 have binding sites on PHACTR1 promoter region. Transcription factor predictions are based on position weight matrices, which quantify the likelihood of binding across various sequences. However, due to structural similarities within the same protein family, pinpointing the exact protein involved in binding can be challenging. To refine our predictions, we used AlphaFold 3 to model protein complex-DNA interactions at rs6903956, focusing on the nearest

amino acid distances between transcription factors and the DNA. Among the potential HOX and MEIS complexes, the HOXA4-MEIS1 combination emerged as the most plausible interactor with the rs6903956 enhancer, based on proximity metrics and structural stability. However, we acknowledge that other members of the HOX and TALE family, such as PBX2 and MEIS2, may also play compensatory roles. Functional redundancy is well-documented among HOX paralogues, and it is possible that multiple complexes contribute to the transcriptional regulation mediated by rs6903956.

HOX and TALE family proteins, such as PBX and MEIS, are well established for conferring target specificity to HOX transcription factors across various biological systems [286-288]. HOX proteins play significant roles in cardiovascular regulation and atherosclerosis pathology[262]. However, the interaction of HOX and TALE family complexes and how it affects cardiovascular diseases is not completely understood. HOXA9 promotes the transcription of *E-selectin* in endothelial cells and enhances endothelial cell migration and tube formation by binding to the *EphB4* promoter [289 290], while HOXA4 acts as a repressor of vascular smooth muscle cell (VSMC) phenotypic switching through inhibition of YAP/TEAD-mediated transcription [291].

MEIS1, while bound to the rs6903956 enhancer region, appear to play a supportive role in the complex as knockdown of MEIS1 did not significantly affect *PHACTR1* expression. Shen et al.[292] were pioneer in reporting the formation of a triple complex between HOXA9, PBX2, and MEIS1 in myeloid leukemia cells. However, their study showed that the inclusion of MEIS1 did not enhance transcriptional activity in a VP16-PBX2 and HOXA9 complex. This is consistent with our findings, where MEIS1 knockdown did not significantly affect *PHACTR1* expression, yet MEIS1 was detected to bind to rs6903956 enhancer by ChIP-qPCR.

While we have not assessed *CXCL12* expression in our base-edited aECs, our prior work showed no significant expression differences in cis-genes *PHACTR1* and *EDN1*, which contrasts we now observe in the base-edited cells. The more precise base editing might yield a more distinct outcome than the broader 63-89bp deletion. Additionally, the observed differences could be attributed to our updated artery-specific endothelial cell differentiation protocol, as opposed to the earlier lateral plate mesoderm differentiation method[178,183].

Our investigation into the 6p24.1 risk locus, identified by GWAS as a contributor to CAD susceptibility, has significantly advanced our understanding of its cis-regulatory mechanisms. By leveraging base editing and aEC models, we have shown that the rs6903956 A allele

acts as an enhancer that drives *PHACTR1* expression through interactions with HOX and TALE family transcription factors. These findings illuminate the complex interplay between non-coding regulatory variants and transcriptional networks, shedding light on the molecular underpinnings of CAD pathogenesis.

Moreover, the allele-specific binding of HOXA4-MEIS1 and the functional validation of their regulatory effects on *PHACTR1* underscore the precision and tissue specificity of transcription factor-driven gene regulation at this locus. Our work highlights how non-coding variants can exert profound effects on endothelial function and inflammation, particularly through pathways involving *PHACTR1*.

By addressing the limitations of previous models and integrating experimental data with *in silico* predictions, Chapter 5 provides a comprehensive framework for dissecting the functional impact of regulatory SNPs in CAD. These insights pave the way for further exploration into transcription factor complexes as potential therapeutic targets, offering new opportunities to mitigate CAD risk by targeting enhancer-promoter interactions at loci like rs6903956.

Conclusion

This dissertation systematically investigates the regulatory role of the East Asian-specific CAD variant rs6903956, revealing its cis- and trans-regulatory mechanisms and providing novel insights into CAD pathogenesis. Each chapter builds upon the last, combining population genomic datasets, creating disease model, and experimental validation methods to connect rs6903956 to molecular and cellular changes driving cardiovascular disease.

Chapter 2 utilized the MESA cohort to establish endothelial-specific effects of rs6903956, demonstrating its association with arterial stiffness and endothelial dysfunction markers. Fine-mapping prioritized rs6903956 as the likely causal variant, with regulatory potential confirmed through ENCODE-derived histone mark analysis. This chapter also identified a LD tandem repeat, rs140361069, unique to East Asians, highlighting the population-specific relevance of this variant in vascular biology.

Chapter 3 established a CRISPR-edited iPSC model by introducing a 63–89 bp deletion encompassing rs6903956. Differentiation into endothelial cells enabled functional and transcriptomic profiling, which revealed dysregulated pathways associated with endothelial dysfunction. Isogenic lines carrying the risk variant exhibited increased ICAM-1 expression and heightened monocyte adhesion, hallmarks of endothelial activation. These findings highlight the role of rs6903956 in mediating endothelial dysfunction, providing a physiologically relevant model for downstream mechanistic studies.

Chapter 4 further investigated the role of rs6903956 using transcriptomic profiling on CRISPR-edited iPSCs, uncovering a trans-regulatory mechanism. Notably, CXCL12, located on chromosome 10q11.21, was identified as a downstream target influenced by rs6903956. Hi-C analysis revealed increased chromatin interactions between the 6p24.1 locus and 10q11.21 in the presence of the risk allele, highlighting long-range enhancer-promoter communication. These findings establish a connection between rs6903956, and pathways involved in vascular inflammation and CAD pathogenesis. However, the broader deletion model used in this study raises the possibility of additional regulatory elements being inadvertently affected, underscoring the need for precise base editing to refine these findings.

Chapter 5 further refined the analysis with precise base editing, converting the risk AA genotype of rs6903956 to the non-risk GG genotype in aECs. Using dual-luciferase assays, we confirmed that

rs6903956 A risk allele creates a TFBS for HOXA4 and MEIS1, which enhances *PHACTR1* expression. Functional assays linked increased *PHACTR1* expression to elevated ICAM-1 levels and monocyte adhesion under disturbed flow conditions, elucidating the cis-regulatory role of rs6903956 in CAD.

The final proposed mechanistic model for rs6903956 in endothelial activation and its contribution to CAD pathogenesis is presented in Figure 26.

While this study resolves key aspects of rs6903956's regulatory role, certain limitations remain. The East Asian-specific effect of rs6903956 remains unexplained, as we ruled out the tandem repeat in LD in our base edited model. Future studies could explore population-specific epigenetic modifications, transcription factor availability, or environmental interactions that may account for this unique effect. Additionally, advanced cellular models, such as 3D organoids or flow-coupled microfluidic systems, could provide a more comprehensive understanding of how rs6903956 influences vascular remodelling under physiologically relevant conditions.

The findings presented in this dissertation open avenues for therapeutic interventions targeting rs6903956-related pathways. For instance, modulating *PHACTR1* or *CXCL12* expression could mitigate pro-inflammatory endothelial phenotypes. Small molecule inhibitors or CRISPR-based gene modulation tools tailored to rs6903956's regulatory network may offer precision medicine approaches for East Asian populations disproportionately affected by CAD.

In conclusion, this work integrates cutting-edge population, epigenetics and cellular approaches to provide a comprehensive understanding of rs6903956 in CAD pathogenesis. By bridging the gap between GWAS findings and mechanistic insights, it establishes a robust framework for studying non-coding variants identified from GWAS to advance precision medicine.

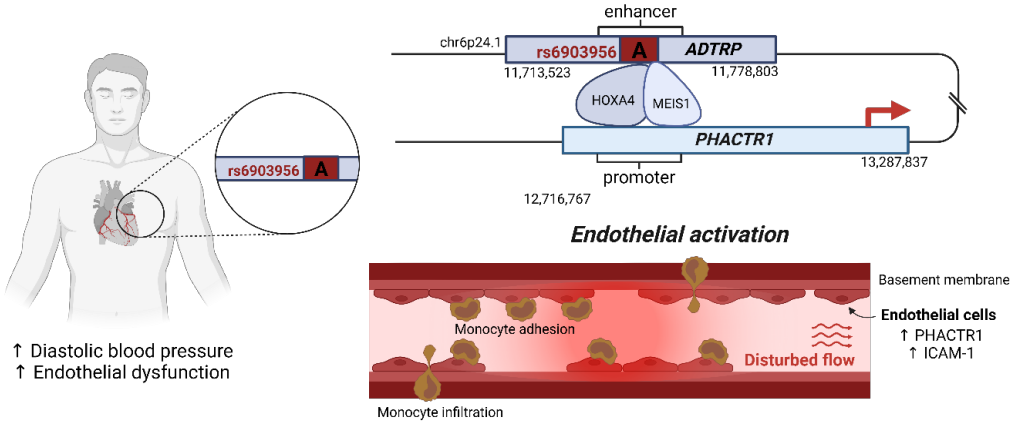


Figure 26: Proposed mechanism of rs6903956 in endothelial activation. Individuals harbouring rs6903956 A risk allele exhibit increased diastolic blood pressure and endothelial activation. Top right panel illustrates the proposed mechanism, where the presence of allele A creates a TFBS for HOXA4 and MEIS1, facilitating an enhancer-promoter interaction leading to increased *PHACTR1* expression. This leads to endothelial activation and heightened monocyte adhesion under disturbed flow conditions, with endothelial cells expressing higher levels of ICAM-1.

References

1. Roth, G.A.*, et al.* Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *J Am Coll Cardiol* **76**, 2982-3021 (2020).
2. Erdmann, J., Kessler, T., Munoz Venegas, L. & Schunkert, H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res* **114**, 1241-1257 (2018).
3. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300-307 (2021).
4. Roadmap Epigenomics, C.*, et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
5. Zhang, K.*, et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001 e5919 (2021).
6. Musunuru, K.*, et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714-719 (2010).
7. Gupta, R.M.*, et al.* A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell* **170**, 522-533 e515 (2017).
8. Krause, M.D.*, et al.* Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. *Proc Natl Acad Sci U S A* **115**, E11349-E11358 (2018).
9. Prestel, M.*, et al.* The Atherosclerosis Risk Variant rs2107595 Mediates Allele-Specific Transcriptional Regulation of HDAC9 via E2F3 and Rb1. *Stroke* **50**, 2651-2660 (2019).
10. Nanda, V.*, et al.* Functional regulatory mechanism of smooth muscle cell-restricted LMOD1 coronary artery disease locus. *PLoS Genet* **14**, e1007755 (2018).
11. Turner, A.W.*, et al.* Functional Analysis of a Novel Genome-Wide Association Study Signal in SMAD3 That Confers Protection From Coronary Artery Disease. *Arterioscler Thromb Vasc Biol* **36**, 972-983 (2016).
12. Bryzgalov, L.O.*, et al.* Detection of regulatory SNPs in human genome using ChIP-seq ENCODE data. *PLoS One* **8**, e78833 (2013).
13. Maurano, M.T.*, et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012).
14. Ernst, J.*, et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49 (2011).
15. Veturi, Y.*, et al.* Modeling Heterogeneity in the Genetic Architecture of Ethnically Diverse Groups Using Random Effect Interaction Models. *Genetics* **211**, 1395-1407 (2019).
16. Martin, A.R.*, et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584-591 (2019).

17. Duncan, L.*, et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* **10**, 3328 (2019).
18. Wang, F.*, et al.* Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat Genet* **43**, 345-349 (2011).
19. Guo, C.Y.*, et al.* Association of SNP rs6903956 on chromosome 6p24.1 with angiographical characteristics of coronary atherosclerosis in a Chinese population. *PLoS One* **7**, e43732 (2012).
20. Meng, J.*, et al.* A coronary artery disease-associated SNP rs6903956 contributed to asymptomatic hyperuricemia susceptibility in Han Chinese. *Lipids Health Dis* **14**, 33 (2015).
21. Tayebi, N.*, et al.* Association of single nucleotide polymorphism rs6903956 on chromosome 6p24.1 with coronary artery disease and lipid levels in different ethnic groups of the Singaporean population. *Clin Biochem* **46**, 755-759 (2013).
22. Dechamethakun, S.*, et al.* Associations between the CDKN2A/B, ADTRP and PDGFD polymorphisms and the development of coronary atherosclerosis in Japanese patients. *J Atheroscler Thromb* **21**, 680-690 (2014).
23. Baird, P.A., Anderson, T.W., Newcombe, H.B. & Lowry, R.B. Genetic disorders in children and young adults: a population study. *Am J Hum Genet* **42**, 677-693 (1988).
24. Baird, P.A. Genetics and health care: a paradigm shift. *Perspect Biol Med* **33**, 203-213 (1990).
25. Strohman, R.C. Ancient genomes, wise bodies, unhealthy people: limits of a genetic paradigm in biology and medicine. *Perspect Biol Med* **37**, 112-145 (1993).
26. Murray, C.J. & Lopez, A.D. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* **349**, 1269-1276 (1997).
27. Mozaffarian, D.*, et al.* Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation* **131**, e29-322 (2015).
28. Nurk, S.*, et al.* The complete sequence of a human genome. *Science* **376**, 44-53 (2022).
29. Genomes Project, C.*, et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
30. Sing, C.F. & Moll, P.P. Genetics of atherosclerosis. *Annu Rev Genet* **24**, 171-187 (1990).
31. Sing, C.F., Haviland, M.B., Templeton, A.R., Zerba, K.E. & Reilly, S.L. Biological complexity and strategies for finding DNA variations responsible for inter-individual variation in risk of a common chronic disease, coronary artery disease. *Ann Med* **24**, 539-547 (1992).
32. Strohman, R. Maneuvering in the complex path from genotype to phenotype. *Science* **296**, 701-703 (2002).

33. Sing, C.F., Haviland, M.B. & Reilly, S.L. Genetic architecture of common multifactorial diseases. *Ciba Found Symp* **197**, 211-229; discussion 229-232 (1996).
34. Plomin, R., Haworth, C.M. & Davis, O.S. Common disorders are quantitative traits. *Nat Rev Genet* **10**, 872-878 (2009).
35. McGue, M., Gottesman, II & Rao, D.C. Resolving genetic models for the transmission of schizophrenia. *Genet Epidemiol* **2**, 99-110 (1985).
36. O'Rourke, D.H., Gottesman, II, Suarez, B.K., Rice, J. & Reich, T. Refutation of the general single-locus model for the etiology of schizophrenia. *Am J Hum Genet* **34**, 630-649 (1982).
37. Risch, N. Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet Epidemiol* **7**, 3-16; discussion 17-45 (1990).
38. Liggett, S.B.*, et al.* A GRK5 polymorphism that inhibits beta-adrenergic receptor signaling is protective in heart failure. *Nat Med* **14**, 510-517 (2008).
39. International HapMap, C. The International HapMap Project. *Nature* **426**, 789-796 (2003).
40. Uffelmann, E.*, et al.* Genome-wide association studies. *Nature Reviews Methods Primers* **1**(2021).
41. Hirschhorn, J.N. & Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**, 95-108 (2005).
42. Barrett, J.C.*, et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* **41**, 703-707 (2009).
43. Xue, A.*, et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* **9**, 2941 (2018).
44. Verstockt, B., Smith, K.G. & Lee, J.C. Genome-wide association studies in Crohn's disease: Past, present and future. *Clin Transl Immunology* **7**, e1001 (2018).
45. Lopez Rodriguez, M., Arasu, U.T. & Kaikkonen, M.U. Exploring the genetic basis of coronary artery disease using functional genomics. *Atherosclerosis* **374**, 87-98 (2023).
46. Kathiresan, S.*, et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**, 189-197 (2008).
47. Nikpay, M.*, et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130 (2015).
48. Stroes, E.*, et al.* Efficacy and Safety of Alirocumab 150 mg Every 4 Weeks in Patients With Hypercholesterolemia Not on Statin Therapy: The ODYSSEY CHOICE II Study. *J Am Heart Assoc* **5**(2016).
49. Sabatine, M.S.*, et al.* Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med* **376**, 1713-1722 (2017).

50. Cerf, M.E. Beta cell dysfunction and insulin resistance. *Front Endocrinol (Lausanne)* **4**, 37 (2013).
51. Emdin, C.A.*, et al.* Phenotypic Consequences of a Genetic Predisposition to Enhanced Nitric Oxide Signaling. *Circulation* **137**, 222-232 (2018).
52. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
53. Tylee, D.S.*, et al.* Genetic correlations among psychiatric and immune-related phenotypes based on genome-wide association data. *Am J Med Genet B Neuropsychiatr Genet* **177**, 641-657 (2018).
54. MacArthur, J.*, et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
55. Crouch, D.J.M. & Bodmer, W.F. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proc Natl Acad Sci U S A* **117**, 18924-18933 (2020).
56. Manolio, T.A.*, et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).
57. Craig, J.E.*, et al.* Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. *Nat Genet* **52**, 160-166 (2020).
58. Khera, A.V.*, et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219-1224 (2018).
59. Mavaddat, N.*, et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* **104**, 21-34 (2019).
60. Wray, N.R.*, et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).
61. Musliner, K.L.*, et al.* Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia With Risk for Depression in the Danish Population. *JAMA Psychiatry* **76**, 516-525 (2019).
62. Zheutlin, A.B.*, et al.* Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *Am J Psychiatry* **176**, 846-855 (2019).
63. Narlikar, L. & Ovcharenko, I. Identifying regulatory elements in eukaryotic genomes. *Brief Funct Genomic Proteomic* **8**, 215-230 (2009).
64. Ghirlando, R.*, et al.* Chromatin domains, insulators, and the regulation of gene expression. *Biochim Biophys Acta* **1819**, 644-651 (2012).
65. Juven-Gershon, T., Hsu, J.Y., Theisen, J.W. & Kadonaga, J.T. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* **20**, 253-259 (2008).
66. Chen, Q.*, et al.* Enhancer RNAs in transcriptional regulation: recent insights. *Front Cell Dev Biol* **11**, 1205540 (2023).

67. Segert, J.A., Gisselbrecht, S.S. & Bulyk, M.L. Transcriptional Silencers: Driving Gene Expression with the Brakes On. *Trends Genet* **37**, 514-527 (2021).
68. Kim, S., Yu, N.K. & Kaang, B.K. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med* **47**, e166 (2015).
69. Kapoor, A.*, et al.* An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval. *Am J Hum Genet* **94**, 854-869 (2014).
70. Lopez Rodriguez, M.*, et al.* Identification and characterization of a FOXA2-regulated transcriptional enhancer at a type 2 diabetes intronic locus that controls GCKR expression in liver cells. *Genome Med* **9**, 63 (2017).
71. Holt, R.J.*, et al.* Allele-specific transcription of the asthma-associated PHD finger protein 11 gene (PHF11) modulated by octamer-binding transcription factor 1 (Oct-1). *J Allergy Clin Immunol* **127**, 1054-1062 e1051-1052 (2011).
72. Rao, S., Yao, Y. & Bauer, D.E. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. *Genome Med* **13**, 41 (2021).
73. Pruim, R.J.*, et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).
74. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
75. Consortium, G.T.*, et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
76. Zhang, W.*, et al.* Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells. *J Biol Chem* **287**, 43137-43155 (2012).
77. Jang, M.K.*, et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell* **19**, 523-534 (2005).
78. Wysocka, J.*, et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90 (2006).
79. Song, L. & Crawford, G.E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* **2010**, pdb prot5384 (2010).
80. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
81. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).

82. Rauluseviciute, I.*, et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174-D182 (2024).
83. Coetzee, S.G., Coetzee, G.A. & Hazelett, D.J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847-3849 (2015).
84. Lieberman-Aiden, E.*, et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).
85. Wei, C.L.*, et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207-219 (2006).
86. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013).
87. Bhattacharya, A.*, et al.* A framework for transcriptome-wide association studies in breast cancer in diverse study populations. *Genome Biol* **21**, 42 (2020).
88. Chen, M., Liu, X., Liu, Q., Shi, D. & Li, H. 3D genomics and its applications in precision medicine. *Cell Mol Biol Lett* **28**, 19 (2023).
89. Hellman, L.M. & Fried, M.G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* **2**, 1849-1861 (2007).
90. Bhattacharyya, S. & Ay, F. Identifying genetic variants associated with chromatin looping and genome function. *Nat Commun* **15**, 8174 (2024).
91. Ong, S.E.*, et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
92. Kim, T.H. & Dekker, J. ChIP-Quantitative Polymerase Chain Reaction (ChIP-qPCR). *Cold Spring Harb Protoc* **2018**(2018).
93. Park, P.J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669-680 (2009).
94. Rossi, M.J., Lai, W.K.M. & Pugh, B.F. Simplified ChIP-exo assays. *Nat Commun* **9**, 2842 (2018).
95. Hainer, S.J. & Fazzio, T.G. High-Resolution Chromatin Profiling Using CUT&RUN. *Curr Protoc Mol Biol* **126**, e85 (2019).
96. Gao, P.*, et al.* Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell* **174**, 576-589 e518 (2018).
97. Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S. & Yang, S.H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
98. Schubert, M.S.*, et al.* Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Sci Rep* **11**, 19482 (2021).
99. Liao, H., Wu, J., VanDusen, N.J., Li, Y. & Zheng, Y. CRISPR-Cas9-mediated homology-directed repair for precise gene editing. *Mol Ther Nucleic Acids* **35**, 102344 (2024).

100. Huang, T.P., Newby, G.A. & Liu, D.R. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat Protoc* **16**, 1089-1128 (2021).
101. Zhang, L., Wu, X. & Hong, L. Endothelial Reprogramming in Atherosclerosis. *Bioengineering (Basel)* **11**(2024).
102. Gutierrez, E.*, et al.* Endothelial dysfunction over the course of coronary artery disease. *Eur Heart J* **34**, 3175-3181 (2013).
103. Pober, J.S. & Sessa, W.C. Evolving functions of endothelial cells in inflammation. *Nat Rev Immunol* **7**, 803-815 (2007).
104. Davignon, J. & Ganz, P. Role of endothelial dysfunction in atherosclerosis. *Circulation* **109**, III27-32 (2004).
105. Widlansky, M.E., Gokce, N., Keaney, J.F., Jr. & Vita, J.A. The clinical implications of endothelial dysfunction. *J Am Coll Cardiol* **42**, 1149-1160 (2003).
106. Farah, C., Michel, L.Y.M. & Balligand, J.L. Nitric oxide signalling in cardiovascular health and disease. *Nat Rev Cardiol* **15**, 292-316 (2018).
107. Moore, K.J., Sheedy, F.J. & Fisher, E.A. Macrophages in atherosclerosis: a dynamic balance. *Nat Rev Immunol* **13**, 709-721 (2013).
108. Kovacic, J.C.*, et al.* Endothelial to Mesenchymal Transition in Cardiovascular Disease: JACC State-of-the-Art Review. *J Am Coll Cardiol* **73**, 190-209 (2019).
109. Andueza, A.*, et al.* Endothelial Reprogramming by Disturbed Flow Revealed by Single-Cell RNA and Chromatin Accessibility Study. *Cell Rep* **33**, 108491 (2020).
110. Bennett, M.R., Sinha, S. & Owens, G.K. Vascular Smooth Muscle Cells in Atherosclerosis. *Circ Res* **118**, 692-702 (2016).
111. Rajendran, P.*, et al.* The vascular endothelium and human diseases. *Int J Biol Sci* **9**, 1057-1069 (2013).
112. Cyr, A.R., Huckaby, L.V., Shiva, S.S. & Zuckerbraun, B.S. Nitric Oxide and Endothelial Dysfunction. *Crit Care Clin* **36**, 307-321 (2020).
113. Durand, M.J. & Gutterman, D.D. Diversity in mechanisms of endothelium-dependent vasodilation in health and disease. *Microcirculation* **20**, 239-247 (2013).
114. Medina-Leyte, D.J.*, et al.* Endothelial Dysfunction, Inflammation and Coronary Artery Disease: Potential Biomarkers and Promising Therapeutical Approaches. *Int J Mol Sci* **22**(2021).
115. Taqueti, V.R. & Di Carli, M.F. Coronary Microvascular Disease Pathogenic Mechanisms and Therapeutic Options: JACC State-of-the-Art Review. *J Am Coll Cardiol* **72**, 2625-2641 (2018).
116. Matsuzawa, Y. & Lerman, A. Endothelial dysfunction and coronary artery disease: assessment, prognosis, and treatment. *Coron Artery Dis* **25**, 713-724 (2014).
117. Sandoo, A., van Zanten, J.J., Metsios, G.S., Carroll, D. & Kitas, G.D. The endothelium and its role in regulating vascular tone. *Open Cardiovasc Med J* **4**, 302-312 (2010).
118. Dart, A.M. & Chin-Dusting, J.P. Lipids and the endothelium. *Cardiovasc Res* **43**, 308-322 (1999).

119. Immanuel, J. & Yun, S. Vascular Inflammatory Diseases and Endothelial Phenotypes. *Cells* **12**(2023).
120. Milosevic, N., Rutter, M. & David, A. Endothelial Cell Adhesion Molecules- (un)Attainable Targets for Nanomedicines. *Front Med Technol* **4**, 846065 (2022).
121. Cotran, R.S. & Pober, J.S. Effects of cytokines on vascular endothelium: their role in vascular and immune injury. *Kidney Int* **35**, 969-975 (1989).
122. Hunt, B.J. & Jurd, K.M. Endothelial cell activation. A central pathophysiological process. *BMJ* **316**, 1328-1329 (1998).
123. Farinacci, M.*, et al.* Circulating endothelial cells as biomarker for cardiovascular diseases. *Res Pract Thromb Haemost* **3**, 49-58 (2019).
124. Burger, D. & Touyz, R.M. Cellular biomarkers of endothelial health: microparticles, endothelial progenitor cells, and circulating endothelial cells. *J Am Soc Hypertens* **6**, 85-99 (2012).
125. Boos, C.J., Lip, G.Y. & Blann, A.D. Circulating endothelial cells in cardiovascular disease. *J Am Coll Cardiol* **48**, 1538-1547 (2006).
126. Blann, A.D.*, et al.* Circulating endothelial cells. Biomarker of vascular disease. *Thromb Haemost* **93**, 228-235 (2005).
127. Heshmatzad, K.*, et al.* Role of non-coding variants in cardiovascular disease. *J Cell Mol Med* **27**, 1621-1636 (2023).
128. Holdt, L.M. & Teupser, D. Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations. *Arterioscler Thromb Vasc Biol* **32**, 196-206 (2012).
129. Holdt, L.M.*, et al.* ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler Thromb Vasc Biol* **30**, 620-627 (2010).
130. Lo Sardo, V.*, et al.* Unveiling the Role of the Most Impactful Cardiovascular Risk Locus through Haplotype Editing. *Cell* **175**, 1796-1810 e1720 (2018).
131. Harismendy, O.*, et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* **470**, 264-268 (2011).
132. Teslovich, T.M.*, et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713 (2010).
133. Myocardial Infarction Genetics, C.*, et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-341 (2009).
134. Rao, A.S. & Knowles, J.W. Polygenic risk scores in coronary artery disease. *Curr Opin Cardiol* **34**, 435-440 (2019).
135. Consortium, C.A.D.*, et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25-33 (2013).
136. 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* **12**, e1001779 (2015).

137. Nagai, A.*, et al.* Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol* **27**, S2-S8 (2017).
138. Koyama, S.*, et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet* **52**, 1169-1177 (2020).
139. Morimoto, T.*, et al.* Significant association of RNF213 p.R4810K, a moyamoya susceptibility variant, with coronary artery disease. *PLoS One* **12**, e0175649 (2017).
140. Park, M.G., Shin, J.H., Lee, S.W., Park, H.R. & Park, K.P. RNF213 rs112735431 polymorphism in intracranial artery steno-occlusive disease and moyamoya disease in Koreans. *J Neurol Sci* **375**, 331-334 (2017).
141. Samani, N.J.*, et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* **357**, 443-453 (2007).
142. Schunkert, H.*, et al.* Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* **117**, 1675-1684 (2008).
143. Sherry, S.T., Ward, M. & Sirotkin, K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* **9**, 677-679 (1999).
144. Luo, C.*, et al.* GATA2 regulates the CAD susceptibility gene ADTRP rs6903956 through preferential interaction with the G allele. *Mol Genet Genomics* **296**, 799-808 (2021).
145. Lupu, C., Zhu, H., Popescu, N.I., Wren, J.D. & Lupu, F. Novel protein ADTRP regulates TFPI expression and function in human endothelial cells in normal conditions and in response to androgen. *Blood* **118**, 4463-4471 (2011).
146. Patel, M.M.*, et al.* Role of ADTRP (Androgen-Dependent Tissue Factor Pathway Inhibitor Regulating Protein) in Vascular Development and Function. *J Am Heart Assoc* **7**, e010690 (2018).
147. Erikci Ertunc, M.*, et al.* AIG1 and ADTRP are endogenous hydrolases of fatty acid esters of hydroxy fatty acids (FAHFAs) in mice. *J Biol Chem* **295**, 5891-5905 (2020).
148. Parsons, W.H.*, et al.* AIG1 and ADTRP are atypical integral membrane hydrolases that degrade bioactive FAHFAs. *Nat Chem Biol* **12**, 367-372 (2016).
149. Ooi, D.S.Q.*, et al.* Detection of ADTRP in circulation and its role as a novel biomarker for coronary artery disease. *PLoS One* **15**, e0237074 (2020).
150. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504 (2018).
151. Yuan, K.*, et al.* Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nat Genet* **56**, 1841-1850 (2024).
152. Bild, D.E.*, et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol* **156**, 871-881 (2002).

153. Rojano, E., Seoane, P., Ranea, J.A.G. & Perkins, J.R. Regulatory variants: from detection to predicting impact. *Brief Bioinform* **20**, 1639-1654 (2019).
154. Srivastava, D. & Mahony, S. Sequence and chromatin determinants of transcription factor binding and the establishment of cell type-specific binding patterns. *Biochim Biophys Acta Gene Regul Mech* **1863**, 194443 (2020).
155. Schnitzler, G.R.*, et al.* Convergence of coronary artery disease genes onto endothelial cell programs. *Nature* **626**, 799-807 (2024).
156. Solomon, C.U.*, et al.* Effects of Coronary Artery Disease-Associated Variants on Vascular Smooth Muscle Cells. *Circulation* **146**, 917-929 (2022).
157. Yeboah, J.*, et al.* Predictive value of brachial flow-mediated dilation for incident cardiovascular events in a population-based study: the multi-ethnic study of atherosclerosis. *Circulation* **120**, 502-509 (2009).
158. Purcell, S.*, et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
159. Nakato, R.*, et al.* Comprehensive epigenome characterization reveals diverse transcriptional regulation across human vascular endothelial cells. *Epigenetics Chromatin* **12**, 77 (2019).
160. Machiela, M.J. & Chanock, S.J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557 (2015).
161. Degtyareva, A.O., Antontseva, E.V. & Merkulova, T.I. Regulatory SNPs: Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases. *Int J Mol Sci* **22**(2021).
162. Maurano, M.T.*, et al.* Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat Genet* **47**, 1393-1401 (2015).
163. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
164. Yu, J.*, et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920 (2007).
165. Takahashi, K.*, et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).
166. Scesa, G., Adami, R. & Bottai, D. iPSC Preparation and Epigenetic Memory: Does the Tissue Origin Matter? *Cells* **10**(2021).
167. Haridhasapavalan, K.K.*, et al.* An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. *Gene* **686**, 146-159 (2019).
168. Rivera, T., Zhao, Y., Ni, Y. & Wang, J. Human-Induced Pluripotent Stem Cell Culture Methods Under cGMP Conditions. *Curr Protoc Stem Cell Biol* **54**, e117 (2020).

169. Baghbaderani, B.A.*, et al.* Detailed Characterization of Human Induced Pluripotent Stem Cells Manufactured for Therapeutic Applications. *Stem Cell Rev Rep* **12**, 394-420 (2016).
170. Zhou, Y.Y. & Zeng, F. Integration-free methods for generating induced pluripotent stem cells. *Genomics Proteomics Bioinformatics* **11**, 284-287 (2013).
171. Knoepfler, P.S. Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine. *Stem Cells* **27**, 1050-1056 (2009).
172. Gu, M.*, et al.* Pravastatin reverses obesity-induced dysfunction of induced pluripotent stem cell-derived endothelial cells via a nitric oxide-dependent mechanism. *Eur Heart J* **36**, 806-816 (2015).
173. Palpant, N.J.*, et al.* Generating high-purity cardiac and endothelial derivatives from patterned mesoderm using human pluripotent stem cells. *Nat Protoc* **12**, 15-31 (2017).
174. Sa, S.*, et al.* Induced Pluripotent Stem Cell Model of Pulmonary Arterial Hypertension Reveals Novel Gene Expression and Patient Specificity. *Am J Respir Crit Care Med* **195**, 930-941 (2017).
175. Yang, L.*, et al.* Differentiation of Human Induced-Pluripotent Stem Cells into Smooth-Muscle Cells: Two Novel Protocols. *PLoS One* **11**, e0147155 (2016).
176. Tousoulis, D., Charakida, M. & Stefanadis, C. Endothelial function and inflammation in coronary artery disease. *Heart* **92**, 441-444 (2006).
177. Pacinella, G., Ciaccio, A.M. & Tuttolomondo, A. Endothelial Dysfunction and Chronic Inflammation: The Cornerstones of Vascular Alterations in Age-Related Diseases. *Int J Mol Sci* **23**(2022).
178. Narmada, B.C.*, et al.* Human Stem Cell-Derived Endothelial-Hepatic Platform for Efficacy Testing of Vascular-Protective Metabolites from Nutraceuticals. *Stem Cells Transl Med* **6**, 851-863 (2017).
179. Miyaoka, Y.*, et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep* **6**, 23549 (2016).
180. Ang, L.T.*, et al.* A Roadmap for Human Liver Differentiation from Pluripotent Stem Cells. *Cell Rep* **22**, 2190-2205 (2018).
181. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1911 (2018).
182. Anzalone, A.V.*, et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157 (2019).
183. Cheung, C., Bernardo, A.S., Trotter, M.W., Pedersen, R.A. & Sinha, S. Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nat Biotechnol* **30**, 165-173 (2012).
184. Doench, J.G.*, et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).

185. Mashiko, D.*, et al.* Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci Rep* **3**, 3355 (2013).
186. Davis, C.M., Fairbanks, S.L. & Alkayed, N.J. Mechanism of the sex difference in endothelial dysfunction after stroke. *Transl Stroke Res* **4**, 381-389 (2013).
187. Vinas, J.L.*, et al.* Sex diversity in proximal tubule and endothelial gene expression in mice with ischemic acute kidney injury. *Clin Sci (Lond)* **134**, 1887-1909 (2020).
188. Joseph, P.G., Pare, G. & Anand, S.S. Exploring gene-environment relationships in cardiovascular disease. *Can J Cardiol* **29**, 37-45 (2013).
189. Talmud, P.J. Gene-environment interaction and its impact on coronary heart disease risk. *Nutr Metab Cardiovasc Dis* **17**, 148-152 (2007).
190. Ang, L.T.*, et al.* Generating human artery and vein cells from pluripotent stem cells highlights the arterial tropism of Nipah and Hendra viruses. *Cell* **185**, 2523-2541 e2530 (2022).
191. Kwon, H.B.*, et al.* In vivo modulation of endothelial polarization by Apelin receptor signalling. *Nat Commun* **7**, 11805 (2016).
192. Steward, R., Jr., Tambe, D., Hardin, C.C., Krishnan, R. & Fredberg, J.J. Fluid shear, intercellular stress, and endothelial cell alignment. *Am J Physiol Cell Physiol* **308**, C657-664 (2015).
193. Chiu, J.J. & Chien, S. Effects of disturbed flow on vascular endothelium: pathophysiological basis and clinical perspectives. *Physiol Rev* **91**, 327-387 (2011).
194. Chien, S. Effects of disturbed flow on endothelial cells. *Ann Biomed Eng* **36**, 554-562 (2008).
195. Bai, J., Wang, Y.J., Liu, L. & Zhao, Y.L. Ephrin B2 and EphB4 selectively mark arterial and venous vessels in cerebral arteriovenous malformation. *J Int Med Res* **42**, 405-415 (2014).
196. Mussbacher, M.*, et al.* More than Just a Monolayer: the Multifaceted Role of Endothelial Cells in the Pathophysiology of Atherosclerosis. *Curr Atheroscler Rep* **24**, 483-492 (2022).
197. Nica, A.C. & Dermitzakis, E.T. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120362 (2013).
198. Yao, C.*, et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *Am J Hum Genet* **100**, 571-580 (2017).
199. Liu, X., Li, Y.I. & Pritchard, J.K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034 e1026 (2019).
200. Pierce, B.L.*, et al.* Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. *PLoS Genet* **10**, e1004818 (2014).
201. Shan, N., Wang, Z. & Hou, L. Identification of trans-eQTLs using mediation analysis with multiple mediators. *BMC Bioinformatics* **20**, 126 (2019).

202. Albert, F.W., Bloom, J.S., Siegel, J., Day, L. & Kruglyak, L. Genetics of trans-regulatory variation in gene expression. *Elife* **7**(2018).
203. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol* **20**, 535-550 (2019).
204. Zhao, M.T.*, et al.* Cell Type-Specific Chromatin Signatures Underline Regulatory DNA Elements in Human Induced Pluripotent Stem Cells and Somatic Cells. *Circ Res* **121**, 1237-1250 (2017).
205. Whyte, W.A.*, et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).
206. Hnisz, D.*, et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).
207. Blayney, J.W.*, et al.* Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell* **186**, 5826-5839 e5818 (2023).
208. Hansen, A.S. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* **11**, 132-148 (2020).
209. Hakim, O. & Misteli, T. SnapShot: Chromosome confirmation capture. *Cell* **148**, 1068 e1061-1062 (2012).
210. Durand, N.C.*, et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
211. Rao, S.S.*, et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
212. Tay, K.Y.*, et al.* Trans-interaction of risk loci 6p24.1 and 10q11.21 is associated with endothelial damage in coronary artery disease. *Atherosclerosis* **362**, 11-22 (2022).
213. Farouk, S.S., Rader, D.J., Reilly, M.P. & Mehta, N.N. CXCL12: a new player in coronary disease identified through human genetics. *Trends Cardiovasc Med* **20**, 204-209 (2010).
214. Mehta, N.N.*, et al.* The novel atherosclerosis locus at 10q11 regulates plasma CXCL12 levels. *Eur Heart J* **32**, 963-971 (2011).
215. Salcedo, R. & Oppenheim, J.J. Role of chemokines in angiogenesis: CXCL12/SDF-1 and CXCR4 interaction, a key regulator of endothelial cell responses. *Microcirculation* **10**, 359-370 (2003).
216. Cambier, S., Gouwy, M. & Proost, P. The chemokines CXCL8 and CXCL12: molecular and functional properties, role in disease and efforts towards pharmacological intervention. *Cell Mol Immunol* **20**, 217-251 (2023).
217. Vagesjo, E.*, et al.* Accelerated wound healing in mice by on-site production and delivery of CXCL12 by transformed lactic acid bacteria. *Proc Natl Acad Sci U S A* **115**, 1895-1900 (2018).

218. Li, W.*, et al.* Peripheral nerve-derived CXCL12 and VEGF-A regulate the patterning of arterial vessel branching in developing limb skin. *Dev Cell* **24**, 359-371 (2013).
219. Janssens, R.*, et al.* Truncation of CXCL12 by CD26 reduces its CXC chemokine receptor 4- and atypical chemokine receptor 3-dependent activity on endothelial cells and lymphocytes. *Biochem Pharmacol* **132**, 92-101 (2017).
220. Mirshahi, F.*, et al.* SDF-1 activity on microvascular endothelial cells: consequences on angiogenesis in in vitro and in vivo models. *Thromb Res* **99**, 587-594 (2000).
221. Sjaarda, J.*, et al.* Blood CSF1 and CXCL12 as Causal Mediators of Coronary Artery Disease. *J Am Coll Cardiol* **72**, 300-310 (2018).
222. Tavakolian Ferdousie, V.*, et al.* Serum CXCL10 and CXCL12 chemokine levels are associated with the severity of coronary artery disease and coronary artery occlusion. *Int J Cardiol* **233**, 23-28 (2017).
223. Döring, Y.*, et al.* CXCL12 Derived From Endothelial Cells Promotes Atherosclerosis to Drive Coronary Artery Disease. *Circulation* **139**, 1338-1340 (2019).
224. Zernecke, A. & Weber, C. Chemokines in Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **34**, 742-750 (2014).
225. Wang, Y., Huang, J., Li, Y. & Yang, G.Y. Roles of chemokine CXCL12 and its receptors in ischemic stroke. *Curr Drug Targets* **13**, 166-172 (2012).
226. Gencer, S., Evans, B.R., van der Vorst, E.P.C., Doring, Y. & Weber, C. Inflammatory Chemokines in Atherosclerosis. *Cells* **10**(2021).
227. Gao, J.H., Yu, X.H. & Tang, C.K. CXC chemokine ligand 12 (CXCL12) in atherosclerosis: An underlying therapeutic target. *Clin Chim Acta* **495**, 538-544 (2019).
228. Gao, J.H.*, et al.* CXCL12 promotes atherosclerosis by downregulating ABCA1 expression via the CXCR4/GSK3beta/beta-catenin(T120)/TCF21 pathway. *J Lipid Res* **60**, 2020-2033 (2019).
229. Hostler, A.C.*, et al.* Endothelial-specific CXCL12 regulates neovascularization during tissue repair and tumor progression. *FASEB J* **38**, e70210 (2024).
230. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
231. Ritchie, M.E.*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
232. Ashburner, M.*, et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
233. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide

expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
234. Kramer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530 (2014).
235. Taberlay, P.C.*, et al.* Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res* **26**, 719-731 (2016).
236. Hagege, H.*, et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**, 1722-1733 (2007).
237. Nora, E.P.*, et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).
238. Mootha, V.K.*, et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).
239. Samani, N.J.*, et al.* Large scale association analysis of novel genetic loci for coronary artery disease. *Arterioscler Thromb Vasc Biol* **29**, 774-780 (2009).
240. Zhang, J.*, et al.* Variants in the CXCL12 gene was associated with coronary artery disease susceptibility in Chinese Han population. *Oncotarget* **8**, 54518-54527 (2017).
241. Wang, X., Cairns, M.J. & Yan, J. Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Res* **47**, 11481-11496 (2019).
242. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* **4**, 895-901 (2007).
243. McCord, R.P., Kaplan, N. & Giorgetti, L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol Cell* **77**, 688-708 (2020).
244. Du, M. & Wang, L. 3C-digital PCR for quantification of chromatin interactions. *BMC Mol Biol* **17**, 23 (2016).
245. Hebbel, R.P. Blood endothelial cells: utility from ambiguity. *J Clin Invest* **127**, 1613-1615 (2017).
246. Chioh, F.W.*, et al.* Convalescent COVID-19 patients are susceptible to endothelial dysfunction due to persistent immune activation. *Elife* **10**(2021).
247. Spielmann, M. & Mundlos, S. Looking beyond the genes: the role of non-coding variants in human disease. *Hum Mol Genet* **25**, R157-R165 (2016).
248. Tan, W.L.W.*, et al.* Epigenomes of Human Hearts Reveal New Genetic Variants Relevant for Cardiac Disease and Phenotype. *Circ Res* **127**, 761-777 (2020).
249. Sun, J.H.*, et al.* Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell* **175**, 224-238 e215 (2018).

250. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).
251. Hill, J.M.*, et al.* Circulating endothelial progenitor cells, vascular function, and cardiovascular risk. *N Engl J Med* **348**, 593-600 (2003).
252. Nadar, S.K., Lip, G.Y., Lee, K.W. & Blann, A.D. Circulating endothelial cells in acute ischaemic stroke. *Thromb Haemost* **94**, 707-712 (2005).
253. Schmidt, D.E., Manca, M. & Hoefer, I.E. Circulating endothelial cells in coronary artery disease and acute coronary syndrome. *Trends Cardiovasc Med* **25**, 578-587 (2015).
254. Icli, B. & Feinberg, M.W. Plasticity of Arterial and Venous Endothelial Cell Identity: Some Nerve! *Circ Res* **119**, 574-576 (2016).
255. dela Paz, N.G. & D'Amore, P.A. Arterial versus venous endothelial cells. *Cell Tissue Res* **335**, 5-16 (2009).
256. Trimm, E. & Red-Horse, K. Vascular endothelial cell development and diversity. *Nat Rev Cardiol* **20**, 197-210 (2023).
257. Lambert, S.A.*, et al.* The Human Transcription Factors. *Cell* **175**, 598-599 (2018).
258. Isbel, L., Grand, R.S. & Schubeler, D. Generating specificity in genome regulation through transcription factor sensitivity to chromatin. *Nat Rev Genet* **23**, 728-740 (2022).
259. Chen, H. & Pugh, B.F. What do Transcription Factors Interact With? *J Mol Biol* **433**, 166883 (2021).
260. Mattioli, K.*, et al.* Cis and trans effects differentially contribute to the evolution of promoters and enhancers. *Genome Biol* **21**, 210 (2020).
261. Paul, S., Zhang, X. & He, J.Q. Homeobox gene Meis1 modulates cardiovascular regeneration. *Semin Cell Dev Biol* **100**, 52-61 (2020).
262. Zhou, Y., Wu, Q. & Guo, Y. Deciphering the emerging landscape of HOX genes in cardiovascular biology, atherosclerosis and beyond (Review). *Int J Mol Med* **53**(2024).
263. Abramson, J.*, et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).
264. Hill, M.S., Vande Zande, P. & Wittkopp, P.J. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet* **22**, 203-215 (2021).
265. Jiang, D.*, et al.* Endothelial PHACTR1 Promotes Endothelial Activation and Atherosclerosis by Repressing PPARgamma Activity Under Disturbed Flow in Mice. *Arterioscler Thromb Vasc Biol* **43**, e303-e322 (2023).
266. Gu, K., Zhang, Y., Sun, K. & Jiang, X. Associations between PHACTR1 gene polymorphisms and pulse pressure in Chinese Han population. *Biosci Rep* **40**(2020).
267. Fujimaki, T.*, et al.* Association of a transcription factor 21 gene polymorphism with hypertension. *Biomed Rep* **3**, 118-122 (2015).

268. Zhang, D.*, et al.* Genome-wide linkage and association scans for pulse pressure in Chinese twins. *Hypertens Res* **35**, 1051-1057 (2012).
269. Surendran, P.*, et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-1161 (2016).
270. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity between motifs. *Genome Biol* **8**, R24 (2007).
271. Mann, R.S. & Affolter, M. Hox proteins meet more partners. *Curr Opin Genet Dev* **8**, 423-429 (1998).
272. Fishilevich, S.*, et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**(2017).
273. Bobola, N. & Sagerstrom, C.G. TALE transcription factors: Cofactors no more. *Semin Cell Dev Biol* **152-153**, 76-84 (2024).
274. Kittke, V.*, et al.* RLS-associated MEIS transcription factors control distinct processes in human neural stem cells. *Sci Rep* **14**, 28986 (2024).
275. Schulte, D. & Geerts, D. MEIS transcription factors in development and disease. *Development* **146**(2019).
276. Zhou, D., Shao, L. & Spitz, D.R. Reactive oxygen species in normal and tumor stem cells. *Adv Cancer Res* **122**, 1-67 (2014).
277. Ma, X.*, et al.* PHACTR1, a coronary artery disease risk gene, mediates endothelial dysfunction. *Front Immunol* **13**, 958677 (2022).
278. Gijsberts, C.M.*, et al.* Ethnicity Modifies Associations between Cardiovascular Risk Factors and Disease Severity in Parallel Dutch and Singapore Coronary Cohorts. *PLoS One* **10**, e0132278 (2015).
279. O'Donnell, C.J.*, et al.* Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation* **124**, 2855-2864 (2011).
280. Anttila, V.*, et al.* Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat Genet* **45**, 912-917 (2013).
281. Debette, S.*, et al.* Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat Genet* **47**, 78-83 (2015).
282. Kiando, S.R.*, et al.* PHACTR1 Is a Genetic Susceptibility Locus for Fibromuscular Dysplasia Supporting Its Complex Genetic Pattern of Inheritance. *PLoS Genet* **12**, e1006367 (2016).
283. Jarray, R.*, et al.* Depletion of the novel protein PHACTR-1 from human endothelial cells abolishes tube formation and induces cell death receptor apoptosis. *Biochimie* **93**, 1668-1675 (2011).
284. Zhang, Z., Jiang, F., Zeng, L., Wang, X. & Tu, S. PHACTR1 regulates oxidative stress and inflammation to coronary artery endothelial cells via interaction with NF-kappaB/p65. *Atherosclerosis* **278**, 180-189 (2018).
285. Rezvan, A. PHACTR1 and Atherosclerosis: It's Complicated. *Arterioscler Thromb Vasc Biol* **43**, 1409-1411 (2023).

286. Merabet, S. & Mann, R.S. To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends in Genetics* **32**, 334-347 (2016).
287. Penkov, D.*, et al.* Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell Rep* **3**, 1321-1333 (2013).
288. Sarno, J.L., Kliman, H.J. & Taylor, H.S. HOXA10, Pbx2, and Meis1 protein expression in the human endometrium: formation of multimeric complexes on HOXA10 target genes. *J Clin Endocrinol Metab* **90**, 522-528 (2005).
289. Bandyopadhyay, S., Ashraf, M.Z., Daher, P., Howe, P.H. & DiCorleto, P.E. HOXA9 participates in the transcriptional activation of E-selectin in endothelial cells. *Mol Cell Biol* **27**, 4207-4216 (2007).
290. Bruhl, T.*, et al.* Homeobox A9 transcriptionally regulates the EphB4 receptor to modulate endothelial cell migration and tube formation. *Circ Res* **94**, 743-751 (2004).
291. Kimura, M.*, et al.* Homeobox A4 suppresses vascular remodeling by repressing YAP/TEAD transcriptional activity. *EMBO Rep* **21**, e48389 (2020).
292. Shen, W.F.*, et al.* HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells. *Mol Cell Biol* **19**, 3051-3061 (1999).